Measuring trainee intent using low-cost, high-impact cognitive models

Sylvain Bruni, Nathan Schurr, Brian Riordan, Jeanine Ayers
Aptima, Inc.
Woburn, MA
sbruni@aptima.com, nschurr@aptima.com,
briordan@aptima.com, jayers@aptima.com

1st Lt Shaun Sucillon
Wright Patterson AFB
Dayton, OH
shaun.sucillon@wpafb.af.mil

ABSTRACT

In order to achieve an increasingly dynamic and nuanced commander's intent, warfighters must understand when and where to apply their skills most effectively. Current training methods, though very effective at producing skilled warfighters, focus primarily on lower level skills and outcome-based performance. However, there is a need to assess the warfighter at the level of intent and how the warfighter factors that into their process of skill selection and skill execution. Cognitive models appear as a promising solution to understanding warfighter processes and intent. Yet, traditional cognitive models designed to replicate human cognitive processes are cumbersome to develop and maintain, requiring large amounts of data.

An innovative capability was designed to address these challenges by leveraging advances in training technology that increase data availability to capture warfighter actions and behaviors during training while applying recent research findings focused on understanding intent from actions (Baker, Saxe & Tenenbaum 2007). This capability integrates a modeling method to infer intent from actions, by employing Markov Decision Processes and Bayesian inverse planning.

This paper will describe initial testing and evaluation of this technology with novice remotely-piloted aircraft operators and show the model's ability to infer intent and predict operator actions with a satisfying level of reliability. Initially implemented in a basic research setting, this modeling method is currently being transitioned to simulation and training environments with gradually increasing level of fidelity, beginning with an operationally-relevant, game-based training environment. This paper will describe the transition plan and discuss how this modeling approach constitutes an example of a new generation of practical, lightweight, and extremely useful cognitive models.

ABOUT THE AUTHORS

Sylvain Bruni is a Human Systems Engineer at Aptima, Inc., where he provides expertise in human-automation interaction, interface design, and the statistical design of experiment. His research targets the design of computer-supported interactive training systems and the conceptualization of human-automation collaboration interfaces for multi-unmanned aerial vehicles (UAV) command and control. Prior to joining Aptima, Mr. Bruni conducted research at the Massachusetts Institute of Technology (MIT), focusing on designing and testing collaborative decision-support systems for UAV mission planning and replanning. His work further included the detection and classification of UAV operators' cognitive strategies as a means to assess interface and system design. Mr. Bruni's background also includes investigating the effects of information display and levels of automation on system performance, as well as defining evaluation tools and methodologies for complex human-computer systems. Mr. Bruni holds a S.M. in Aeronautics and Astronautics from MIT and a Diplôme d'Ingénieur from the Ecole Supérieure d'Electricité (Supélec, France). He is currently a doctoral candidate in Aeronautics and Astronautics at MIT. He is a member of the Human Factors and Ergonomics Society, the IEEE Systems, Man, and Cybernetics Society, the Association for Computing Machinery, and the DoD Human Factors Engineering Technical Advisory Group.

Nathan Schurr is a Human-Agent Collaboration Scientist and the Team Lead for the Artificial Intelligence group at Aptima, Inc. Nathan Schurr's interests are in allowing humans and artificially intelligent entities to collaborate and interact. He has applied this expertise to domains ranging from software personal assistants to human-multirobot teams. Dr. Schurr is the creator of an Incident Command training system, DEFACTO, which has been tested by the Los Angeles Fire Department. The system allows human incident commanders to coordinate with fire fighter agents during a large-scale disaster response. Dr. Schurr holds a Ph.D. in Artificial Intelligence from the University of Southern California, where he was awarded the Viterbi School of Engineering Homeland Security Center Doctoral Fellowship. He holds a M.S. in Computer Science from the University of Southern California and a B.S. in Computer Engineering from California Polytechnic State University San Luis Obispo. He is a member of the Association for the Advancement of Artificial Intelligence.

