Supporting Technology Enabled Learning with Artificial Intelligence and Cognitive Modelling

Dr. Bruno Emond
National Research Council Canada
1200 Montreal Road, Ottawa, ON. K1A 0R6
CANADA

Bruno.Emond@nrc-cnrc.gc.ca

Dr. Jerzy Jarmasz
Defence Research and Development Canada
1133 Sheppard Ave. W., Toronto, ON, M3K 2C9
CANADA

Jerzy.Jarmasz@drdc-rddc.gc.ca

ABSTRACT

Emerging technological trends such as cloud computing, artificial intelligence, data analytics, virtual and augmented reality, and advanced speech recognition technology will play a role in defining modelling and simulations (M&S) and the future training systems. Agent-based modelling and simulation is an important element for the development of the next generation of training simulators where human communication and interaction demand high cognitive fidelity and natural language processing capabilities. The challenges faced in this area include the need to conduct more effective simulation-based training in team tasks, reduce reliance on live training, as well as increasing the readiness of individuals to participate in team training events, thereby making these events more effective and efficient. The development of agent-based modelling and simulation for training in the form of adaptive training systems offer a valuable approach to support interoperability and standardization, as well as a means to measure learning and training effectiveness. The current paper presents a set of projects for developing adaptive training simulations using synthetic teammates for the Royal Canadian Navy. Each project used an agile approach to system development for teamwork training, specifically using iterative, requirements-guided approaches to developing synthetic agents to play well-defined team roles, which resulted in "shallow agent" modelling. The projects are described in terms of their intended functionalities, and the means to assess if the goals were reached.

1.0 INTRODUCTION

Emerging technological trends (Cearley, Walker, & Burke, 2017) such as cloud computing (NATO-MSG-131, 2015), artificial intelligence, data analytics, virtual and augmented reality, and advanced speech recognition technology will play a role in defining modelling and simulations (M&S) and the future training systems. Agent-based modelling and simulation (Macal & North, 2010) is an important element for the development of the next generation of training simulators where human communication and interaction demand high cognitive fidelity, (Liu, Macchiarella, & Vincenzi, 2009), including natural language processing capabilities , which must be measured not only by the avatars' physical appearance but also by their psychological and cognitive realism from a trainee's point of view. The integration of synthetic agents into training simulations touches on a number of technological components including speech recognition, dialog management, agents design and implementation, and multimodal interfaces. All of these technological components, just like any other training intervention, require an evaluation of its training effectiveness (Boldovici, Bessemer, & Bolton, 2002; Carroll, Thompson, & Deaton, 2008).

However, from the perspective of simulation-based training systems, the rapid pace of technological innovations is often causing a shift from issues related to human performance and skill acquisition towards technology and device fidelity (Hodges, 2014). This shift away from human factors is compounded by demands to achieve more rapid development and deployment of training solutions, which means organizations no longer have the luxury of taking months to conduct a detailed needs assessment (Bell, Tannenbaum, Ford, Noe, & Kraiger, 2017). These pressures (strong emphasis on technological fidelity, and less detailed training needs analysis) are exacerbating the challenges that training practitioners already face in assessing training simulations utility at the end of the development process, and make more urgent the need for agile and ongoing utility assessment beyond traditional learning transfer studies (see an extensive discussion in Boldovici et al., 2002). Training professionals are interested in evidence-based interventions that have been shown to positively influence transfer (Alexander, Brunyé, Sidman, & Weil, 2005). Also, teamwork training often does not need highly technical simulations, but rather brings out training design considerations instead (Beaubien & Baker, 2004). Although we have evidence-based findings, one of the predominant limitations of the research literature is that studies stop at the point of identifying, describing, or measuring factors that may influence transfer without investigating how those factors might be effectively changed or managed within the workplace or training context to enhance transfer (Ford, Baldwin, & Prasad, 2018), which points towards an agile, customer-based research agenda to better support training design (Baldwin, Ford, & Blume, 2017; Granek, Jarmasz, Boland, Guest, & Bailey, 2016).

The current paper presents a set of projects for developing adaptive training simulations using synthetic teammates. The development of agent-based modelling and simulation for training in the form of adaptive training systems offer a valuable approach to support interoperability and standardization (Brawner, Goodwin, & Sottilare, 2016), as well as a means measuring learning and training effectiveness (Brawner, Sinatra, & Sottilare, 2017). Each project followed an agile development methodology to iteratively refine training goals, and make technology design and implementation decisions. Agile software development promises benefits such as on-time delivery and customer satisfaction, thus aiming to deliver business value in short iterations. In addition, effective and efficient simulation-based training systems require adequate verification, validation and accreditation from human and system performance data. Some of the projects exemplify different means to exploit artificial intelligence and data analytics methods to support claims about the acceptability of intended application functionality.

