



**SCHOOL OF INFORMATION SCIENCES AND TECHNOLOGY
THE PENNSYLVANIA STATE UNIVERSITY**

**Improving Interfaces for CGFs through
Multidisciplinary Evaluations: A New, Broad
Approach**

Frank E. Ritter and Marios N. Avraamides

ritter@ist.psu.edu marios@psu.edu

Technical Report No. ACS 2001-1

15 November 2001

acs.ist.psu.edu

Phone +1 (814) 865-4453 Fax +1 (814) 865-6426

School of IST, Thomas Building, University Park, PA 16802

20 November 2001

Improving Interfaces for CGFs through Multidisciplinary Evaluations: A New, Broad Approach

Frank E. Ritter and Marios N. Avraamides

ritter@ist.psu.edu marios@psu.edu
School of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802

ACS #2001-1

15 November 2001

Abstract

Typically, the design of cognitive models has not emphasized the role of interfaces for describing the model's behavior. Models that populate synthetic environments are particularly complex and need support in this area. Using subject-matter experts we evaluated the use of the Situation Awareness Panel (SAP) as a tool for inspecting the behavior and reasoning of Soar agents in a JSAF simulation. We gathered suggestions on how to improve future implementations of the SAP from experts in a variety of disciplines (e.g., military pilots, cognitive psychologists, an HCI specialist, a logistic specialist, and a software designer). We used their behavior and reports to develop a task analysis that can be used as a general guide for future designs of user interfaces for cognitive models in general and for the design of interfaces for models in synthetic environments in particular. We suggest this approach of having multiple types of experts review an interface as a general method for improving complex interfaces such as interfaces to cognitive models.

20 November 2001

Table of Contents

1. Introduction: Expert evaluations of interfaces.....	1
2. Expert evaluations of the Situation Awareness Panel (SAP)	2
2.1 Cognitive walk throughs, heuristic evaluation, and semi-structured interviews ...	3
2.2 The Situation Awareness Panel	4
2.3 The study	6
2.3.1 Participants	6
2.3.2 Material and equipment	6
2.3.3 Procedure.....	6
2.4 Analysis and Results	10
3. Conclusions	12
4. References.....	14
Appendix A. Feedback on the initial Situation Awareness Panel	16
Appendix B. The new Situation Awareness Panel.....	21
Appendix C. Feedback on the new Situation Awareness Panel.....	22

1. Introduction: Expert evaluations of interfaces

Newell's (1990) call for a unified theory of cognition has led to a new way of research in cognitive psychology. During the last two decades many researchers have started examining psychological phenomena through situation-specific theories that they implement as computer programs. These theories, called cognitive models, remain confined to the constraints imposed by grand theories of cognition, generally known as cognitive architectures.

Cognitive architectures and cognitive models have been used extensively as means for exploring the mechanisms involved in human cognition. However, cognitive models have been also built with the goal of being used as surrogate users. Models with human-like behavior can replace humans in many situations ranging from cognitive tutoring (Anderson, Corbett, Koedinger, & Pelletier, 1995) to usability testing of interfaces (Ritter & Young, 2001; St. Amant, 2000).

The use of cognitive models as surrogate users is especially appealing for situations in which human expertise is either costly or difficult to recruit. Military training has been one of those fields that typically require a great amount of human resources. Cognitive modeling provides an alternative avenue for supporting military training. Cognitive models as intelligent agents can populate synthetic environments representing some or all of the entities involved in real combats, thus enabling the use of realistic environments for training purposes (Pew & Mavor, 1998; Ritter, Shadbolt, Elliman, Young, Gobet, & Baxter, in press). One such attempt has been the TacAir-Soar system (Tambe, Johnson, Jones, Koss, Laird, Rosenbloom, et al., 1995), which employs cognitive models developed with the Soar architecture (Laird, Newell, & Rosenbloom, 1987; Newell, 1990) to simulate the behavior of military personnel in fixed-wing aircraft missions. The benefits of using TacAir-Soar are particularly evident in large-scale simulation exercises in which many of the entities involved can be driven by Soar models instead of human users. For example, in the Stow '97 exercise, up to 3,700 computer-generated forces were involved as both friendly and enemy entities (Jones, Laird, Nielsen, Coulter, Kenny, & Koss, 1999).

While using cognitive models to either answer psychological questions or to replace human users provides great advantages, serious problems have been identified as well. One of the problems is the limited reuse of cognitive models. It seems fair to say that cognitive models are not typically used by researchers other than the ones who developed them. Part of the problem can be attributed to the lack of graphical-user interfaces for many of the models developed (Ritter, Jones,

& Baxter, 1998). Without graphic displays, observing and understanding the behavior of the cognitive models is restricted, which can contribute to limiting their adoption by others.

