Conversational Agents in Virtual Worlds: Bridging Disciplines

Abstract

This paper examines the effective deployment of conversational agents in virtual worlds from the perspective of researchers/practitioners in cognitive psychology, computing science, learning technologies, and engineering. From a cognitive perspective, the major challenge lies in the coordination and management of the various channels of information associated with conversation/communication and integrating this information with the virtual space of the environment and the belief space of the user. From computing science, the requirements include conversational competency, use of nonverbal cues, animation consistent with affective states, believability, domain competency, and user adaptability. From a learning technologies perspective, the challenge is to maximize the considerable affordances provided by conversational avatars in virtual worlds balanced against ecologically valid investigations regarding utility. Finally, the engineering perspective focuses on the technical competency required to implement effective and functional agents, and the associated costs to enable student access. Taken together, the four perspectives draw attention to the quality of the agent-user interaction, how theory, practice, and research are closely intertwined, and the multidisciplinary nature of this area with opportunities for cross fertilization and collaboration.

Veletsianos, G., Heller, R., Overmyer, S., & Procter, M. (2010). Conversational Agents in Virtual Worlds: Bridging Disciplines. *British Journal of Educational Technology*, *41*(1), 123-140.

Conversational Agents in Virtual Worlds: Bridging Disciplines

Conversational agents able to interact with users have been deployed in numerous technology-enhanced learning environments, including video games, standalone applications, and virtual worlds. Recent developments in virtual world technology have brought virtual characters (or, avatars) at the forefront of research and development, with researchers proposing that knowledge gained from virtual world predecessors (e.g., MOOs) can assist researchers and designers in implementing contemporary avatars and virtual characters (Mazar & Nolan, 2008). Indeed, virtual world researchers can learn from research on MOOs and MUDs, but they can also learn quite a lot from research and development work in diverse disciplines including online learning, instructional design, educational psychology, psycholinguistics, information systems, and human-computer interaction. This is the basic premise of this manuscript. Since research and practice in virtual worlds transpires across numerous disciplines that are varied but interconnected, we can learn a lot from each other by collaborating across disciplinary lines. Thus, in this paper, we hope to provide a space where four researchers/practitioners from four diverse but converging disciplines have come together to answer one question: What are the barriers to the effective deployment and implementation of conversational avatars in virtual worlds (in teaching and learning contexts)? Effective and engaging deployment of virtual characters in educational settings has so far proven difficult (Dehn & van Mulken, 2000, Gulz & Haake, 2006), partly because collaboration across disciplinary lines seems to be limited. We hope that the ideas presented herein will lead to synergies and collaborative work across disciplines.

To ensure consistency across the contributions we situate the question in the context of geriatric avatar-patients in medical simulations. More specifically, the contributions from each discipline will assume that the avatar discussed will be a geriatric avatar with which medical students (represented by avatars themselves) can hold conversations such that they engage in the diagnosis of certain conditions based on the geriatric avatar's input.

The paper proceeds by presenting the perspectives of a cognitive psychologist, a computer scientist/human-computer interaction expert, an engineering practitioner, and a learning technologies researcher. Contributions are then summarized and future research and development work is proposed. It is important to note that within this paper, research and practice are inextricably intertwined. A second point to keep in mind while reading this paper is the fact that references to animated pedagogical agents, conversational agents, chat bots, conversational avatars, and virtual characters, represent the same thing: virtual representations embedded in learning environments that serve pedagogical purposes. The variability in the terminology stems partly from lack of collaboration across disciplinary lines and partly from researchers wanting to highlight a specific part of the character. Although we could have chosen a uniform term to use across the manuscript, we decided that using these varied terms highlights the status quo and serves to illustrate the fact that different disciplines use different terminology for essentially similar items. Finally, while the word avatar will probably remind readers of virtual personas of real people, in this manuscript we are concerned with autonomous avatars whose activities (e.g., movement, responses, and speech) are controlled by a back-end technology such as an artificial intelligence engine.

Conversational Agents in Virtual Worlds: A Cognitive Perspective

The author of this section was trained as an Experimental Psychologist in memory and language, with an information processing perspective on cognitive function. Hence, much of his focus was on the user's internal experiences defined and measured using experimental methods.

Since this author's training, theories of cognitive function have either reached outside of the black box to recognize the critical role of the environment in the creation of meaning (i.e. situated cognition, see Brown, Collins, and Duguid, 1989) or deeper inside the black box to connect with neurons as the biological bases of knowledge (i.e. connectionism, see Rummelhart & McClelland, 1978). The author's interests tend to be outside the box, especially how the environment can be extended to support limited cognitive function in special populations (Heller, Dobbs, & Strain, 2009; Waller, 2007).

Conversational agents are well-poised to support cognitive function at any level. There are, however, barriers to their effective deployment that ironically are related to the strengths they enable. Specifically, conversational agents that are anthropomorphised as avatars in a virtual environment provide multiple channels of information that must be managed and coordinated such that agent-user interactions unfold naturally and efficiently. It is the presence of anomalous information or conflicting evidence between channels that can usurp the cognitive resources intended for a learning experience. In other words, the additional information afforded by a conversational agent avatar in a virtual space should not be a distraction but rather support the cognitive processes being trained. Information that is superfluous to the task at hand needs to be examined for ways of increasing its relevance or possible elimination. To illustrate, we developed a historical figure application of a conversational agent based on Sigmund Freud and evaluated its performance under three conditions; agent without any visual representation, agent with a static image of Freud, and agent with an animated image of Freud. We found that the ratings of learner satisfaction and conversational agent performance were highest in the first condition in which no visual or auditory information was provided and significantly different than the other two conditions with visual information (Heller & Procter, 2009). Other researchers have reported this all or nothing effect in the investigation of pedagogical agents (Dirkin, Mishra, & Altermatt, 2005). We hypothesized that the visual information in the latter two conditions was poorly integrated with the conversational dialogue and subsequently distracted the user leading to poorer ratings. Thus, the informationally rich environment afforded by a virtual world with an animated avatar needs to be carefully integrated with the conversational agent's performance as the consequence of mismatches can be significant and may undermine the learning activity. Moreover, the user's expectations and beliefs must also be managed and coordinated with the learning activity since mismatches at this level can also be a distraction to the task at hand.