The projects' efforts to provide an evidence-based approach to assess the feasibility of training with synthetic teammates, could be portrayed as an example of the Generic Methodology for Verification and Validation (GM-VV) (NATO/STO-TR-MSG-073, 2015; SISO, 2013). GM-VV offers a framework to tailor verification and validation efforts either as either post-hoc, concurrent, iterative, or recursive (SISO, 2013), and shares attributes of agile methodologies such as a strong focus on user needs, and design implementation decisions being supported on an evidence-based process to generate arguments for M&S features and/or results acceptance goals and criteria. The GM-VV framework is rooted in both goal-oriented requirements engineering and claim-argument-evidence safety engineering principles. Under GM-VV, the quality of the V&V efforts will establish the value of acceptance recommendations, which requires the development of a structured approach where the reasoning about the supporting arguments is traceable, reproducible, and explicit (SISO, 2013).

In the case of synthetic teammates though, the range of supporting evidence for simulated human communication can be very difficult to obtain, and could amount for a system to the equivalent of passing a

Turing test (Turing, 1950). However, when dealing with task-oriented artificial intelligence as opposed to general artificial intelligence, other methods of evaluation are possible including human discrimination, problem benchmarks, and peer confrontation (Hernández-Orallo, 2017). The methods rely respectively on human subjective judgments, performance on data processing, and AI agents competing against one another. In spite of the relative objectivity of the latter two methods, it is often not the case that a data set is available to tell apart alternative synthetic teammates options, or that a competition can be set to determine which agents is the winner. At least at the prototyping phase of synthetic teammates development, human discrimination can be used to focus on the believability of artificial agents as one of the validation measures (Togelius, Yannakakis, Karakovskiy, & Shaker, 2012). To complement the subjective validation measures, test-driven software development for software function verification can be applied.

The rest of the paper is divided in two main sections. The first section briefly defines the training problems to be addressed using synthetic teammates, namely the need to conduct more effective simulation-based training in team tasks while reducing reliance on live training, as well as for individuals to increase their live team training readiness so that live team training could be more effective and efficient. The second section describes three projects that have used synthetic teammates for training: a) landing signal officers, b) change of direction and conning orders for the Officer of the watch, and c) tactical voice procedures by Naval Combat Information Operators (NCIOP). Each of the projects is described in terms of its intended functionalities (goals), and the means to assess if the goals were reached (evidences). Finally, the last section summarizes the main elements of the paper, and identifies areas for future research.

2.0 TEAM TRAINING READINESS AND VERBAL COMMUNICATION

Preparing individuals for participating in team and collective training and rehearsal events is a stated goal of military training systems, as expressed in documents such as NATO's Education and Training Directive (NATO-Bi-SC-75-2, 2013). This is often challenging, in particular when a team is composed of individuals having different duties. In addition to developing proficiency on their individual roles and skills, trainees must learn and develop some degree of proficiency in teamwork-related skills before they can effectively participate in collective exercises, where the objective is typically to rehearse team skills under stressful conditions (Wilson, Salas, Priest, & Andrews, 2007). Researchers in other training domains stress the importance in allowing trainees to overtrain team-related skills before they participate in "full mission" team training events (Beaubien & Baker, 2004). At the same time, it is crucial for militaries to make most effective use of live collective training events, as these are typically costly in personnel time (in particular for exercise support staff and role players) and consumables (ammunition, fuel), and impose a burden on the operational vehicle fleet and the availability of personnel for other duties. Thus, the use of simulations to hone trainees' teamwork skills, and not just their individual duties, before they participate in collective training events is an attractive proposition (Grant, 1999). One of the challenges for simulation-based training for team skills is the fact that natural language communication is a key element in operational team coordination. The main problem to be solved is therefore how to provide initial team communication training for novices without the cost of requiring humans to play teammate roles.

However, focusing solely on the communication skills is not enough. For one thing, effective teamwork also involves developing shared attitudes and objectives, and effective anticipation and coordination of teammates' behaviours and information needs (Beaubien & Baker, 2004; Wilson et al., 2007). More importantly for our discussion, communication in military missions always happens concurrently with other tasks (e.g., monitoring or manipulating equipment), and it is important to prepare individuals in teams for this type of pervasive "multitasking" (Beaubien & Baker, 2004). Furthermore, a key aspect of teamwork seems to be communicating about information provided to the team by their situation awareness systems: empirical studies show that teams building up a common operating picture communicate more about information available on their common systems rather than about information gaps (Durlach, Bowens, Neumann, & Carnahan, 2004), and even when teams have a common operating picture, the combination of verbal reports with tracked sensor data results in a better situation awareness outcome than what it would have been if the data sources are taken individually (Grueneberg et al., 2013). Also, training on the technical use of a combat management interface, as well as on message formats are not sufficient to build expertise, because the ability to understand the battlefield situation and integrate critical elements of information is also required (Dyer, Vaughn, & Blankenbeckler, 2004). Thus, it is crucial to provide personnel opportunities to train not only on technical and teamwork skills separately, but also to practise performing these two types of skills together, before engaging them in large team training events.