Brian Riordan is a Research Scientist at Aptima, Inc. He specializes in machine learning, statistical natural language processing, and cognitive modeling. His recent research includes machine learning approaches to inferring user intent and predicting human performance, large-scale text-analytics, and pattern recognition in multi-source intelligence data. Dr. Riordan holds a Ph.D. in Linguistics and Cognitive Science from Indiana University, an M.A. in Computational Linguistics from Indiana University, and a B.A. in Linguistic Anthropology and East Asian Studies from New York University. He is a member of the Association for the Advancement of Artificial Intelligence (AAAI) and the Association for Computational Linguistics (ACL).

Jeanine Ayers is a Software Architect and Team Lead for the Intelligent Distributed Systems Team in Aptima's Software Engineering Division. Her experience includes leading small teams of software engineers in many development efforts at Aptima, Inc. focusing on architecting and building data-driven software applications in support of multi-agent systems, computer based training systems and performance measurement and assessment systems. Ms. Ayers has an M.B.A from Boston University and a B.S. from Carnegie Mellon University in Industrial Management/Information Systems.

1st Lt Shaun Sucillon is a behavioral scientist assigned to the Air Force Research Laboratory's 711th Human Performance Wing, Warfighter Readiness Research Division at Wright-Patterson Air Force Base in Dayton, Ohio where he manages their Remotely Piloted Aircraft (RPA) Training Research Program. This multiyear effort is exploring areas where learning science, tools, and M&S technology can be developed, leveraged, and applied to improve the quality of preparation, mission planning, execution, AAR and refresher training for the Air Force's MQ-1, MQ-9, and RQ-4 RPA communities. Lt Sucillon earned his Bachelor of Science degree in Behavioral Sciences from the United States Air Force Academy in Colorado Springs, CO.

Measuring trainee intent using low-cost, high-impact cognitive models

Sylvain Bruni, Nathan Schurr, Brian Riordan, Jeanine Ayers
Aptima, Inc.
Woburn, MA
sbruni@aptima.com, nschurr@aptima.com, briordan@aptima.com, jayers@aptima.com

1st Lt Shaun Sucillon
Wright Patterson AFB
Dayton, OH
shaun.sucillon@wpafb.af.mil

INTRODUCTION

In order to achieve an increasingly dynamic and nuanced commander's intent, warfighters must understand when and where to apply their skills most effectively. Consequently, a need to understand how the warfighter factors intent into their process of skill selection and skill execution has emerged. While cognitive models are a promising solution to assess the warfighter at the level of intent, they are traditionally designed to replicate human cognitive processes, are cumbersome to develop and maintain, and require large amounts of data.

To address these limiting factors, we have designed an innovative capability: the Mixed Initiative Machine for Instructed Computing, a modeling method to infer intent from actions, by employing Markov Decision Processes and Bayesian inverse planning.

This paper will describe initial testing and evaluation of this technology with novice remotely-piloted aircraft operators and show the model's ability to infer trainee intent and predict trainee actions. Initially implemented in a basic research setting, this modeling method is currently being transitioned to simulation environments with gradually increasing level of fidelity, beginning with an operationally-relevant, game-based training environment. This paper will describe the results of this transition and discuss how this modeling approach constitutes an example of a new generation of practical, lightweight, and useful cognitive models.

BACKGROUND

Current training methods, though very effective at producing skilled warfighters, focus primarily on lower level skills and outcome-based performance. However, considering the crucial need to satisfy commander's intent, assessing how the warfighter translates intent into action has become a necessity: How do warfighters process intent? What priorities or goals constitute high-order cognitive constraints that drive skill selection and execution? Cognitive models are a promising solution to understanding warfighter processes and intent. Traditional cognitive models such as Soar (Laird et al., 1987) or ACT-R (Anderson and Lebiere, 1998) have typically been constructed to reproduce in details how human cognitive processes operate: these architectures generally focus on modeling fine-grained human behavior and decision-making and include low-level characteristics of human cognitive processes. Consequently, they are fastidious to apply and maintain, and require large amounts of data to be useful.