In managing and coordinating the multiple channels of information afforded by an anthropomorphised conversational agent in a virtual world, it may be useful to examine findings from research on Animated Pedagogical Agents (APAs). In particular, the auditory channel is noteworthy given the report of a modality effect such that learning outcomes are improved when APAs use spoken voice as opposed to text only (Moreno & Mayer, 2002; Craig, Gholson, & Driscoll, 2002). Moreover, the quality of the voice, as well as its expressive capabilities (Veletsianos, 2009; Veletsianos, Miller & Doering, 2010) can further influence learning outcomes (Mayer, Sobko, & Mautone, 2003). Clearly, an age appropriate voice would be required for a geriatric patient. However, using voice places additional constraints on avatar behaviour as voice needs to be coordinated with lip synchronization (matching the phonological properties of text with the associated lip movements). Voice without lip synchronization may be seen as distracting and likely experienced as less satisfying and immersive. Voice also needs to be modulated to reflect the natural speech patterns around syntactic units and pauses, which, both filled and unfilled, should be inserted and consistent with their pragmatic function.

In addition to managing the auditory properties of voice, the semantic properties of the response need to synchronized or coordinated with the facial expressions, gaze location, and/or body gestures. For instance, if the conversation focuses on the avatars arm, the avatar should look at her arm, perhaps hold it, or examine it. If the arm is causing pain, the facial expression should somehow reflect the degree of discomfort. In particular, facial expressions incongruous with semantic content could be seen as distracting and perhaps incorrectly diagnostic for disorders related to mood.

Finally, pragmatic aspects of the conversation also need to be managed and coordinated. Cassell and Thorisson (1999) found that nonverbal behaviours related to turn taking (i.e. gaze locations, head movements, body gestures) in the process of a conversation were associated with positive agent evaluations. They used the term Envelope Feedback and found that such feedback was even more important than the emotional expressions provided by the agents. Nonverbal behaviours can also be used to support other pragmatic aspects of communication like humour and indirect requests. Wallis and Norling (2003) argue that pragmatic functions or social intelligence is a key area for development if conversational agents, anthropomorphized or not, are to make any headway in the simulation of social relations.

In addition to these internal features ostensibly embedded within the patient avatar, there are the external features both directly reflected in the virtual environment and indirectly in the beliefs of the user about their role in the simulation. The objects and features of the environment must be built to support the simulation. The choice of objects and environmental features should extend and enhance the simulation and should not distract or be irrelevant to the simulation. The intelligence of the patient should recognize these objects and features when referenced in speech or behaviour and respond accordingly. The objects and features are an important part of the shared world between the patient and the user and must be used to support and enhance the interaction.

Finally, the user's beliefs and expectations regarding their interaction with the patient play a critical role in how the simulation unfolds. The user's expectations about the patient's conversational capability need to be clear as do the rules for interacting. Although users are provided with this information up front, in the course of an interaction, a different set of beliefs may come to operate. The virtual patient needs to act in ways that direct users to consider their responsibilities in the completion of the simulation. It is another aspect of the shared world view (albeit a small one) that needs to be reflected in the strategies virtual patients use in novel situations. In particular, the virtual patient needs to know when the user is completing a clinical interview or delivering the results of a diagnostic test.

In sum, there are multiple channels of information that can be used to support the cognitive operations required for natural and efficient agent interactions. These channels were identified from a cognitive perspective and it is argued that their integration and coordination is a critical task for the effective deployment of conversational agents in virtual worlds.

Conversational Agents in Virtual Worlds: Perspectives from Computer Science and Human-Computer Interaction

The author of this section has a background in software engineering and human-computer interaction, in both academia and industry.

Conversational and pedagogical agents in 3D virtual worlds and intelligent tutoring systems are receiving considerable attention from researchers in human-centred computing. There are a number of approaches to the creation of pedagogical agents ranging from very simple deterministic solutions to complex cognitive simulations. We are very encouraged by

Silverman et al (2006), who conclude that useful and realistic models already exist for synthetic agents, but have yet to be in widespread application. These authors further conclude that agent reuse is important to reduce development time and effort. Finally, their analysis suggests that human performance simulators are worth doing, and feasible.

The nature of the design of pedagogical agents has also become a subject of considerable interest. Using pedagogical agents as anthropomorphized learning companions in a similar manner as we are discussing here, Kim and Baylor (2006), found that the effectiveness of agents and perception of students of the agents depended on the level of competence of the agent in relationship to the level of competence of the student. This finding was extended in Kim (2007) which revealed that students prefer agents similar to themselves, and that low competency agents may help low competency students gain confidence and achieve better learning outcomes. This strongly suggests that pedagogical agents should be adaptive, and adapt according to some sort of knowledge or academic competence model of the student.

Baylor (2001) suggests a number of possible roles for pedagogical agents, with agents also being able to serve as domain authorities and repositories of conversational knowledge. This kind of agent might be the manager of an activity or enterprise, who serves to be the "gatekeeper" of information and authority regarding their domain of expertise. The role of this agent is to provide information when asked about a domain which a student is motivated to explore, such as, in the present context, patient symptoms and pathology. This domain information is contributory to the primary learning objective, and essential to student learning. Baylor reaches additional conclusions regarding the nature of the learner-agent relationship, and suggests that the learner must have confidence in the agent for the agent to be effective. In order to gain the learner's confidence, she concludes that, "the agent must be believable, motivational, competent, and trustworthy, all of which require the agent to assert control in the learning process" (p. 413).