3.0 SYNTHETIC TEAMMATES FOR TRAINING

Efforts to substitute human teammates by synthetic teammates in training has been pursued from different perspectives including team cognition (Cooke, Gorman, Myers, & Duran, 2013), cognitive modelling (Ball et al., 2010; Demir et al., 2015), human behaviour representation (Gunzelmann, Gaughan, Alexander, & Tremori, 2014), and conversational agents and chatbots (Berg, 2014; Grant, 1999; Jurafsky & Martin, 2017). Synthetic teammates, as human imitation relies on four characteristics: 1) the use of natural language to listen and speak, 2) a mixed-initiative condition, where both humans and artificial agents can influence the course of the dialogue in comparison to a system-initiative dialogue where only artificial agents control the flow, or a user-initiative system, where the user makes all the requests and the system only answers, 3) flexibility, where all dialogue partners can express information in a variation of ways, and 4) adaptation, where partners in the dialogue cooperate and give feedback to help resolve ambiguities, misunderstanding and confirm what was said (Berg, 2014).

Even though all of these characteristics are essential to simulate humans in open dialogs, the issue of how much fidelity is required to simulate teammates depends highly on the team operation context. For teamwork training, as noted above, methodologies that prioritize instructional goals/design and interaction with the user may be more important/relevant than focus on technologies. In addition, very structured operational context where verbal communication is expected to follow a well-defined format will require less adaptation than an operational situation where verbal communication is to support problem solving. An early study on synthetic teammates for training by Grant (Grant, 1999) showed that even with very limited linguistic "skills," an agent-based system could provide adequate team training so long as the scenario was well constrained and there were also opportunities for trainees to interact with the agents through non-linguistic means (e.g., changing the states of a vehicle). In this respect, synthetic teammates could be "shallow entities" in the sense that they only appear to have human-like properties without having much general intelligence. This distinction between complex and simple agent is reflected in agent-based modelling, where one can find either a large number of relatively simple and highly interactive agents; or a smaller number of agents with more complex internal structures (Orr, Lebiere, Stocco, Pirolli, & Kennedy, 2018). Certainly, natural language comprehension and production seem to require agents with complex internal structures, but some acceptable fidelity could be obtained with simple agents. For example, shallow synthetic teammates can provide synthetic teammates support by mapping learner speech input to change in environment conditions or to performance assessments without the need to specify knowledge representation in synthetic teammates. Communication from shallow synthetic teammates is also a simple mapping from text strings to sound using text-to-speech technology. Shallow synthetic teammates are rigid in their behaviour and have little internal autonomy, even though they can display a range of conversation options and behaviour adaptation.

Another factor which determines the level of complexity required for synthetic teammates is the sequence of training scenario complexity. An empirical study of marine docking operations by Hjelmervik, et al. suggests that even though the technology and computational power provide for new possibilities in training simulators with high fidelity, new features that make the tasks more complicated should not be included too early in the training without sufficient investigation. Their results indicate that increasing functional fidelity of the simulation during training improved the performance of participants on later complex tasks, as compared to those training with the highest fidelity from the beginning (Hjelmervik, Nazir, & Myhrvold, 2018).

3.1 Landing Helicopters on Ship Decks

An approach combining human behaviour representation with relatively "shallow" synthetic agents was developed and used to train landing signal officers (LSO) to conduct landing manoeuvres for helicopters on ship landing decks at sea for the Canadian Armed Forces (Cain, Magee, & Belyavin, 2011). Agents playing the roles of the helicopter pilot and ship command and control were integrated into a virtual environment, and communicated with the trainee via synthetic speech recognition and production. The "shallow" agent approach was suitable because the range of actions and communications between the various roles in the helicopter landing task were well constrained; using a Human Behaviour Representation (HBR) approach limited to representing only the behaviours required for the task and representable in the virtual environment enabled the agents to be implemented relatively easily and quickly, and to produce credible interactions with the trainee. Cain et al. evaluated the training environment thus designed using a reverse-transfer-of-training methodology whereby the performance of expert LSOs on the landing task in the training system was compared to that of novice LSOs. The rationale behind this approach is that, assuming the training environment captures the task characteristics adequately, experts should achieve proficient performance levels much more quickly than novices, who should exhibit a more prolonged learning curve as they learn the task. Cain et al.'s data showed that expert LSOs did achieve proficient performance much more quickly than novices, thereby providing an empirical demonstration of the viability of using relatively "shallow" synthetic teammates developed with rapid HBR techniques for training team skills. Given the success of this approach in the LSO context, it is now being investigated to provide training for other military personnel who require role players to practise their team-related skills before engaging with their team, for instance, commanders of Canadian Army combat vehicle crews (Dubreuil, 2016).