MODEL DEVELOPMENT

In order to enable the assessment of warfighter intent without entailing cumbersome model development, we took a radically different approach and developed a low-cost, high-impact cognitive model of intent that leverages a framework of Bayesian inference for understanding intent from actions (Baker, Saxe & Tenenbaum 2009) as well as advances in training technology that make it feasible to capture warfighter actions and behaviors during training. In contrast with existing cognitive architectures, we seek to model human high-level goals and priorities by making inferences from observed data. In turn, using the model's inferences, rather than predict the detailed time-course of cognitive processing, our aim is to accurately predict the content of future human behavior (e.g., whether the operator will choose one possible action versus another).

Domain

In order to scope the development of the cognitive model, we focused on the domain of multiple-remotely piloted aircrafts (RPAs) mission planning and execution. We developed a use-case to illustrate how this modeling method will be used to support a mission planner as they are tasked with the creation of an air

***Distribution A**: Approved for public release; distribution unlimited per 88ABW-2011-3701, 28 June 2011.*

tasking order (ATO) for four RPAs in an intelligence, surveillance and recognition (ISR) mission. Using a resource allocation planning interface (Figure 1), the operator develops a plan that specifies what RPA goes to what target, in what order, at what speed, and with what purpose. The operator is provided with rules of engagement (ROEs) and instructions reflective of the commander's intent, i.e., a set of constraints and priorities (such as "target X requires a covert approach" or "avoid weather as much as possible"), as well as objectives to maximize or minimize (such as "maximize the number of targets covered" or "minimize fuel consumption").

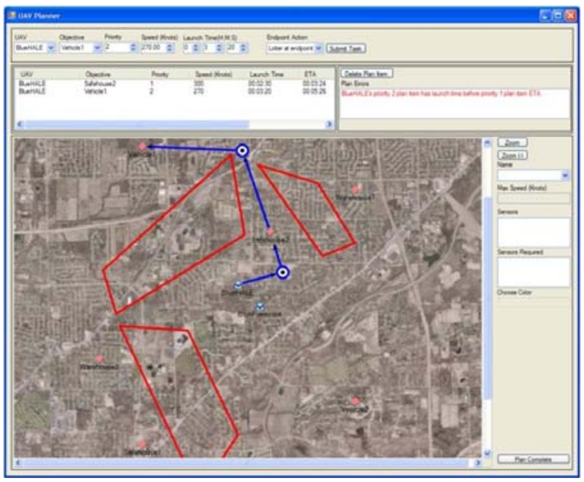

Figure 1. Multi-RPA Mission Planning Interface

Model Structure

At the core of the system is a model capable of inferring an operator's mission planning priorities. Planning priorities refer to those outcomes that operators try to rank and optimize over, during the mission planning process. These include, for example, "assign all targets to at least one RPA" (coverage priority) or "avoid weather zones" (weather avoidance priority). We hypothesize these priorities can be inferred from the observable actions performed by an operator during mission planning: the model seeks to capitalize on regularities in planning actions to predict an operator's likely priorities and future actions.

The machine learning component of the model specifies the causal relation between planning priorities and planning actions as a Markov Decision Process based on rational probabilistic planning. Bayesian inference is used to invert the causal relation between priorities and actions, using observed action sequences to infer the most probable priorities that led to these actions (Baker et al., 2009).

The principal properties of the Bayesian inverse planning framework are:

- Operators can be effectively modeled as approximately rational planners. In attempting to capture the priorities and behaviors of agents, we assume they will choose the actions that most efficiently lead them to the accomplishment of their goals while maintaining their priorities.

- Inverse planning can be accomplished by integrating bottom-up information from observed data and top-down constraints from a hypothesis space of possible goals. This approach allows inference of an agent's latent goals and preferences, as well as prediction of the agent's future actions.