Emerging 3D virtual training environments (e.g., OLIVE; see Greenburg, 2008) have been largely without intelligent thinking and talking agents, but rather have relied upon humans as part of the training simulation. This in itself is an interesting and useful application of virtual world technology. However such environments require a large cast of characters to be online and are not useful enough as standalone intelligent tutoring systems. This requirement is usually too onerous to be met, for reasons including the cost and unpredictability of human specialists as well as the diversity of roles to be met. While there has been a great deal of discussion about the use of conversational and pedagogical agents for tutoring, results have been sparse, and largely developed on a one-off basis for each situation. Further examples can be found in the literature (e.g., Chatham, 2007; Nielsen et al, 2006), but not all of them contain a direct pedagogical approach (e.g., Weiland et al, 2003).

Requirements

From an implementation perspective, then, what does this research suggest about requirements for the nature of conversational and pedagogical agents? Agents should be:

- Competently conversant
- Able to both produce and respond to non-verbal cues
- Able to detect and exhibit affective state (Silverman et al, 2006)
- Animated in a manner that is consistent with the interaction scenario
- Believable, motivational, seen as competent, and trustworthy (Baylor, 2001).
- Competent in the area of inquiry and cognizant of, and matched to, the level of competence of the learner (Kim & Baylor, 2006; Kim, 2007)

These requirements are rather formidable, and to date have only been implemented in simulators produced for single-purpose applications, with mixed success. To detail some of the challenges inherent in these requirements, the first three are discussed. The fourth requirement is delineated in the first section of this paper, while aspects of the fifth and sixth requirement are presented throughout the manuscript.

Competently Conversant

This requirement is a rather subjective one from an acceptance perspective, since what is satisfactory for one tolerant end user might be intolerable for another. There are currently a few solutions, all of which are, in one way or another, seemingly flawed.

For an agent to be competently conversant (or appear to be so), it must be able to, at a minimum, respond to questions put to it in a manner that satisfies the expectations of the questioner. At present, there are a few solutions that provide this capability, and they generally fall into one of two categories, either pattern matching or cognitive model-based. From an implementation perspective, the simpler of the two is pattern matching. One of the most common pattern matching engines for conversation is A.L.I.C.E., developed by Richard Wallace. The A.L.I.C.E. standard and most engines are Open Source under GNU Public License, and use AIML for encoding knowledge. A standard AIML set is provided to which additional domain specific knowledge may be added.

Relatively easy-to-use authoring tools exist online for simple applications; however, for more advanced capabilities such as staying on topic or responding with emotionality, additional capabilities must be added to the AIML engine, and these are not straightforward, especially for non-programmers. Furthermore, adding new AIML to the standard AIML set usually causes pattern matching conflicts, which are difficult to find and resolve. For this requirement, a new manner of creating and adding knowledge is needed beyond that which currently exists, especially if novice developers (e.g., educators, physicians, etc.) are to conveniently create new virtual patients with novel symptoms and complaints. More complex cognitive models (e.g., ACT-R) show promise as knowledge engines for conversational agents, however, such tool may be unavailable to casual users due to their complexity of creating and loading models.

Produce and Respond to Non-Verbal Cues

Implementation of this requirement implies detection and categorization of end user gestures. There are currently two general approaches to implementation. One approach is to use body position sensors that are then tracked by cameras and other position receptors. This approach is intrusive and expensive, but very accurate. The more viable approach is using a camera as an input device to computer vision software. To detect the gestures of an end-user, some kind of image analysis is required, generally involving a camera on the user's computer, and real-time image analysis and gesture categorization. These capabilities are still in their infancy, but are on the radar of researchers in the domain.

The second side of this requirement is interpretation of the end user's gestures, which is also a difficult problem, even though a number of applications that involve gesture interpretation and use are already in existence (e.g., Betke, Gips & Fleming, 2002; Bradski, 2002; Licas & Szirany, 2004; Wilson & Cutrell, 2005; Wu, Shah & Lobo, 2000). Few of these examples, however, mention the recognition and interpretation of non-verbal communication cues, which is of primary interest for satisfying this requirement. These cues are readily perceptible to most humans, but often very subtle and difficult to interpret by software. The use of gesture in human-computer interaction clearly has a future; however, it is equally clear that this future has yet to arrive, so this requirement may be difficult to satisfy in the near term.

Able to detect and exhibit affective states

Some systems are currently functioning with some success in both detecting and adapting to changes in emotional/affective state, such as *Eve* in the domain of elementary mathematics tutoring (Sarrafzadeh et. al., 2008), and *Edu-Affe-Mikey* (Alepis et al, 2008) in the domain of medical tutoring. Several other research and development groups are also working towards this goal as well (e.g., D'Mello & Graesser, 2009; Kort, Reilly & Picard, 2001), giving us reasons to believe that this path has reasonable possibilities in producing more realistic and effective conversational and pedagogical agents.