3.2 Learning Ship Manoeuvres and Conning Orders

The RCN-ULEARN project (Emond et al., 2016) aimed to deliver to the Royal Canadian Navy (RCN) a proof of concept of a ship's bridge ubiquitous learning application allowing for part-task training support using serious game concepts, and learning analytics to foster self-directed learning in trainees. During the first phase of the project, the focus was on the implementation of a relatively complex scenario, whereby an advanced trainee would manoeuvre his or her ship from astern to the starboard quarter of a guide ship, as an early test of the feasibility of using speech to control a simulated ship and to interact with simulated agents on the bridge. The scenario required communication between the officer of the watch (trainee) and six synthetic agents (guide ship, naval communicator, range finder, relative velocity (RelVel), helmsman, and commanding officer). This is an example of using synthetic teammates to learn communication skills that would normally require the use of expensive equipment (either a real ship or an expensive naval bridge simulator) and the sourcing of numerous role players. The officer of the watch (trainee) was located on the bridge, with a view on the sea and the states of a gyrocompass. The second phase efforts aimed at addressing the other end of the spectrum, that is, learning to execute simple change of course tasks using basic conning orders, with no prior knowledge of the officer of the watch role. The goal was to determine if trainees' performance data exhibited learning patterns and if execution strategies could be identified, which could inform future learning design.

3.2.1 Assessing the usability of speech recognition

The acceptability of the speech interface (recognition and synthesis) was assessed using subjective opinions from subject matter experts (SMEs) during software demonstrations and speech error metrics during software development. The system was qualified as acceptable if SMEs, upon using the system or observing a demonstration, judged the speech interactions on the bridge to have met the required level of realism. A dozen demonstrations were given to SMEs, who deemed the level of realism in the training scenario to be adequate. In addition to the subjective assessment of experts during demonstrations, speech recognition experience metrics were applied during the iterative process of grammar specification and software development to assess the quality of the speech recognition module. The grammar development process was iterative and aimed at extracting task relevant knowledge form SMEs. Five error measurements were taken by the software developers while SMEs were performing the manoeuvring task in order to refine the speech recognition grammar. These measures were: 1) concept-error, when a trainee uses phraseology that is in the system and in scope (i.e, relevant for the task context), but the utterance is not recognized by the system; 2) fault-tolerant, when a trainee uses phraseology that is in the system and in scope and a concept error occurs, then the trainee can correct the concept error with a single repeat of the original phraseology, the second utterance is classified as a concept-pass; 3) input-error, when a trainee uses phraseology that is in the system and in scope, but the utterance is not recognized due to external factors that impact the quality of the signal input such as microphone set up, background noise, or partial press-to-talk interactions; 4) validation-error, when a trainee uses phraseology that is not in the system but should be in scope (undocumented valid requirement); 5) training-error, when a trainee uses phraseology that is not in the system and is not in scope. The data collected during speech recognition grammar development led to a reduction of validation-errors, which is anticipated given the iterative development process, and a desired system property (all relevant speech element are in the system). In addition to the development of the grammar in collaboration with SMEs, a usability study with a subject matter expert not familiar with the system indicated no training-error, but repeated input-errors related to the use of press-to-talk procedures, which validate the need to prepare a trainee well to some of the technical factors involved in speech recognition technology or remove the necessity for the press-to talk button which brings technical challenges (e.g. increased computing power, increased input errors). Overall the utterances required by the trainee (officer of the watch) were mostly short and fall-back strategies were easily available in the form of alternative ways to say something. However, during the plan-briefing task, the officer of the watch must give a relatively long report with many possibilities for performance and speech recognition errors. The solution developed involves supporting the learning with a template containing blanks for the trainee to fill in. Moreover, the trainee could access a manoeuvring board to help them identify the missing values for the plan form. The board was based on the initial and final station coordinates, and the headings of the guide and own ships. After completing the template, the trainee would simply read it out load for the speech recognition module.