Goal-based planning as a Markov Decision Process

MDPs are a machine learning framework for sequential decision-making under uncertainty (Puterman, 2005). Given an environment and a goal, a trained MDP model specifies a course of action from any state of the world that maximizes the rewards to a participant.

By employing an MDP to model an operator's planning behavior, we assume they act rationally in making planning decisions, choosing actions likely to bring them closer to achieving their goal.

A Markov Decision Process comprises a state variable, a model of the environment, and a set of rewards or costs. The state variable includes information about the operator's state and the configuration of the environment. The environment specifies what actions the operator can perform, and a causal model of how these actions change his state and the environment. Actions are associated with a reward or cost which may be received upon performing an action or for performing an action and transitioning to the next state (Puterman, 2005).

In the MDP formulation, actions are chosen probabilistically. This allows the model to account for noise and variation in how operators create plans given the same scenarios. The MDP maintains a probability distribution over actions P(Actions | Goal, Environment).

Bayesian inverse planning

In an MDP-based goal inference model (Figure 2), the inferences that can be made about operators' goals during planning depend on the structure of the goal hypothesis space and the prior probabilities of goals assigned within the model. In the current implementation of the model, we make the simplifying assumption operators have a single optimal plan state, or goal, which they attempt to achieve throughout each mission scenario. A priority is associated with a number of goal states, the only requirement being that a goal state must carry out at least one priority (e.g., "avoiding weather zones"). Thus a single goal state may involve carrying out a single priority or multiple priorities simultaneously.

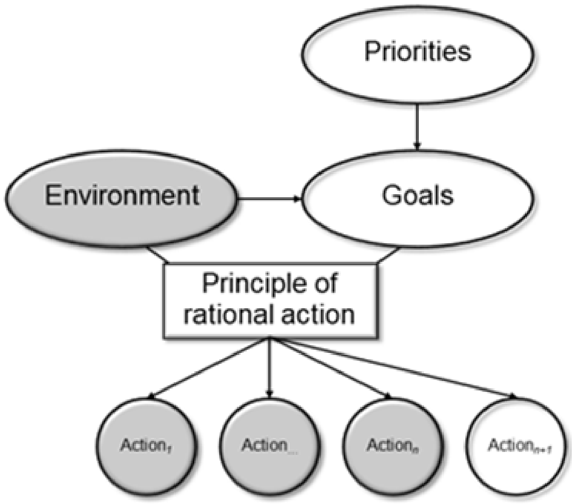

Figure 2. The hypothesized causal structure of goal- and priority-based planning. Shaded nodes represent observed variables; unshaded nodes are latent variables whose values must be inferred.

The model includes only one parameter, β, which controls the level of determinism with respect to optimal action selection that the MDP model is able to fit. High values for β will fit participants whose actions follow the optimal sequence of actions to their goal, while low values for β will better fit participants who deviate from this sequence. In future work, we plan to implement more complex models in which goals may change throughout the course of planning.

The technical details of the implementation of the model are described elsewhere (Riordan et al., 2011).

Model Population and Training

In order to populate the model with appropriate features (namely, the priorities, goals and constraints to look for) and, subsequently, to train the model on a body of relevant data, we conducted a controlled, human-in-the-loop experimentation where participants used the planning interface to complete 18 scenarios. Planning actions, in the form of XML log files, were collected automatically and a cognitive walkthrough was performed after each scenario to elicit from participants what priorities, goals, and constraints they employed while creating the mission plans.

Method

Forty-two undergraduate and graduate students, aged 18-69 years (average 30 years ± 12.5), participated in the experiment. Six had prior military experience (from ROTC student to Commissioned Officer), seven had prior aviation background (from pilot licenses to flying courses), eleven were familiar with the RPA domain, five had mission planning experience (in the military or civilian world), and eight had participated in other controlled experiments featuring RPAs. All were compensated $10/hour.