The non-optimal design or the total absence of graphical displays that are needed to make the behavior of the models visible make the validation of the models problematic as well. Subject-matter experts, who are not programmers themselves, will have difficulties evaluating the behavior of the model based on traces of the running program. In order to understand the model, these users need a clearer form.

The need for improved user interfaces for cognitive models is particularly important for models that populate synthetic environments. These environments are typically loaded with a great number of computer-generated forces that necessitates that their behavior is easily observable if it is to be understood. As Ritter, Jones, and Baxter (1998) point out, graphical user interfaces have led in the past to new understanding about the behavior of models. When a graphical interface was added in Soar (Ritter & Larkin, 1994) it became evident that the Soar models searched through the problem spaces rather than in them.

The present study tested one such graphical interface. We recruited a number of subject-matter experts from a wide variety of related fields and asked them to observe the behavior of Soar agents that fly fixed-wing aircraft in a JSAF simulation. Their understanding of the model in some cases led to a type of Turing-like test, where they were attempting to judge if the model's behavior was similar to a human's.

We have used comments on ways to improve the specific graphical interface as well as their behavior with the interface to provide a list of suggested improvements. We have also created a task analysis that can be used to improve the design of the interface we studied and of modeling interfaces in general, based on their comments and our own experience with models.

2. Expert evaluations of the Situation Awareness Panel (SAP)

The goal of the project was to understand and improve the SAP as a tool for inspecting the behavior and reasoning of the Soar agents that populate the JSAF simulation environment. Our attempt was focused on the validity and usability of the SAP.

Validity refers to whether the types of information displayed on the SAP is truly in the awareness of actual pilots engaged in air combat. Usability refers to whether people using the SAP can understand the model based on what they can see through the SAP.

In order to examine these issues we recruited people with a variety of expertises and asked them to perform a number of basic tasks while observing the awareness of the agents during a preprogrammed scenario. In addition to expert pilots, our list of participants included experts from various other domains that we thought were related to different aspects of the JSAF simulation. Such domains included Cognitive Psychology, Geographical Information Systems (GIS), Human-Computer Interaction (HCI), software development, and the military. Our goal was to assemble a multidisciplinary subject pool in order to get feedback about the functioning of the SAP from a variety of perspectives.

This work is similar to a variety of evaluation techniques, including heuristic evaluation, cognitive walk-throughs, and semi-structured interviews with the addition that we used a variety of experts. This work resulted in a task analysis rather than being based on a task analysis.

2.1 Cognitive walk throughs, heuristic evaluation, and semi-structured interviews

There are a variety of approaches that we could have used to examine the usability of the SAP interface. We could have done a task analysis (Schraagen, Chipman, & Shalin, 2000) if we had a list of what tasks users were performing, but we were attempting to create such a list. We were not trying to optimize performance, per se, so timing users on tasks was not appropriate either. We were not just looking at learning, as our users would either just be watching it (with very little learning, we would hope), or would be working with it for a while (with quite extensive learning) and did not have a task analysis in hand, so cognitive walk-throughs (Polson, Lewis, Rieman, & Wharton, 1992) seemed not quite appropriate either.

In the end, we did what could be described as guided heuristic evaluation. We prepared a subset of tasks that we knew the interface would be used for. We had potential and existing users and a variety of usability experts broadly defined perform these tasks with the interface. We observed these users and also had them comment on the problems they had. After performing these tasks, we debriefed them in order to find out what other tasks they would like to have been able to perform with the interface. In some cases, they could do these tasks, in others, we were able to add these tasks to our developing task analysis. In some ways, our approach was similar to semi-structured knowledge acquisition interviews (see, for example, SigArt ACM Special Interest Group on Artificial Intelligence, 1989).

We believe that using a multidisciplinary participant pool for validating interfaces of military simulations is a necessity due to the variety in the nature of the information that is typically

displayed. For example, a situation-awareness display, such as the one we evaluated, contains information that varies from the execution of standard combat routines to awareness about the terrestrial terrain, memory for past events, perception of various sorts of input, aircraft logistics (i.e., fuel and weapon status) and so on. Instead of relying on our own common sense to evaluate the way the various types of information are presented, we have employed subject-matter experts from fields that relate to the nature of information contained in the interface. We believe that this approach is preferred over relying on common sense and we agree with Jones et al. (1999) in that "...what is common sense to an experienced pilot is quite different from the common sense of an AI researcher" (p. 8).