There are a number of approaches to detecting and processing affective states, including facial expression recognition, dialog analysis, and analysis of human psycho-physiological response. For instance, *Easy with Eve* (Sarrafzadeh et al., 2008) adapts to students via a lifelike animated agent called Eve, who is able to detect student emotion through facial expression analysis. Conati (2002) has developed a probabilistic model of determining student affect that has been applied in an educational game for 11 year old mathematics students. The model relies on a Dynamic Decision Network (DDN) to identify student affect. When fully complete, the system will use both facial expression analysis and biometric sensors to feed real-time information into the DDN. Furthermore, Burleson (2006) has developed an *Affective Learning Companion* that is able to mirror several student non-verbal behaviors and detect frustration/help-seeking behavior from students using input from nonverbal sensors. Although the accuracy of this system is at 79%, the result is based on offline computation. Future work is needed to refine the system to accurately sense these frustration/help-seeking behaviors in real time, so that the agent can be continuously aware of the student’s affective state. Finally, other researchers work with emotion-aware tutoring systems where students wear a Bluetooth-enabled sensor ring on their finger, able to transmit heart rate, blood pressure and skin conductance data to a tutoring system to assess the user's boredom and confusion levels (Simonite, 2007). Affective intelligence on the part of conversational pedagogical agents will serve to make them both more believable and attentive to the needs of the student participants. This is especially the case in which the agent is directly simulating a human patient as in the present example.

There are several aspects of conversational agents that have rather critical human-computer interaction and user interface design considerations. One aspect of conversational agents that is often overlooked however, is the interface provided to authors of learning systems. This interface covers not only the authoring of agents themselves, but the simulation that creates the opportunity for learning to occur for students who use the agent, and the environment in which the agent "resides". Since authors may well not be computer scientists, but could be medical faculty members, or others with only a peripheral knowledge of the underlying technology, the interface must provide the authors with the capability of authoring learning scenarios, intelligent agents, and the context or environment in which they operate, without deep knowledge of the underlying technology. Figure 1 shows an idealized authoring task description surround the use of the agent-patient in our current example.

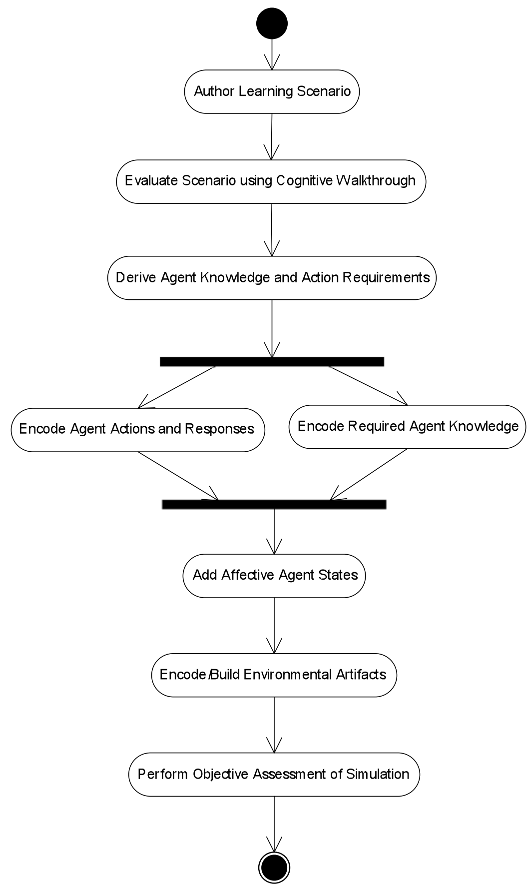

Figure 1. Authoring Task for a Learning Simulation

Notice that there are two opportunities for evaluation during the authoring task, one associated with the learning scenario that can be done subjectively, and one associated with the overall learning simulation, including the patient agent, that is done objectively. Either of these evaluations can be performed by the author, or by an external evaluator if desired. In the former case, the evaluator is walking through the scenario assuming the role of a student, or is employing a student or content expert to walk through the scenario, talking aloud or annotating the scenario with comments and questions. In the objective assessment task, the scenario is executed by the student and objective measures are collected on the effectiveness of the simulation in meeting the learning objectives of the author. Possible measures are pre-test, post-test scores, meeting interim learning objectives, asking the “right” questions of the conversational agent-patient, correctness of the final diagnosis and/or treatment plan and time on task. Finally, another critical assessment to be performed is of the usability of the interface to the authoring tool itself. It is not reasonable to assume that the use of conversational agents such as the agent-patient will ever come into widespread use if their knowledge is not easy to encode, and if simulations using agents are not easy to author.

Overall, implementation of effective conversational and pedagogical agents will be difficult from both human-computer interaction and software engineering perspectives. Results to date, however, are encouraging, and lead one to believe that effective agents are in the realm of possibility. Right now, using pattern matching techniques and AIML, we are able to implement superficially believable conversational agents. The technological details of creating agents that users regularly mistake for humans, is a challenge for future developers.

Conversational Agents in Virtual Worlds: Perspectives from the Learning Technologies

The author writing this section of the paper has a background in Learning Technologies and considers himself to be a “crossover scholar” (Carr-Chellman and Hoadley, 2004), subscribing to ideas from the Learning Sciences, with special emphasis on sociocultural forms of learning in context (e.g., Collins, Brown and Duguid, 1989), and Instructional Design (ID).

An important barrier that needs to be overcome to allow for the effective and engaging implementation of conversational agents in three-dimensional learning and teaching contexts is the tendency to focus on one single aspect of the agent. For instance, conversational pedagogical agents are not simply front-end interfaces to artificial intelligence engines whose sole purpose is to deliver information to learners; they serve socio-cultural purposes as well. While this fact is recognized in the literature, it is seldom accounted for in research practice as we try to vary individual variables to observe their impact and implications. A geriatric agent-patient for example, not only needs to have knowledge about his/her condition to reply to learner questions , but he/she also need to *look* like a geriatric patient (Veletsianos 2007), while, at the same time, avoiding perpetuating stereotypes usually associated with the elderly. As argued by Veletsianos (2007), contextual relevance is a problem in the current literature as wizards and sorcerers have been often been employed to teach such subjects as economics and English. As indicated by Haake (2009), stereotyping is also a problem in the pedagogical agent literature, especially as it relates to gender. What is required therefore is a holistic approach aimed at enhancing agent-learner interaction and the design of conversational agents, by focusing on such diverse issues as agent role, demeanour, pedagogical purpose, appearance, credibility, message composition, and others. One step towards this direction is proposed by Veletsianos, Miller & Doering (2010) who put forth the *Enhancing Agent Learner Interactions* (EnALI) framework. Instead of focusing on a single issue or a single variable, this framework raises numerous interrelated issues that designers need to consider when considering the integration of pedagogical agents in teaching and learning.