The experiment duration averaged three hours and included three phases: training, data-gathering, and debriefing. First, participants underwent a 45-minute training session, which included a series of PowerPoint training modules and the completion of two practice scenarios, under experimenter supervision. Then, participants played multiple simulation scenarios in which they were tasked with creating a mission plan for two RPAs in a hostile environment based on a set of rules of engagement (ROE). Participants were responsible for assigning RPAs to targets, setting parameters and constructing flight routes. Finally, a cognitive walkthrough protocol was implemented.

This experiment included two independent variables: planning time (that is, how long the participant was allowed to plan the RPAs' missions) and scenario complexity. There were two levels of planning time: short (3 minutes) and long (6 minutes). There were three levels of scenario complexity: low (10 targets and few ROE constraints), medium (20 targets and few ROE constraints), and high (20 targets and many ROE constraints). ROE constraints typically included the number of weather zones to avoid, the maximum total flight time, fuel allowance, and the number and duration of actions to perform at targets. A repeated-measures design was implemented, in which all participants saw three replicates of the six conditions (two planning times by three scenario complexities), yielding a total of 18 trials per participant. Blocking of the randomized replications was counterbalanced across participants.

During each trial, the simulation testbed recorded interface interactions and plan state in XML log files. This data was pushed to a database for subsequent use by the model.

Additionally, at the conclusion of the 18 trials, a cognitive walkthrough (CWT) protocol was performed and included the following elements: (1) a review of a video of the last scenario was played while participants commented on their decision-making processes, their strategies and the objectives, constraints and priorities they tried to satisfy; (2) a series of questions was administered to reveal how their decision-making process and strategies changed under various scenario conditions, and to specify what their main priorities were, what trade-offs they considered, and what constraints were of higher importance to them.

Results

Seven hundred and fifty-six XML log files (18 scenarios by 42 participants) were recorded during this experiment to constitute the initial training data for the machine learner.

Among other questions, the CWT protocol asked participants to mention and rank-order what they considered to be their planning priorities during the various scenarios they played. Fourteen subjective planning priorities were described by participants: “Avoid weather zones,” “Use closest targets,” “Minimize threat level,” “Monitor planning time,” “Monitor mission time,” “Use adequate covertness,” “Satisfy ROE objectives,” “Address biggest threat,” “Optimize speed,” “Optimize vehicle use,” “Minimize distance traveled,” “Balance vehicle load,” “Create complete plan,” and “Minimize fuel usage.”

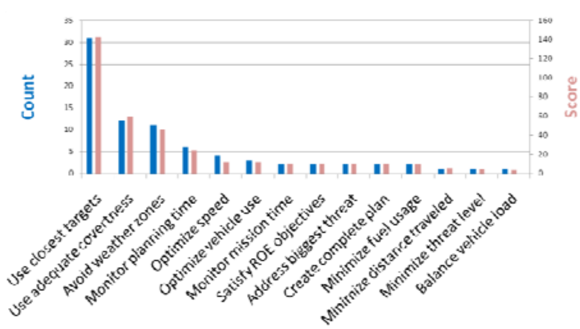

Figure 3. Count and Score for each priority mentioned.

Figure 3 displays the count of participants who mentioned each priority, and each priority’s total score based on the rankings provided by participants. The scoring metric is a weighted function where the score is increased by 5, 4, 3, 2 or 1 point(s) when the priority is ranked respectively 1^{st}, 2^{nd}, 3^{rd}, 4^{th} or 5^{th}. It appears that “Use closest targets” was, by far, the prevalent priority, measured by count or score. “Use adequate covertness” and “Avoid weather areas” came in second and third position respectively. This trend existed regardless of the participants’ backgrounds. Table 1 describes the trade-offs associated with these three top planning priorities.

Table 1. Selected planning priorities.