2.2 The Situation Awareness Panel

The Situation Awareness Panel is a graphical tool that enables the user of the JSAF simulation to observe a Soar agents' understanding of a situation, their goals, and their history (Jones, 1999). It is, in essence, a window into the Soar agent's awareness of the current situation. Figure 1 shows a screenshot of the SAP of an agent named *Shooter2*.

The panel is useful for examining and verifying the behavior of the Soar agent. With the SAP, the user is able to observe the awareness of an model and do things like examining whether the model's behavior reflects its understanding of the situation or its intentions for action, evaluating the model's current actions within the context of its history, and so on.

This version of the SAP is realized in Tcl/Tk. Tcl/Tk is an extension language (Ousterhout, 1994).that is jointly compiled with Soar. The SAP interacts with Soar agents running in a variety of JSAF scenarios. Any of the Soar agents can be explored with it, and there is nothing to preclude it being applicable to other Soar agents, although the plan-view display would not be useful for many of them.

Detail about the functions of the various displays of the SAP is provided by Jones (1999). In short, the SAP is divided into four displays:

- (a) The **Active Goal Display** is located at the top left part of the SAP and it contains the model's current stack of goals and subgoals. By enabling the user to observe what the model is trying to achieve at any moment, a comparison of the internal intentions of the model and its external behavior can be made.

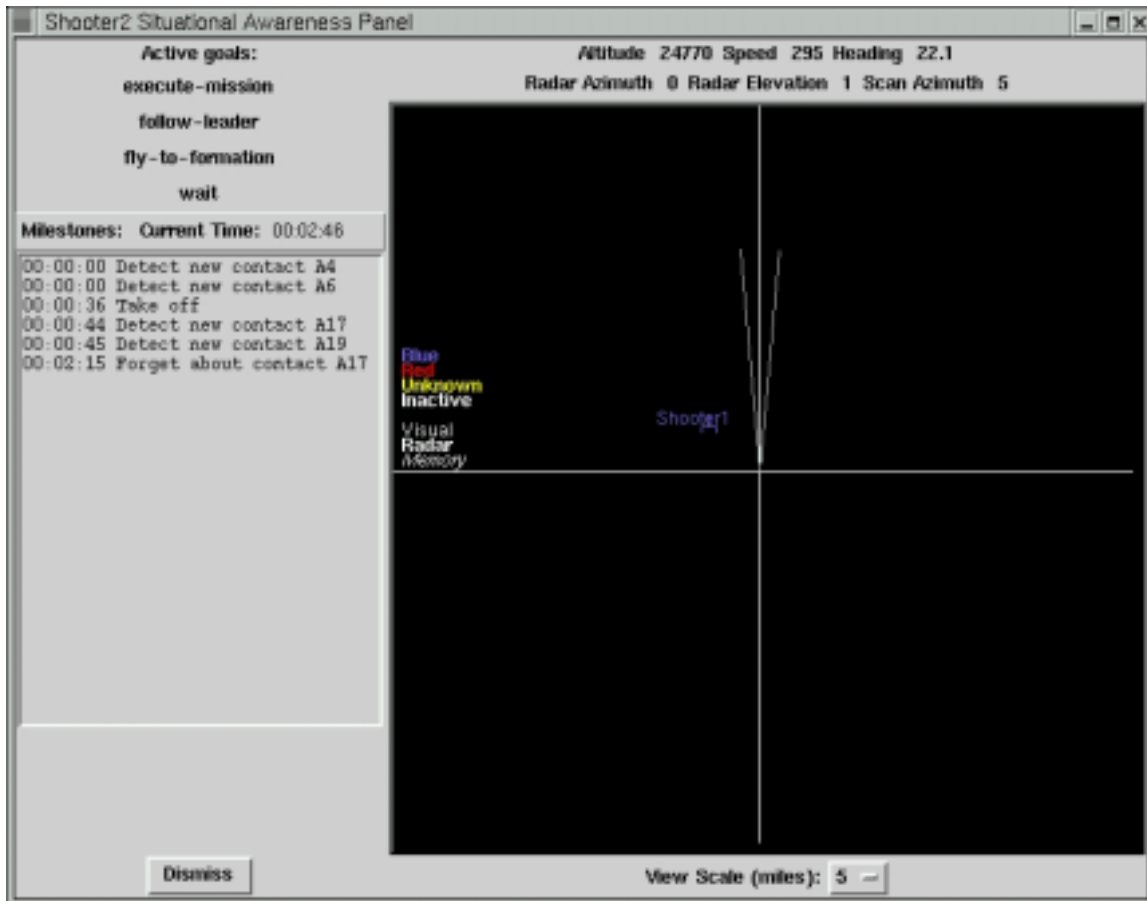


Figure 1. A screenshot of the SAP.