3.2.2 Using cognitive models to understand novice performances on conning tasks

One of the objectives of the RCN-ULEARN project was to develop a simple game for supporting basic conning acquisition, collect data on human performance and learning, and identify some cognitive processes in issuing conning orders. The game had five levels of increasing difficulty, each with trials following a particular scenario. Once a learner had successfully completed a certain number of trials in succession, the learner advanced to the next, more difficult level. If the learner failed to complete a trial, the virtual captain provided corrective feedback. Feedback was also provided upon successful trial completion. A data collection protocol was applied to determine the learnability of conning orders. Twelve novice human subjects participated in the study. A learning session was limited to 30 minutes. Only some learners had enough time to progress through all levels. A scenario was comprised of a series of trials of equivalent difficulty. Each trial began with a text-to-speech command from the virtual captain. The learner then responded by speaking one (scenario 1) or several conning orders (other scenarios) to have the captain's command executed by the virtual crew. This then generated the appropriate visual and auditory feedback. The first scenario required learners to issue one of three conning orders in response to the captain's command. Learners were first instructed on the three order types: port fifteen turned the ship left, starboard fifteen turned it right, and midships straightened the ship's head. The sequence of turning trials was randomized. However, after every order to turn the ship, the next trial required the learner to bring the ship out of the turn with a midship order. Consequently, learners gave twice as many midships orders than either turn order.

As a means to determine the cognitive behaviour of novices performing conning tasks, a set of ACT-R cognitive models (Anderson, 2007; Anderson et al., 2004) were written to model performance data. This cognitive modelling method consists of using artificial intelligent agents to simulate human performance on given tasks, with different model parameters leading to different performance profiles. The cognitive models that best fit the performance data (task execution time) from twelve novices processed the captain's command utterance to the end before triggering the relevant conning order, even when a relevant order could be retrieved based on part of the command only, suggesting novices follow the same strategy, and that retrieval of conning orders requires a mnemonic strategy, at least for novices (Emond & Vinson, 2017).

3.3 Learning Tactical Voice Procedures

The RCN-ASPO project was particularly interested basic and highly perishable skills, such as tactical procedure verbal block reports. As discussed above, these communications skills need to be performed under stress and while operating other equipment, and are thus of the type identified by Beaubien and Baker (2004) as likely requiring overtraining prior to collective training, which would be particularly suited to training with synthetic agents in a virtual environment. One of the project objectives was to be able to enable measures and assessments of continuous multi-sentence speech productions, which was identified as an issue during the RCN-ULEARN project. To this end, the project developed a prototype synthetic training environment that allows Naval Combat Information Operators (NCIOP) to learn to effectively deliver block reports. Synthetic teammates were developed to provide the NCIOP trainee with requisite input and responses, and a scenario manager was developed to provide instructional feedback. This allows the NCIOP to train without involving live instructors or role players. The objective of the project was to develop a prototype training system using the Anti-Submarine Plotting Operator (ASPO) as a use case, which could eventually be integrated with the RCN's concept for a Multi-role reconfigurable trainer (MRT; described in (Hazen & Gillis, 2017)). Like the RCN-ULEARN project, the prototype application uses speech recognition and speech synthesis for simulating crew members in the operations room, as well as a simple Combat Management System (CMS) to support submarine marking and tracking tasks. The prototype training system operates with a scenario and assessment manager in the role of an Electronic Coaching and Evaluation System, identified as an essential component of the RCN's MRT concept (Hazen & Gillis, 2017), and made use of Commercial Off-The-Shelf (COTS) and open-source software to provide the RCN with a base to engage in further in-house development (Hazen & Gillis, 2017).

3.3.1 Determining best methods for capturing user performance on tactical verbal procedures

For the development of the training application for tactical voice procedures, the project utilized an initial contact report as a use case. The scenario, developed in consultation with RCN SMEs, consisted of a series of verbal communication by different agents following the identification of a new sonar object. The ASPO duties are to monitor these reports, and then assist in command by changing some CMS object properties, as well as provide a tactical verbal report on the situation. The initial contact report used to develop the simulation contained four parts/sentences. The first sentence identifies the intended recipient and the source of the message on the external channel. The second sentence repeats the new track information known at this point in time. The third sentence classifies the new track with a certain level of certainty. The last sentence closes the verbal report.

The design of the training application included three main elements related to the tutor interventions, the user interface, and specific task required by the execution of the training scenario. The tutor interventions include giving instructions, sequencing scenario events, and providing feedback on performance; the interface included various elements, such as buttons, that are extraneous to the execution of ASPO duties, but are necessary to assist the training/learning function (e.g., start/stop scenario). Finally, the specific ASPO duties task elements include assistance commands and tactical voice procedures.