A second barrier to the effective use of conversational avatars in virtual worlds stems from the fact that implementations tend to resemble what can be accomplished in others virtual environments and/or in real life. In other words, use of avatars in virtual worlds tends to replicate familiar activities that fail to take into account the possibilities of avatars in virtual worlds. If we return to our geriatric agent example, one can easily see that the familiar activity to implement in this context would be the development of a simulation where the student asks questions and the agent answers those questions while sitting in the doctor’s office. While the immersive aspects of the virtual world may differentiate this experience from an agent simulation that occurs within the confines of a browser, such an implementation is void of the added-value and affordances of the virtual world and the avatar. Affordances are defined by Norman (1988) as “perceived possibilities for action” and represent *suggestions* to the user/observer based on a multiplicity of factors ranging from cultural to social to contextual considerations. More specifically to educational settings, virtual worlds and avatars may offer pedagogical, social, and technological affordances (Kirschner, Strijbos, Kreijns, & Beers, 2004). For example, instead of students being faced with an avatar that they can diagnose via conversation, they could examine the avatar by using virtual world gestures developed for this purpose. The geriatric avatar could then respond to this examination by text, movement, or voice. One can easily see the potential impact of this learning experience if the geriatric avatar yells out in excruciating pain or if its whole body jerks away in response to the student’s choice of gesture. To truly enhance practice we therefore need to consider avatars in virtual worlds as affordances for our specific teaching and learning purposes as opposed to merely using them because they are new and exciting. For example, designers and researchers need to ask themselves what opportunities this medium provides to them. For example, avatars in Second Life can respond to actions, initiate gestures, engage in synchronized behaviour, and simulate flying. Do these actions and activities offer any opportunities for improving teaching and learning? In some situations, it is quite likely that they do. In other situations though, these opportunities may not be applicable in which case the researcher/designer may need to reconsider the reasons of using this medium.

Finally, current conversational agent research seems to follow the same path as research in other emerging technologies used in education. Specifically, due to the emerging nature of the technology, Veletsianos (2010) suggests that research tends to investigate the advantages and disadvantages of the tool, take a case-study approach whose findings apply to limited contexts, and tends to be comparative in nature (often comparing the new technology with a face-to-face equivalent). To truly learn how to implement and integrate conversational agents in virtual worlds, we need ecologically valid (Gulz, 2004) and longitudinal research studies that focus on solving real-world problems in situ. Veletsianos & Miller (2008) note that there is a lack of pedagogical agent research occurring outside of the research lab – their research for example (Veletsianos & Miller 2008; Veletsianos, Scharber, & Doering 2008), highlights the complexity of pedagogical agent implementations in real-world settings and brings to surface issues that are usually ignored when pedagogical agents are investigated in the experimental lab (e.g., how does one consistently motivate learners to continue interacting with conversational avatars over the long-term). Conducting research in real-world contexts is therefore one important step towards the effective integration of conversational pedagogical avatars.

Conversational Agents in Virtual Worlds: Practical Perspectives from Engineering

The author of this section has a background in Electrical Engineering and system integration. From an engineering and software development point of view, many of the barriers to the deployment of conversational agents in a virtual environment for educational purposes are the same as those associated with the implementation of any new technology for education (Kay, 2006). The most significant of these being access to appropriate technical skills, cost of infrastructure to support access for students, and uncertainty concerning the future of the technology.

The degree to which these barriers act as obstacles to development may depend on the level of sophistication of the deployment. A typical implementation of a conversational agent can be broken down into three major components: the conversational/AI engine, an autonomous virtual character avatar (a geriatric patient in our example), and a virtual space within which the avatar operates, such as a hospital, doctor's office, or examining room.

Consider the implementation of the virtual character in more detail. It is important to recognize that there are a number of levels of functionality that can be implemented. It is relatively straightforward to bring an existing conversational agent into a virtual setting such as Second Life, as long as movement, expressions, and automatic detection of other avatars are not required. It is possible to purchase applications, such as the Pandorabot Client Kit, or SAN Chat Bot, which are capable of providing this basic level of functionality for Second Life. These can be used in conjunction with external services such as Pandorabots, which provides a simple to use conversational engine based on AIML (Kerly et al, 2007). In fact, numerous levels of implementation can be identified:

- conversational agent attached to avatar – conversation initiated by user or open chat
- ability to identify users and initiate conversations based on location or observed actions of users (e.g., sitting on a chair) – simple animation
- interaction with scripted objects that exist within the virtual world
- ability to identify users based on visibility and direction in which they are facing (e.g., is there an obstacle, such as a wall, between them?), activity (e.g., standing quietly vs. speaking to someone else), and history (e.g., has this person been asked and already declined to converse?)
- simple movement based on navigation around static paths in the virtual world
- advanced path finding and navigation based on an analysis of surrounding obstacles and other avatars (recognition/identification of objects, such as walls, openings, dimensions of objects, trajectory of moving objects and avatars, knowledge of operation of doors).

Note that this list does not address the cognitive and social behaviours that will need to be implemented in virtual conversational avatars, as these issues have been examined in detail in other sections of this paper.