- (b) The Milestone Display is located underneath the Active Goal Display. Each milestone event is recorded as a new line in the window. A time stamp for each milestone event is also recorded. This display enables the user to review quickly the model's past activity and reasoning.
- (c) The Aircraft Status Display is located at the top of control panel and stretches to the right. It is a short strip that provides some basic aircraft information that is available to the model. The Altitude, Speed, Heading, Radar Azimuth, and Radar Elevation are displayed in this strip.
- (d) The Agent Awareness Display occupies the rest of the SAP. This display enables the user to inspect the current state of the model's awareness. It is basically a view of the model's reasoning about what is going on in its world (which is not necessarily an accurate depiction of what is really going on). Entities with which the agent had contact (through vision, radar, or radio) are all represented in the display and marked with different colors

to indicate whether they were friendly, hostile, unknown, or inactive. The type of contact is represented by different type styles. The user can adjust the scale of the Agent Awareness Display by choosing a different number from the "View Scale" drop-down menu.

2.3 The study

2.3.1 Participants

The participants were twelve experts coming from different disciplines. Table 1 lists the area of expertise of our participants. The first eight participants completed the study in our laboratory, while the last four were did so under the same equipment but on the site of their employment. All participants received monetary reimbursement in exchange for their participation. Participants were run individually with each experimental session lasting between an hour and two hours. The last four participants also observed a new and improved version of the SAP that became available to us shortly before the time they were run. Appendix B presents a screenshot of the new SAP.

2.3.2 Material and equipment

The JSAF simulation environment was presented on 19-inch monitor attached to a Dell Optiplex computer running Red Hat Linux 6.1. All experimental sessions were videotaped with the use of a SONY TRV-120 Hi-8 digital camcorder. In addition, the computer desktop activity of the first eight subjects was videotaped on VHS tape. Participants read and signed an informed consent form prior to the beginning, and they were debriefed upon completion of the study.

2.3.3 Procedure

Each experimental session started by providing the participant with a short description of the SAP taken from Avraamides's manual (Avraamides, 2001) and a description of the scenario within the JSAF environment that would executed by the models. Soar Technology provided us with three pre-programmed scenarios, from which we selected the *Defensive Counter Air (DCA)* scenario for our testing purposes. The description of the DCA scenario that subjects read before the study is presented in Figure 2.

Table 1. List of expert participants.

	Area of expertise
1	Plan view/geographic information systems specialist.
2	Graduate student in AI and cognitive modeling.
3	Marine Major, specializing in logistics and infantry.
4	Former software developer in Silicon Valley with Fortune 100 companies.
5	Former merchant marine officer and expert on social and group processes.
6	Navy fixed and rotary wing pilot. RWA instructor.
7	Cognitive psychologist.
8	Cognitive psychologist with some amateur flying experience.
9	Former military aviator from BMH Associates.
10	Former military aviator from BMH Associates.
11	Former military aviator from BMH Associates.
12	Former military aviator from BMH Associates.

DEFENSIVE COUNTER AIR (DCA)

The Defensive Counter Air mission involves defending an area against airborne threats. An Airborne Early Warning (AEW) aircraft is used for its long-range radar to watch for distant threats. When threats arise, the AEW dispatches an airborne 2-ship flying a Combat Air Patrol (CAP) to engage the bogeys.

Of Interest:

-The fighters's Situational Awareness Panel (SAP) will demonstrate the agent's attention to the oncoming air threats.

-Coordination between AEW and fighters.

Figure 2. Screenshot of the instructions provided to participants.

The study was divided into two parts. The goal of the first part was to familiarize our subjects with JSAF. We therefore asked them to perform a number of actions such as the ones shown in Table 2. This part lasted longer for the first eight participants because they were completely unfamiliar with JSAF.

Table 2. Example activities for familiarization with JSAF.

- As soon as the Plan View Display (see Appendix [of the instructions] for a screenshot) becomes visible, zoom into the map and locate the position of the AEW. Please report the approximate longitude and latitude of the AEW.
- When the Agent Window's appear select "1:2,000,000" from the Scale menu of the Plan View Display. Report the names of the agents you can see.