Priority	Description & Tradeoffs
Use closest targets	The operator may assign UASs to the targets closest to them. Doing so may involve tradeoffs with covertness or entry into dangerous weather zones.
Use adequate covertness	The covertness capability is not shared by all UASs; hence the operator must consider the impact of routing a covert UAS to one target on its availability to cover other targets.
Avoid weather areas	Going around weather zones might take more time than going through them. The operator must consider this trade-off in the route planning.

These three priorities were selected to be represented in the model. In the model’s state variable, we encoded how well participants maintained these priorities with each action they took, along with other features describing the configuration of the environment. Participants’ goals were represented as states in the model in which the mission was completed and one or more of the three priorities was maintained (e.g., no weather zones were crossed in any plan). We assumed each action taken by a participant was an effort to maintain a priority and achieve a goal state. Based on how the participant’s actions change the state of the environment, the model classifies the action into one of the following predefined actions:

- **New plan: avoid weather**. A UAS is assigned to an available target such that the route avoids weather zones.
- **New plan: proximity**. A UAS is assigned to an available target that is closer to it than to the other UAS.
- **New plan: maintain covertness.** A UAS is assigned to an available target such that the UAS chosen meets the covertness requirements specified in the scenario.

- **Plan modification: avoid weather**. The operator alters an existing plan so the UAS's trajectory avoids weather zones.

- **Plan modification: proximity**. The operator implements modifications based on proximity. These modifications include both the objective (i.e., change the target to a different, unassigned target to minimize distance travelled) and the assigned UAS (i.e., assign a different UAS to fly to the target to minimize distance travelled).

- **Plan modification: maintain covertness.** The operator changes the target of an existing plan so a UAS is assigned to a target based on the covertness requirements specified in the scenario.

MODEL EVALUATION

Following the first experiment that led to the selection of model feature and to the training of the model's algorithms, we conducted two additional experiments, replicating exactly the first one, to evaluate the ability of the model to infer operator priorities and to predict operator actions.

Priority Inference

Method

To assess the model's performance at priority inference, we repeated the first experiment with fifteen undergraduate and graduate students, aged 18-43 years (average 26 years ± 7.5). While the experimental procedure was the same, at the end of each scenario, we asked participants to indicate their subjective assessment of what planning priorities they used in the scenario they just completed: each priority was rated for importance and time spent attending to it, and their order of presentation was fully randomized to avoid response bias.

Table 2. Correlation between operator ratings and model predictions for several values of β

β	0.5	1	1.5	2	2.5
r	0.315	0.283	0.264	0.256	0.248

Results

Using the algorithms trained using data from the first experiment, the model was applied to data from this second experiment, and we compared the model outputs (i.e., operator priorities inferred by the model's algorithms) to the operators' subjective assessments (i.e., operator priorities as stated by the operator). Several values of the β parameter were tested. Table 2 presents the correlation values obtained for five values of β. $\beta = 0.5$ led to the highest correlation $r = 0.315$.

Subsequently, we broke down the results by planning priorities, focusing on the three priorities featured in the model. The correlation results are presented in Table 3: the "covertness" priority led to the highest correlation of $r = 0.562$.

Table 3. Correlation r between operator ratings and model predictions for each planning priority

Priority	Covertness	Proximity	Avoiding weather
r	0.562	0.365	0.359

The results of this second experiment show the model can infer operators' priorities at a reasonable level of reliability. This work extends the results obtained by Baker et al. (2009), and shows Bayesian inverse planning can both account for operator goals and the preferences that guide the selection of these goals.

Action Prediction

Method

To assess the model's performance at predicting operator actions, we repeated the first experiment with eight undergraduate and graduate students, aged 18-54 years (average 27 years ± 11.7). While the experimental procedure was the same, after the experiment, two human coders observed all videos of all scenarios for all participants and manually coded each of their actions in a database by assigning a probability that each participant action fell into each of the six action types represented in the model. For this experiment, the low-level actions with the planning interface were grouped together into sequences that corresponded to the high-level priorities encoded in the model. Inter-coder reliability using Cohen's Kappa was 0.737.