Although not strictly ordered, in general, the further down the list one goes the more sophisticated the level of implementation and the greater the expertise and associated cost (Liang et al 2008; Stytz & Banks 2003). However, what remains to be determined is the level of functionality actually required to achieve the desired benefits of deploying a conversational agent for educational purposes. Although realistic behaviour has been linked to engagement (Strassner & Langer 2005; Yuan & Chee 2005) the evidence supporting this is far from definitive (see Gulz, 2004 and the cognitive psychology perspective within this paper). In many cases, lack of functionality can be mitigated by greater control of the environment, such as reducing the number of objects, or restricting the number of users who can access the agent at one time. In other words, the degree to which lack of access to expertise is a barrier to deployment depends largely on what level of implementation is required, and we may not have enough information to determine that at this time.

Furthermore, the level of functionality may vary with application and the type of user that is being targeted. Whereas it may be appropriate for a patient simulation, expected to engage a post-secondary student, to mimic human appearance and behaviour as closely as possible, an iconized figure may be sufficient (Gulz & Haake, 2006; Haake & Gulz, 2009). Similarly the measures of success will vary according to the specific goals of each application and some of these are identified in the final discussion of this paper.

Uncertainty also exists with regards to the future of virtual worlds: Which ones will exist in the long term? For example, while Second Life has existed for 6 years, Google Lively was discontinued 4 months after its release. Additionally, there’s the question of which virtual worlds are most popular and which ones only attract a small user base. KZero, a market research company that specializes in virtual environments, has data on over 90 virtual worlds as of Q4 2008 (Warburton, 2009). These considerations may lead designers and researchers to wait to see how this technology matures and which platforms thrive before making a significant investment. Until a standard platform, and associated Application Programming Interface[1] emerges, there is a clear risk associated with developing conversational pedagogical avatars for one platform when the ability to port them to another is not present.

Finally, there remains the issue of developing the knowledge base or content associated with the conversational component of the avatar. Although this issue is not specific to virtual worlds, the success of conversational avatars in any environment is also dependent on how well the avatar is able to interact and convey useful information through conversation, i.e. to be “competently conversant” as described earlier. Regardless of the technical challenges associated with automating an avatar in a virtual world, the task of achieving conversational competency represents a significant portion of the work. In our work with historical figure actor agents (Authors, submitted) and virtual patient agents, it’s estimated that as much as 90 per cent of the development time has gone to the conversational component of our virtual characters. The student's trust in, and thereby, continued engagement with, the process is directly affected by the perceived intelligence of the agent. The level of conversational competency is one of the most important contributors in the perception of intelligence (Norman, 1997). Additionally, a strong conversational component can assist in defining the personality of a virtual character (Gulz, 2004) and possibly making up for shortcomings in the avatar's animation or ability to navigate

[1] Application Programming Interface (API) refers to procedures and functions that allow communications between multiple applications/programs. In our context, it would be processes that allow portability of conversational agents between virtual worlds

freely. The complexity of course arises when we have to consider the limits of intellect and conversational ability embedded within the conversational avatar and the expectations that are developed as a result of the avatar's perceived intelligence.

As with learning objects, (Kay 2007) the development of the conversational component would benefit greatly from the ability to standardize the associated conversation-specific data, preferably independent of any particular platform, allowing for the sharing of conversational content. Taking the geriatric patient example, attempts to standardize information relevant to virtual patients exist (Ellaway et al, 2006), but this does not currently extend to conversational information, such as the types of questions a doctor may ask a patient in trying to determine a diagnosis. Our own work with historical figure actor agents (Authors, submitted) and virtual patients has benefited from the adoption of this concept and the development of such standards is underway.

Summary and Concluding Thoughts

In this paper, we provided a space where four researchers/practitioners from four diverse but converging fields to come together to discuss the barriers hindering the effective deployment of conversational avatars in virtual worlds for educational purposes. While it is obvious that each field has different foci points, it is also evident that there exist coordinates of convergence. These points of convergence identify the ways in which all disciplines can be informed and benefit from each other, and are identified below:

- Agent-Learner Interaction: All four disciplines highlighted the fact that researchers and practitioners need to pay attention to the *interaction* between agents and learners. The suggestion to enhance agent-learner interactions arises at many levels and seems to be at the forefront of what needs to be attained to push the field forward. Authors have identified multiple levels at which agent-level interactions can be enhanced ranging from ensuring naturalness while avoiding cognitive overload, to enhancing the agent's conversational capability, to enabling the agent to respond to both verbal and non-verbal cues.

- Intertwined Theory, Practice, and Research: It is obvious from reading the above that the conversational avatar field represents a unique combination of theory and research intertwined with practice. Importantly, practical considerations drive research (and at times theory) and theoretical and research issues drive practice. In addition, it is easy for researchers to find themselves embroiled in discussions regarding the perceived superiority of research approaches and methodologies. For these reasons, we believe that the field will benefit from approaching research on this topic under a Design Based Research (DBR) umbrella. Such an approach is iterative, flexible, and multi-paradigmatic and allows for the simultaneous design, development, and implementation of interventions in real-world contexts (The Design-Based Research Collective, 2003). The ideas outlined in this paper lend themselves well to DBR because modifications to the intervention are easy to implement without disrupting current practice (e.g., a conversational avatar's interaction patterns can be modified without disrupting existing practice). Additionally, as this paper has demonstrated, the close relationship between practice and research in this field, necessitates a collaboration between practitioners and researchers, and we believe that the DBR approach is the best option for supporting these collaborations.
- Related research and current opportunities: The discussion above highlights the fact that the field can learn a lot from prior research that has occurred in related contexts (e.g.,

APAs). Nevertheless, current practice offers further ground for theorizing and development. For instance, conversational avatars may have the ability to identify learners based on the direction in which they are facing, thus avoiding engaging in conversation with them in inappropriate times. Such behaviour may enhance the user's experience in the learning environment, ultimately leading to improved performance and satisfaction.