These simple tasks were introduced in order to guide our subjects to explore the various windows of the JSAF simulation and learn some important functions, for example, zooming into the map with either pressing simultaneously the two mouse buttons or using the Scale menu. We believe that some familiarity with the other windows of JSAF is needed in order to make possible the efficient use of the SAP. For example, inspecting the Plan View Display a user could determine which of the agents is more likely to have a target in its awareness, and then use the SAP to examine this feature.

The second part of the experiment had tasks that required that subjects observe the four displays of the SAP at the time the agents were engaged in combat as defined by the on-going scenario. Some example tasks are presented in Table 3.

As can be seen from the list of tasks in Table 3, we asked our subjects a variety of questions that differed in terms of both what they were required to do and what aspect of the SAP was brought into focus. Some of the questions required that subjects simply observed what was going on and others required some interaction of the subject with the interface. Observation questions were aimed at assessing whether the SAP was successful at conveying the information that it was supposed to convey. We were primarily interested in seeing whether our subjects could easily pick up from each panel of the SAP the information that the SAP was meant to provide.

Table 3. Example tasks used for exploring and interacting with the SAP.

- Identify the two aircraft that are dispatched by the AEW to engage the enemy and bring up the SAP of one of them. What do you think that each part of the SAP does?
- What is the altitude and speed of the agent that you are inspecting? Is the agent currently ascending or descending?
- What friendly agents are in your agent's awareness? What is their status?
- Are any enemy aircraft present in the Agent Awareness Display? If not, go back to the map to inspect the positions of the enemy aircraft. Bring up the SAP of the agent that you think is the most likely to be aware of the enemy
- If the enemy is within awareness, is it also within the Radar range of the agent?
- Are there any inactive or unknown agents present in the awareness of the agent you are inspecting. If yes, can you identify those agents from the Plan View Display?
- Go to the Plan View Display, go to the "On order" menu and select "MIG29 FWA Sweep Base". Find the agent that you think the enemy would appear in its SAP and report its most active goal.
- What is the most recent history event of the agent you are inspecting?
- Close the SAP.

Difficulties with and sometimes misunderstandings of information were of particular interest since they provide points to consider for future implementations of the SAP. The feedback we got from our subjects enabled us to determine ways of improving the visual interface of the SAP in future designs. Questions that required the subjects to interact with the simulation in order to initiate some action (i.e., "MIG29 FWA Sweep Base") provided a way for our subjects to evaluate the reasoning and the behavior of the agent, while it was engaging in action to achieve or prevent a user-initiated goal. The primary focus of the study was to determine the success of the SAP interface at revealing to users the reasoning of the agents and pinpoint their limitations, particularly in assessing whether the reasoning of the Soar models was realistic.

2.4 Analysis and Results

All videotapes were reviewed at a later time by the experimenters and a list of potential problems with the SAP interface was generated. The problems noted along with feedback from our subjects on how to deal with them enabled us to generate a set of suggestions for the improvement of the interface. Because our study focused on providing feedback for the improvement of the SAP, the present paper does not address any of the positive feedback received from our participants, which was substantial.

A total of 35 problematic issues for the SAP display were identified by our experts. The table in Appendix A presents a list of the main problems that were identified and the suggestions we were able to provide for overcoming them. Suggestions are subjective to the experimenters but they depend wholly on feedback obtained from our subjects.

Appendix C presents feedback we received about the new version of the SAP. Eleven different problems were noted by the four participants who observed the new SAP.

Table 4 reports the disposition of these two sets of suggestions. The totals do not equal the number of suggestions because some suggestions were put into two categories, particularly where the suggestion included multiple aspects. About half of the suggestions have been (done) or are being considered to be included (possible) into the revised SAP. Some are unlikely to be done (unlikely), and several of the suggestions represent problems in the old SAP that have been avoided rather than fixed in the revised SAP. There are also several of the more complex suggestions that remain under discussion.

Table 4. Disposition of the usability problems found for the original and revised SAPs.

	Original SAP	Revised SAP
Done	5	3
Possible and/or planned	11	6
Unlikely to be done	5	0
No longer applicable	10	0
Under discussion	6	2

Table 4. Set of tasks found in this analysis method.

Perception (Inputs) - What inputs does the model get?

Inputs does the model get from instruments
Radar and IFF values (if from display), and visual input
Voice input/communication from other agents
Other perceptual events
Constants in perception, e.g., due north
Self-perception, physical status of pilot: healthy, tired, bored
Where is our agent's attention (for analyst) -- perhaps with a spotlight metaphor (this was used by Chong in the AMBR project to good reviews)

Actions (Outputs) - What actions has the model done?