In agreement with findings from the automatic speech recognition research literature, and using iterative user testing, the prototype was developed with the purpose to minimize the issues related to continuous speech recognition. The work flow for the execution of the training scenario imposed a strict step-by-step process in execution of the scenario which offers better monitoring of incoming information, and improves speech recognition robustness. The approach to develop the prototype was to minimize the need to develop artificial agents and focus on mapping trainees' speech input to expected utterances, using a scenario engine to provide text-to-speech feedback based on the mapping assessment. The mapping is performed by a scenario manager which relies on a training scenarios configuration file where scenario events are sequenced, and can potentially contain information related to hints and acceptable speech inputs. The results of the development efforts provided an acceptable level of realism for the synthetic agents, while keeping the complexity of the agent design and implementation very low.

3.3.2 Using simulated data to assess learning analytic methods of student learning

Another goal of the project was to demonstrate the potential of learning analytics to inform training designers. A set of simulated data was generated in order to present some examples of learning analytics methods that could be applied to monitor student learning and training materials. Student responses were simulated using a Bayesian knowledge tracing (BKT) model (Corbett & Anderson, 1995). A BKT model defines four parameters for individual skills or knowledge components to be acquired by a student, and allows to predict students' performance. The four parameters include two learning parameters (probability of knowing on the first trial, and probability of knowing on subsequent trials), and two performance parameters (probability of guessing when not knowing, and the probability of making an error when knowing). A BKT model generated responses for one hundred (100) simulated students on the task of giving an initial contact report. The simulated data were then processed by three learning analytics methods: Hidden Markov model, additive factors model, and a heterogenous data reduction method. The results offered stakeholders different data visualization options related to the demonstrated methods, as well as help determine data capture requirements to support skills monitoring related to performing sentence by sentence tactical voice procedures, or other training simulations aiming to capture learning data on specific skills.

4.0 CONCLUSION

Both new technologies and the move to use simulations to prepare people for teamwork are putting pressures on training simulations. Here we examined the use of agile development methods tailored to training requirements for developing synthetic teammates to support training for teamwork. We discussed the development and validation approaches and lessons learned. Particular challenges still exist in area of agile methods for tasks/behaviour modelling and the robustness of voice production/recognition.

Simulation-based training has been expanding in recent years from a near-exclusive emphasis on individual training focusing on task skills using part-task and full-mission simulators to an increasing inclusion of training for teamwork skills (mainly team communication and shared situation awareness). This shift in emphasis is leveraging advances in computational technologies, such as artificial intelligence, machine learning, cloud computing, and natural language processing, as well as a renewed interest in more “traditional” methodologies such as instructional design and training requirements analyses. This combination is leading researchers to investigate more agile approaches to system development for teamwork training, specifically using iterative, requirements-guided approaches to developing synthetic agents to play well-defined team roles, using a “shallow entity” modelling approach.

In the work presented above, we discussed a few projects wherein we applied such an agile development approach to a particular use case, namely preparing individuals to participate in team training events for the Royal Canadian Navy. The training requirements involved using synthetic agents to allow individuals to practise well-defined communications procedures while performing secondary tasks (maintaining situation awareness, or operating sonar equipment). In developing the instructional environments and synthetic teammates for these applications, we examined the use of cognitive modelling (e.g., ACT-R) and learning analytics methods (e.g., Bayesian Knowledge Tracing) as tools for assessing and predicting trainees’ performance (rather than for developing synthetic teammates to interact with directly), and explored the challenges with capturing trainee performance data (verbal and system interactions) to support these analytic methods. One lesson learned was that, despite recent advances in synthetic speech recognition and production, natural language interactions with synthetic agents are still a challenge, and workarounds were often required. Nevertheless, using simulated data, we were able to determine that the synthetic agent-based approach for training communications skills in multi-demand task environments can be effective.

Enabling individual trainees to “overtrain” fundamental communication skills that need to be performed concurrently with other mission-critical tasks, without requiring the use of team training resources (team simulators, live collective training, etc.) could be a game changer for military organizations such as the RCN. More work is required to advance this application area. One key effort needs to be the validation of the training effectiveness of such training technologies through human experimentation, as Cain et al. did for synthetic agents in another training use case (Cain et al., 2011). Another important area of investigation is an assessment of the applicability of the “shallow entity” modelling approaches beyond those involving well-defined teamwork interactions. For instance, is such an approach to modelling teammates sufficient to support scenarios where teammates commit errors or behave in more spontaneous was, or where other role players (e.g., an adversary) behave in unexpected ways to confuse and gain an advantage over the trainees? Or must the more traditional and onerous cognitive modelling approaches (Anderson, 2007; Anderson et al., 2004) be used in those cases? As the projects for the RCN described above indicate, “shallow” agents and more complex cognitive models can be profitably combined in ways other than directly modelling agents in a team task. Understanding the limits of the applicability of the more agile, “shallow entity” approach may help researchers and training practitioners understand how to combine these “deep” and “shallow” approaches more effectively in order to design systems for training teamwork skills that are more responsive to, and more scalable across, training requirements, and whose training effectiveness is easier to evaluate.