Table 4. Correlation r between coder ratings and model predictions for several values of β

β	0.5	1	1.5	2	2.5
r	0.551	0.572	0.577	0.576	0.570

Results

The trained model was applied to the data of Experiment 3 and the prediction output was compared to the coders' assignment of probability distributions to actions. This evaluation was performed for a range of values of the β parameter. Table 4 lists the resulting correlation values obtained for five values of β. The highest correlation (r =0.577) was obtained for $\beta = 1.5$.

The results of this third experiment show the model can reasonably predict an operator's action during planning. Future enhancements to the model will seek to improve these scores.

TRANSITION TO A SIMULATION ENVIRONMENT

Motivation

The modeling method described in the previous section helps embed into the Remote Piloted Aircraft (RPA) novel models of tactical decision making and control from well-selected exemplars and interactions with human operators. This method is designed to benefit RPA by enabling them to reach their full autonomous potential. More specifically, this method will enable the following: (1a) expert human operators to teach the RPA to overcome problems normally caused by lack of context in automated control systems, (b) RPA to adapt more easily to dynamic and uncertain environments (through use of model restructuring software), and (c) learning in real-time in test and simulation environments, which will enhance and accelerate knowledge transfer of tactical maneuvers and new autonomous behaviors.

Transition Environment

The Gaming Research Integration for Learning Laboratory (GRILL) is one of several Air Force Research Laboratory (AFRL) sponsored research programs to evaluate gaming for training. More specifically, the GRILL is interested in answering questions such as how effective is game-based training, what aspects of gaming can be applied to the training realm, and what are the implications for games as part of a family of complimentary trainers in a Live, Virtual, and Constructive (LVC) architecture. The GRILL contains multiple types of training technologies that are integrated in a distributed network infrastructure.

The current GRILL environment as configured for the model integration is illustrated in Figure 4. The GRILL is composed of the following modules: scenario and environment management, training platforms, and a distributed network infrastructure. The training platforms are the center of the GRILL environment. There are three training platforms that are of interest to this integration. Each of the platforms is equipped with software that allows the simulation (based on the trainee actions) to publish relevant information over the Distributed Interactive Simulation (DIS) protocol. In addition to publishing data over DIS, each trainee platform is equipped with an environment manager. The environment manager provides instructions and information regarding the scenario to the players during the training. The first platform is the Joint Terminal Attack Controller (JTAC), which is for a