What plane/pilot/RIO has said and done
details of those actions if complex
What milestones are there, and what's the range of types of milestones, i.e., what could have been there but are not, and thus why are they not?

Physical environment features that affect the agent's body

Weather
Terrain, including base location, feet wet/dry, ground threats, places to land for RWA
Unknown but suspected ground threats will be an interesting thing to display

Mental environment -- Current Goals and Active Plans

Active goals, and their current status
Inactive goals, and why inactive (complete list of all possible goals and plans, and their status)
Details of the goals
Remaining steps in a goal/plan with associated physical location
Distance to target or other key events that agent would keep in attention and update, such as time left on CAP (thus timing)
Long term memory contents and active elements
Structure of memory and other mental objects
Contents of short term memory
Contents of perceptual (iconic) memory
Capacity remaining in each capacity, e.g., working memory, idle (slack) time in central processor.
Pop-up display of changes/targets of goals for turning, climbing, accelerating, i.e., when the plane starts to do any of these, a pop-up window appears over or in the SAP indicating what is being attempted
An articulate model that comment on its behavior
what other operators were available
why operators were or were not chosen (cf. Lewis Johnson's work)

Social environment

Cultural/political/historical facts that influence behavior (declarative facts)
Rules of engagement (perhaps available but not displayed if they don't change often)
Other social context of team, broadly defined

Mental models of other agents

(actual vs. mental may indeed be different)
x, y, z, heading, roll, yaw, pitch, speed, weapons
dx, dy, dz, d(heading),d(roll, yaw, pitch, speed)
Model of what the other planes are and what they are doing and what they are going to do
(this list repeated one level down based on what they think you are going to do!)
What other planes have said
What other planes have been told, perhaps from a specific range of time
What other planes can see (their radar might not be as good, and you might know it)
Physical status of other plane, damaged or not, fuel status, munitions, etc.
Physical status of other plane pilot, healthy, tired, bored

Military environment (task and hardware of own agent)

Written instructions
x, y, z, heading, roll, yaw, pitch, speed, weapons
dx, dy, dz, d(heading),d(roll, yaw, pitch, speed)
What other planes have said to agent
Physical status of own plane, damaged or not, fuel status, munitions, etc.
Physical status of pilot, healthy, tired, bored
Munitions capabilities if novel (otherwise, assumed or reconstructed), and range

Based on the feedback we got from our participants and our experience building interfaces for cognitive models (Ritter et al., 1998; Ritter & Larkin, 1994) we created the task-analysis shown in Table 4. We believe this task analysis can be used as a guide for designing interfaces for cognitive models of military content. These tasks include what all users need to know to understand their models, so interfaces that supported these tasks would also be useful more generally.

This task analysis includes tasks that would be expected. That the Perceptions and Actions of the model are needed by analysts will not be a surprise to many modelers. Likewise, access to the Mental environment of the model should not be surprising, but this is not fully supported by every modeling environment. Similarly, because the models are increasingly becoming embodied and subject to their environment, the modelers need to know what aspects of the environment influence a model.

Including the Social environment of the model explicitly is somewhat novel. These agents clearly have social aspects to their behavior and reasoning. This appears to be a different type of knowledge and processing, a type that interfaces should make available to modelers.

The Mental models of other agents is of increasing importance for cognitive models as they become team members, and this is particularly true for models in synthetic environments that need to understand colleagues and advisories. Finally, the model and the modeler need to keep in mind aspects of the environment related to their specific task. In the case of these models, the domain is a military one. Other models are likely to require additional information related to their domain of performance.

3. Conclusions

Using cognitive models as surrogate users in military simulations provides the capability of training military personnel even individually. In a JSAF simulation, thousands of entities can be represented as computer-generated forces providing the feel of a realistic environment without the need to recruit great numbers of humans to participate in the simulation. The success of simulators depends greatly on how realistic the behavior of the cognitive models is.

Graphical interfaces that make the behavior of these models visible to human users enable the validation and the improvement of the models. So far, not much emphasis has been placed on the design of graphical interfaces for cognitive models. Even when graphical interfaces have been supplied along with models, they have been designed based on the “common sense” of their developer.

We believe that graphical interfaces are very important for cognitive models. By making their internal state more visible, graphical interfaces allow a better understanding of the reasoning and actions of the model and therefore lead to easier debugging, better validation, and more powerful demonstrations of the models.