5.0 REFERENCES

1. Alexander, A., Brunyé, T., Sidman, J., & Weil, S. a. (2005). From gaming to training: A review of

studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group*, (November), 14. http://doi.org/10.1016/j.athoracsur.2004.02.012

2. Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York, NY, NY: Oxford University Press.
3. Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.
4. Baldwin, T. T., Ford, J. K., & Blume, B. D. (2017). The State of Transfer of Training Research: Moving Toward More Consumer-Centric Inquiry. *Human Resource Development Quarterly*, *28*(1), 17–28. http://doi.org/10.1002/hrdq.21278
5. Ball, J., Myers, C., Heiberg, A., Cooke, N. J., Matessa, M., Freiman, M., & Rodgers, S. (2010). The synthetic teammate project. *Computational and Mathematical Organization Theory*, *16*(3), 271–299. http://doi.org/10.1007/s10588-010-9065-3
6. Beaubien, J. M., & Baker, D. P. (2004). The use of simulation for training teamwork skills in health care: how low can you go? *Quality and Safety in Health Care*, *13*(suppl_1), i51–i56. http://doi.org/10.1136/qshc.2004.009845
7. Bell, B. S., Tannenbaum, S. I., Ford, J. K., Noe, R. A., & Kraiger, K. (2017). 100 years of training and development research: What we know and where we should go. *Journal of Applied Psychology*, *102*(3), 305–323. http://doi.org/10.1037/apl0000142
8. Berg, M. M. (2014). *Modelling of Natural Dialogues in the Context of Speech-based Information and Control Systems*. University of Kiel. Retrieved from http://macau.uni-kiel.de/receive/dissertation_diss_00016739
9. Boldovici, J. A., Bessemer, D. W., & Bolton, A. E. (2002). *The Elements of Training Evaluation*. Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.
10. Brawner, K. W., Goodwin, G., & Sottilare, R. (2016). Agent-Based practices for an intelligent tutoring system architecture. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9744, pp. 3–12). Springer, Cham. http://doi.org/10.1007/978-3-319-39952-2_1
11. Brawner, K. W., Sinatra, A. M. A. M., & Sottilare, R. (2017). Motivation and research in architectural intelligent tutoring. *International Journal of Simulation and Process Modelling*, *12*(3–4), 300–312. http://doi.org/10.1504/IJSPM.2017.085562
12. Cain, B., Magee, L., & Belyavin, A. (2011). Validating an HBR simulation for training within a virtual environment. In *20th Annual Conference on Behavior Representation in Modeling and Simulation 2011, BRiMS 2011* (pp. 154–161). Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84865379597&partnerID=40&md5=ccced83686a87d3c8195ff5112430032
13. Carroll, M., Thompson, T., & Deaton, J. (2008). Justification for Use of Simulation. In D. A. Vincenzi, J. A. Wise, A. Mouloua, & P. A. Hancock (Eds.), *Human Factors in Simulation and Training* (pp. 39–48). Boca Raton, FL: CRC Press. http://doi.org/10.1201/9781420072846.ch2
14. Cearley, D. W., Walker, M. J., & Burke, B. (2017). *Top 10 Strategic Technology Trends for 2017*. Stamford, CT.
15. Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, *37*(2), 255–285. http://doi.org/10.1111/cogs.12009
16. Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modelling and User-Adapted Interaction*, *4*, 253–278. http://doi.org/10.1007/BF01099821
17. Demir, M., McNeese, N. J., Cooke, N. J., Ball, J. T., Myers, C., & Friedman, M. (2015). Synthetic teammate communication and coordination with humans. In *Proceedings of the Human Factors and*