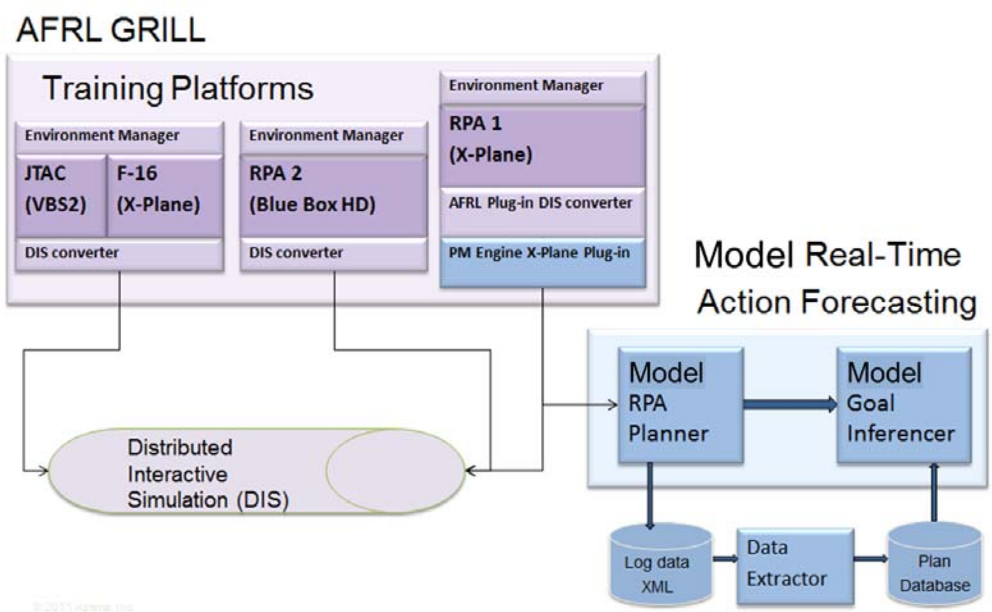

Figure 4. Integration Environment

forward observer on the ground. The JTAC training platform is developed as a Virtual Battle Space 2 (VBS2) ground simulation. In addition to the JTAC, there are also F-16 X-Plane platforms that are the supporting entities in the environment. The F-16 pilots use a modified version of X-Plane. The integration we focus on concerns RPA simulations. Currently, RPA training platforms are found in two configurations. The first RPA simulation is developed by L-3 Link and simulation training using their proprietary Blue Box HD system. The second RPA simulation is a modified predator model using X-Plane. The X-Plane environments use a custom plug-in created by AFRL to publish data over the DIS protocol. Each of these training platforms run on their own PC's which are networked through a gigabit switch. Each simulation platform is using a common set of terrain and model databases correlated with real-world data. These programs are tied together through DIS and through the environment manager software. The environment manager sends instructions and interacts with the players in a controlled manner. This software was originally created by AFRL. In the future, the plan is to replace the custom controller with a new environment manager provided by a third party vendor.

DISCUSSION

The integration of the model in the GRILL environment is continuing throughout the summer of 2011. At this point, an integration plan and a training scenario have been developed. The work of integrating the technology with the data sources in the GRILL laboratory is ongoing. By the end of the summer, the hope is to demonstrate how a model that is able to infer operator intent and goals can lead to more responsive adaptive training environment including: (1) better and targeted adversarial behaviors, (2) more directed decision support that enables effective decision making by the trainee, and (3) real-time scenario modifications to enrich the training environment. We believe this modeling approach constitutes an example of a new generation of practical, lightweight, and useful cognitive model that can be used to improve training effectiveness.

CONCLUSION

We described the development and evaluation of a new modeling method in the multi-RPA mission planning domain and showed how the Bayesian inverse planning framework can be used to infer intent from human user actions.

We described our current efforts to transition this technology to the training domain by integrating the model into AFRL's GRIL laboratory.

This integration, scheduled to be completed by the fall of 2011, will enable improved training effectiveness in simulated training environments.

ACKNOWLEDGEMENTS

The authors wish to thank Jennifer Roberts and Jonathan Lansey of Aptima for their contributions to model development and evaluation, Dr Nancy Cooke, Noel Rima and Ivonne Murray of the Cognitive Engineering Research Institute, as well as Dr Marc Steinberg of the Office of Naval Research (ONR) for his support of this research. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of ONR or the U.S. government.

REFERENCES

Anderson, J., & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum.

Baker, C. L., Saxe, R., and Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition,* 113, 129-349.

Laird, J., Rosenbloom, P., and Newell, A. (1987). Soar: An Architecture for General Intelligence. *Artificial Intelligence*, 33:1-64.

Ng, A., and Russell, S. (2000). Algorithms for inverse reinforcement learning. In *Proceedings of the 17th International Conference on Machine Learning (ICML 2000).*

Puterman, M. L. (2005). Markov decision processes: Discrete stochastic dynamic programming. New York: John Wiley and Sons.

Ramachandran, D., and Amir, E. (2007). Bayesian inverse reinforcement learning. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007)*.

Riordan, B., Bruni, S., Schurr, N., Freeman, J., Ganberg, G., Cooke, N.J., Rima, N. (2011). Inferring user intent with Bayesian inverse planning: Making sense of multi-Unmanned Aerial Systems mission management*. 20th Annual Conference on Behavior Representation in Modeling Simulation*, Sundance, UT.