Given the importance of graphical interfaces, we argue that they should undergo rigorous testing to assess their usability. The present study provides an example of how this can be done. The provided task-analysis can be used to guide the design of future modeling interfaces. Keeping and extending this task analysis will help design better interfaces so that they support the user performing their tasks.

Acknowledgments

We need to acknowledge the help of our subject matter experts who provided the feedback that we summarize. The experts at BMH were particularly helpful at a distance. Glenn Taylor and others at Soar Technology provided invaluable help in installing and maintaining the SAP for our use. Comments from Angie Barnhill have improved this report.

4. References

- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4(2), 167-207.
- Avraamides, M. N. (2001). A brief manual for running a JSAF Demo and examining the Situation Awareness Panel (Tech. Note No. 2001-1). Applied Cognitive Science Lab, School of Information Sciences and Technology, Penn State. [available upon request].
- Jones, R. M. (1999). Graphic visualization of situation awareness and mental state for intelligent computer-generated forces. In *Proceedings of the Eighth Conference on Computer Generated Forces and Behavioral Representations*. 219-222. Orlando, FL: Division of Continuing Education, University of Central Florida.
- Jones, R. M., Laird, J. E., Nielsen, P. E., Coulter, K. J., Kenny, P., & Koss, F. V. (1999). Automated intelligent pilots for combat flight simulation. *AI Magazine*, 20(1), 27-41.
- Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, 33(1), 1-64.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Ousterhout, J. K. (1994). *Tcl and the Tk toolkit*. Reading, MA: Addison-Wesley.
- Pew, R. W., & Mavor, A. S. (Eds.). (1998). *Modeling human and organizational behavior: Application to military simulations*. Washington, DC: National Academy Press. books.nap.edu/catalog/6173.html.
- Polson, P. G., Lewis, C., Rieman, J., & Wharton, C. (1992). Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies*, 36, 741-773.
- Ritter, F. E., Jones, R. M., & Baxter, G. D. (1998). Reusable models and graphical interfaces: Realising the potential of a unified theory of cognition. In U. Schmid, J. Krems, & F. Wysotzki (Eds.), *Mind modeling - A cognitive science approach to reasoning, learning and discovery*. 83-109. Lengerich, Germany: Pabst Scientific Publishing.
- Ritter, F. E., & Larkin, J. H. (1994). Using process models to summarize sequences of human actions. *Human-Computer Interaction*, 9(3), 345-383.
- Ritter, F. E., Shadbolt, N. R., Elliman, D., Young, R., Gobet, F., & Baxter, G. D. (in press). *Techniques for modeling human performance in synthetic environments: A supplementary review*. Wright-Patterson Air Force Base, OH: Human Systems Information Analysis Center.

- Ritter, F. E., & Young, R. M. (2001). Embodied models as simulated users: Introduction to this special issue on using cognitive models to improve interface design. *International Journal of Human-Computer Studies*, 55, 1-14.
- Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (Eds.). (2000). *Cognitive task analysis*. Mahwah, NJ: Lawrence Erlbaum.
- SigArt ACM Special Interest Group on Artificial Intelligence (1989). Knowledge Acquisition. *Sigart Bulletin*(Special issue).
- St. Amant, R. (2000). Interface agents as surrogate users. *intelligence*, 11(Summer 2000), 29-38.
- Tambe, M., Johnson, W. L., Jones, R. M., Koss, F., Laird, J. E., Rosenbloom, P. S., & Schwamb, K. (1995). Intelligent agents for interactive simulation environments. *AI Magazine*, 16(1), 15-40.

Appendix A. Feedback on the initial Situation Awareness Panel

	Subjects who identified problem	Problem	Suggestion for dealing with problem
1.	1, 2, 4, 6, 7, 8, 10, 12	Labels for the agents in the Agent awareness display are not very clear.	Use different fonts, or font size, or icons to distinguish the status of each agent.
2.	10	Labels take much space in the Agent Awareness Display.	Labels do not need to be visible all the time. Provide the user with the opportunity to hide them.
3.	9	Label “ukn” is not needed.	The symbol used (square with a speed vector) is used in cockpits to represent unknown entities. Use the symbol of unknown and a different one for known.
4.	All subjects	SAP resizes as more goals are added on the Active Goal panel.	Use scrolling windows for the Active Goal panel and the History panel so that a fixed number of goals/events is visible at any point, while the user can use the scroll bar to examine any goal or event.
5.	1, 4, 6, 10	The most recent goal/events are added on the bottom of the Active Goal/History panel. This is counterintuitive given that in goal-stack most recent items are added on the top.	Reverse the order of events in the Active Goal and the History panels so that most recent events are added on the top. Do this along with adding scroll-bars so that users scroll down to view previous goals and history events.
6.	1, 6, 7	The line-vectors that accompany the agent symbols in the Agent Awareness display are not very good indicators of heading.	Replace lines with arrows that will mark heading more clearly. Manipulate the length of the arrow to provide a visual indicator of the speed of movement of the agents.