Ergonomics Society (pp. 951–955). http://doi.org/10.1177/1541931215591275

18. Dubreuil, A. (2016). *Human Factors Research and Modelling : Task 4 Investigation of Computer-Generated Forces Modelling of Information Layer Effects Final Report*.
19. Durlach, P. J., Bowens, L. D., Neumann, J. L., & Carnahan, T. J. (2004). *Coding verbal interactions in a prototype future force command and control simulation*. Alexandria, VA.
20. Dyer, J. L., Vaughn, A. W., & Blankenbeckler, P. N. (2004). *Training on Common Military Messages*. Alexandria, VA.
21. Emond, B., Maugeais, M., Vinson, N. G., Shaikh, K., Fournier, H., Lapointe, J.-F., & Martin, C. (2016). Adaptive training simulation using speech interaction for training navy officers. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)* (pp. 2924–2934). Orlando, FL: National Training and Simulation Association.
22. Emond, B., & Vinson, N. G. (2017). Modelling Simple Ship Conning Tasks. In *15th Meeting of the International Conference on Cognitive Modelling* (pp. 42–44). Coventry, UK.
23. Ford, J. K., Baldwin, T. T., & Prasad, J. (2018). Transfer of Training: The Known and the Unknown. *Annual Review of Organizational Psychology and Organizational Behavior*, *5*(1), 201–225. http://doi.org/10.1146/annurev-orgpsych-032117-104443
24. Granek, J. A., Jarmasz, J., Boland, H., Guest, K., & Bailey, L. (2016). Mobile Applications for Personalized Mental Health Resiliency Training. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)* (p. 14). Orlando, FL: National Training and Simulation Association. Retrieved from http://cradpdf.drdc-rddc.gc.ca/PDFS/unc269/p805213_A1b.pdf
25. Grant, S. C. (1999). Assessing intelligent software agents for training maritime patrol aircraft crews. In *Interservice/Industry Training, Simulation, and Education Conference* (pp. 830–837).
26. Grueneberg, K., de Mel, G., Braines, D., Wang, X., Calo, S., & Pham, T. (2013). Context-rich semantic framework for effective data-to-decisions in coalition networks. In *Proc. SPIE 8742, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IV*. http://doi.org/10.1117/12.2018022
27. Gunzelmann, G., Gaughan, C., Alexander, T., & Tremori, A. (2014). In Search of Interoperability Standards for Human Behaviour Representations. *I/Itsec*, (14027), 1–12.
28. Hazen, M. G., & Gillis, A. (2017). *Multi-role Reconfigurable Trainer Concept Final Report on Requirements and Architecture Investigation*. Defence Research and Development Canada.
29. Hernández-Orallo, J. (2017). *Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement*. *Artificial Intelligence Review* (Vol. 48). Springer Netherlands. http://doi.org/10.1007/s10462-016-9505-7
30. Hjelmervik, K., Nazir, S., & Myhrvold, A. (2018). Simulator training for maritime complex tasks: an experimental study. *WMU Journal of Maritime Affairs*, *17*(1), 17–30. http://doi.org/10.1007/s13437-017-0133-0
31. Hodges, G. A. (2014). *Identifying the Limits of an Integrated Training Environment Using Human Abilities and Affordance Theory*. Naval Postgraduate School, Monterey, CA. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA608057
32. Jurafsky, D., & Martin, J. H. (2017). *Speech and Language Processing*. Stanford, CA; Boulder, CO: Draft Third edition.
33. Liu, D., Macchiarella, N. D., & Vincenzi, D. A. (2009). Simulation fidelity. In D. A. Vincenzi, J. A. Wise, A. Mouloua, & P. A. Hancock (Eds.), *Human factors in simulation and training* (pp. 61–73). Boca Raton, FL: Taylor & Francis Group.
34. Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, *4*(3), 151–162. http://doi.org/10.1057/jos.2010.3
35. NATO-Bi-SC-75-2. (2013). *Education and Training Directive (E&TD)*. Neuilly-sur-Seine Cedex, France.

36. NATO-MSG-131. (2015). *Modelling and Simulation as a Service: New Concepts and Service-Oriented Architectures*. Neuilly-sur-Seine Cedex, France, France.
37. NATO/STO-TR-MSG-073. (2015). *Generic Methodology for Verification and Validation (GM-VV) to Support Acceptance of Models, Simulations and Data*. Neuilly-sur-Seine Cedex, France.
38. Orr, M. G., Lebiere, C., Stocco, A., Pirolli, P., & Kennedy, W. (2018). Multi-Scale Resolution of Cognitive Architectures : A Paradigm for Simulating Minds and Society. In R. Thomson, C. Dancy, A. Hyder, & H. Bisgin (Eds.), *Social, Cultural, and Behavioral Modeling* (pp. 3–14). Springer.
39. SISO. (2013). *Reference for Generic Methodology for Verification and Validation (GM-VV) to Support Acceptance of Models, Simulations and Data. GM-VV Vol. 3: Reference Manual*. Orlando, FL.
40. Togelius, J., Yannakakis, G. N., Karakovskiy, S., & Shaker, N. (2012). Assessing believability. In *Believable Bots: Can Computers Play Like People?* (Vol. 9783642323, pp. 215–230). http://doi.org/10.1007/978-3-642-32323-2_9
41. Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460.
42. Wilson, K. A., Salas, E., Priest, H. A., & Andrews, D. (2007). Errors in the Heat of Battle: Taking a Closer Look at Shared Cognition Breakdowns Through Teamwork. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(2), 243–256. http://doi.org/10.1518/001872007X312478