- Evaluation and Assessment: While varied disciplines may seek similar goals and outcomes, the way that conversational avatars are assessed and evaluated across disciplines varies. According to Dehn & van Mulken (2000), measures of success could include subjective self reports following the interaction, objective behavioural measures taken during the interaction, or performance measures of competency taken after the interaction. In addition the conversational record is another source of data that could be mined for measures relevant to the learning objectives. The above discussion clearly supports a multitude of measures although in each case, the measures generally reflect the disciplinary objectives and focus. Thus, it is important to identify the critical measures of success relevant to the learning objectives so that conversational avatars are properly evaluated and assessed ideally within a DBR approach.

Teaching and learning with conversational avatars is a multidisciplinary endeavour with a myriad of concepts and methods. Nevertheless, there is strong convergence on improving the quality of learner–agent interactions as the primary consideration. To meet the objectives of smooth, effective, and natural interaction will require guidance from empirical evidence and design knowledge from practitioners from fields as diverse as psychology, linguistics, sociology, graphic design, psychology, physics, and education. Only by capitalizing on the expertise of different disciplines will we be able to fully and effectively utilize pedagogical avatars in education.

References

Alepis, E., Virvou, M. and Kabassi, K (2008) Requirements Analysis and Design of an Affective Bi-Modal Intelligent Tutoring System: The Case of Keyboard and Microphone, *Studies in Computational Intelligence* (SCI) *104*, 9–24.

Baylor, A. L. (2001). Permutations of control: Cognitive considerations for agent-based learning environments. *Journal of Interactive Learning Research. 12*(4), 403-425.

Betke, M., Gips, J. and Fleming, P. (2002) The Camera Mouse: visual tracking of body features to provide computer access for people with severe disabilities, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *10*, 1-10.

Bradski, G. R. and Davis, J.W. (2002) Motion segmentation and pose recognition with motion history gradients, *Machine Vision and Applications*, 13, 174-184.

Burleson, W. (2006, September) "Affective Learning Companions: Strategies for Empathetic Agents with Real-Time Multimodal Affective Sensing to Foster Meta-Cognitive and Meta-Affective Approaches to Learning, Motivation, and Perseverance," MIT PhD Thesis.

Brown, J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32-41.

Carr-Chellman, A., & Hoadley, C. (2004). Learning sciences and instructional systems: Beginning the dialogue. *Educational Technology*, *44*(3), 5-6.

Cassell, J., & Thórisson, K.R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13, 519-538

Conati, C. (2002) Probabilistic Assessment of User's Emotions in Educational Games. *Applied Artificial Intelligence,16 (7-8*), 555-575.

Craig, S.D., Gholson, B., & Driscoll, D.M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. Journal *of Educational Psychology*, 94(2), 428-434.

Dehn, D.M., & van Mulken, S. (2000). The impact of animated interface agents: A review of the empirical literature. *International Journal of Human-Computer Studies*, *52*, 1-22.

D'Mello, S. and Graesser, A. (2009) Automatic Detection of Learner's Affect From Gross Body Language. *Applied Artificial Intelligence*, *23*(*2*), 123-150.

Dirkin, K.H., Mishra, P., & Altermatt, E. (2005). All or nothing: Levels of sociability of a pedagogical software agent and its impact on student perceptions and learning. *Journal of Educational Multimedia and Hypermedia*, *14(2),* 113-128.

Friedman, D., Steed, A., & Slater, M. (2007). Spatial social behavior in Second Life. *Lecture Notes in Computer Science*, *4722*, 252-263.

Gulz, A. (2004). Benefits of virtual characters in computer based learning environments: Claims and evidence. *International Journal of Artificial Intelligence in Education, 14*, 313-334.

Gulz , A., & Haake, M. (2006). Design of Animated pedagogical agents: A look at their look. *International Journal of Human-Computer Studies, 64(4),* 322-339.

Haake, M. (2009). Embodied Pedagogical Agents: From Visual Impact to Pedagogical Implications. Unpublished doctoral thesis, Lund University, Lund, Sweden.

Haake, M., & Gulz, A. (2009). Visual stereotypes and virtual pedagogical agents. *Educational Technology & Society, 11*(4), 1-15.

Heller, R., Dobbs, B. & Strain, L. (2009). Video programming for individuals with Dementia: Assessing cognitive congruence. American Journal of Alzheimer's Disease and Other Dementias, 24(2), 122-128.

Heller, R.B. & Procter, M. (2009). Animated pedagogical agents: The effect of visual information on a historical figure application. *International Journal of Web-based Learning and Teaching Technologies, 4(1)*, 54-65.

Kay, R. (2006). A Systematic Evaluation of Learning Objects for Secondary School Students. *Journal of Educational Technology Systems*, *35*(4), 411-448.

Kerly, A., Hall, P., & Bull, S. (2007). Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems*, *20*(2), 177-185.

Kim, Y. (2007). Desirable characteristics of learning companions. *International Journal of Artificial Intelligence in Education, 17*(4), 371-388

Kim, Y., & Baylor, A. L. (2006). Pedagogical agents as learning companions: The role of agent competency and type of interaction. *Educational Technology Research and Development, 53*(3), 223-243.

Kirschner, P., Strijbos, J., Kreijns, K., & Beers, P. J. (2004). Designing electronic collaborative learning environments. *Educational Technology Research and Development, 52*(3), 47-66.

Kort, B., Reilly, R., and Picard, R.W. (2001) An Affective Model of Interplay between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion, *Proceedings of the Second IEEE International Conference on Advanced Learning Technologies* (ICALT'01), 43-46.

Licsar A. and Sziranyi, T. (2004) Hand Gesture Recognition in Camera-Projector System, in N. Sebe et al. (Eds.): HCI/ECCV 2004, LNCS 3058, pp. 83-93.

Mayer, R.E., Sobko, K., & Mautone, P.D. (2003). Social cues in multimedia learning: Role of Speaker's Voice. *Journal of Educational Psychology*, 95(2), 419-425.

Mazar, R., & Nolan, J. (2008). Hacking say and reviving ELIZA: Lessons from virtual environments. *Innovate* 5 (2). Retrieved on May 12, 2009 from http://www.innovateonline.info/index.php?view=article&id=547

MedBiquitous Virtual Patient Architecture - Virtual Patient Working Group - Confluence. (n.d.). . Retrieved May 28, 2009, from http://groups.medbiq.org/medbiq/display/VPWG/MedBiquitous+Virtual+Patient+Architecture.

Moreno, R., & Mayer, R.E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media.*Journal of Educational Psychology*, 94(3), 598-610.

Norman, D. (1997). How might people interact with agents. In J. M. Bradshaw (Ed.), *Software agents* (pp. 49-56). Menlo Park, CA: MIT Press.

Norman, D. (1988). *The psychology of everyday things*. New York: Basic Books.

Ortony, A., Clore, G. L., and Collins, A. (1988) *The Cognitive Structure of Emotions*. Cambridge University Press.

Ronghua Liang, Mingmin Zhang, Zhen Liu, & Krokos, M. (2008). Cognitive and synthetic behavior of avatars in intelligent virtual environments. *Virtual Reality*, *12*(1), 47-54.

Rumelhart, D.E. & McClelland, J.L. (1986).Parallel distributed processing: Explorations in the microstructure of cognition, MIT Press Cambridge MA.

Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C. and Bigdeli, A. (2008) “How do you know that I don't understand?" A look at the future of intelligent tutoring systems. *Computers in Human Behavior*, *24* (*4*), pp. 1342-1363.

Silverman, B.G., Johns, M., Cornwell, J., and O’Brien, K. (2006) Human Behavior Models for Agents in Simulators and Games: Part I: Enabling Science with PMFserv. *Presence: Teleoperators & Virtual Environments*, 15(2), 139–162.

Simonite, T. (2007) Emotion-aware teaching software tracks student attention. New Scientist, Retrieved on December 17, 2008 from http://www.newscientist.com/article/dn10894-emotionaware-teaching-software-tracks-student-attention.html.

Strassner, J., & Langer, M. (2005). Virtual humans with personalized perception and dynamic levels of knowledge. *Computer Animation & Virtual Worlds*, *16*(3/4), 331-342.

Stytz, M. R., & Banks, S. B. (2003). Progress and Prospects for the Development of Computer-Generated Actors for Military Simulation: Part 1—Introduction and Background. *Presence: Teleoperators & Virtual Environments*, *12*(3), 311-325.

The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Vegso, J. (2005), Interest in CS as a Major Drops Wilson, A.D. and Cutrell, E. (2005) FlowMouse: A Computer Vision-Based Pointing and Gesture Input Device, In M. F. Costabile and F. Paterno (Eds.) *Proceedings of the International Conference on Human-Computer Interaction (INTERACT 2005)*, Rome, Italy: Springer-Verlag Berlin Heidelberg, *3585*, 565-578.

Veletsianos, G. (2010). A Definition of Emerging Technologies for Education. In G. Veletsianos (Ed.), *Designing, Delivering, Developing, and Evaluating Emerging Technologies for Education*. Athabasca University Press.

Veletsianos, G., Miller, C., & Doering, A. (2010). EnALI: A Research and Design Framework for Virtual Characters and Pedagogical Agents. *Journal of Educational Computing Research, 41*(2), 171-194.

Veletsianos, G. (2007). Cognitive and Affective Benefits of an Animated Pedagogical Agent: Considering Contextual Relevance and Aesthetics. *Journal of Educational Computing Research, 36*(4), 373-377.

Veletsianos, G. (2009). The impact and implications of virtual character expressiveness on learning and agent-learner interactions. *Journal of Computer Assisted Learning, 25*(4), 345-357.

Veletsianos, G., Scharber, C., & Doering, A. (2008). When Sex, Drugs, and Violence Enter the Classroom: Conversations between Adolescent Social Studies Students and a Female Pedagogical Agent. *Interacting with Computers, 20*(3), 292-301.

Veletsianos, G., & Miller, C. (2008). Conversing with Pedagogical Agents: A Phenomenological Exploration of Interacting with Digital Entities. *British Journal of Educational Technology, 39*(6), 969-986.

Waller P.A. (2007) Extended television: a study of how investigations of use can inform design processes in nursing homes. *PsychNology Journal, 5*, 105-132.

Wallis, P. & Norling, N. (2003). The trouble with chatbots: Social skills in a social world.

Warburton, S. (2009). Second Life in higher education: Assessing the potential for and the barriers to deploying virtual worlds in learning and teaching. *British Journal of Educational Technology*, *40*(3), 414-426.

Wu, A., Shah, M., and Lobo, N. d. V. (2000) A Virtual 3D Blackboard: 3D Finger Tracking Using a Single Camera, in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*: IEEE Computer Society, p. 536-543

Xiang Yuan, & Yam San Chee. (2005). Design and evaluation of Elva: an embodied tour guide in an interactive virtual art gallery. *Computer Animation & Virtual Worlds*, *16*(2), 109-119.