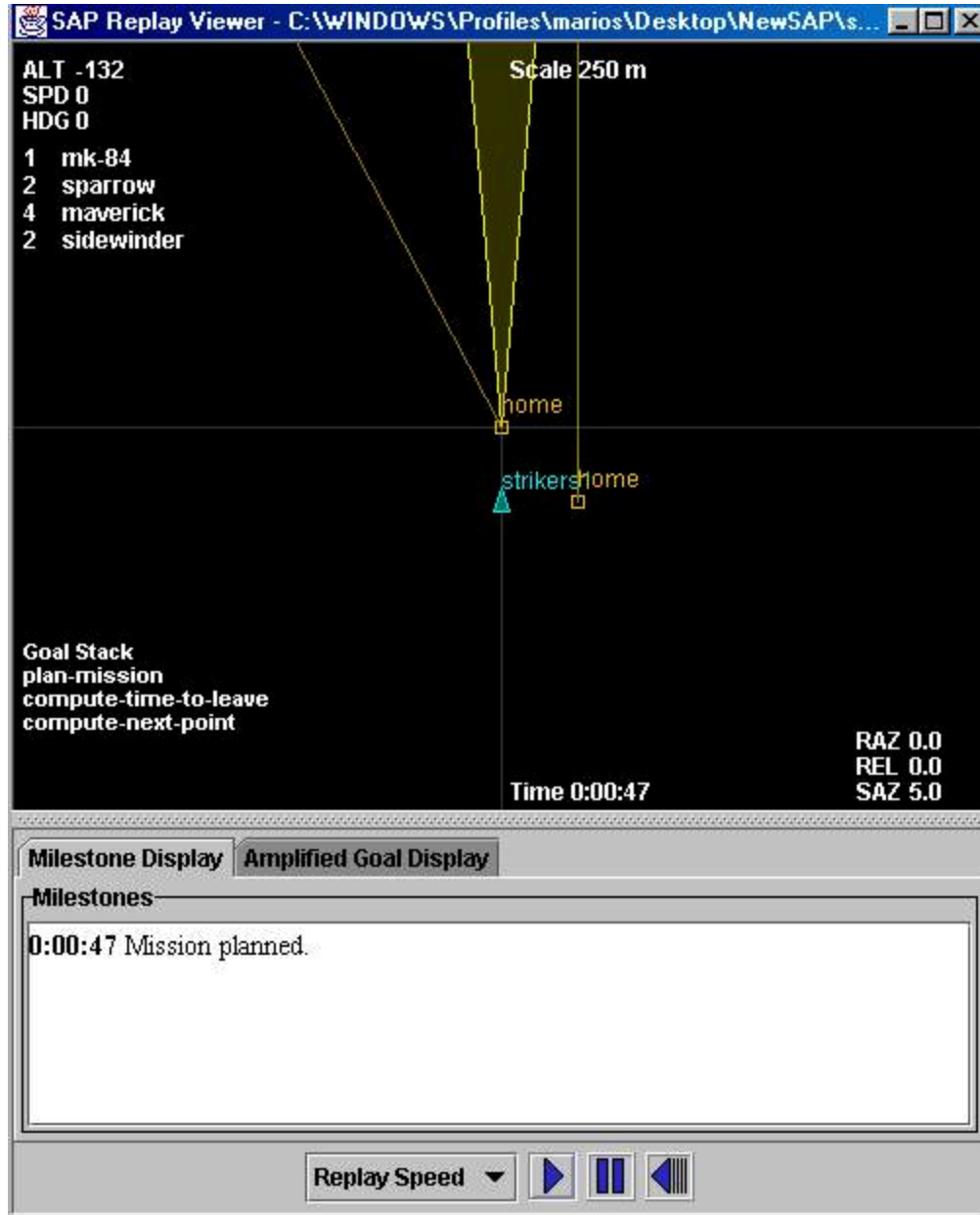
7.	1, 9, 10, 12	No ascending/descending information available to the awareness of the agent. (It can be inferred from the changes of the altitude reading).	Put an arrow next to the altitude reading to denote whether the plane is ascending or descending. Use a visual speed indicator (VIS) to convey change of altitude. Knowing whether the plane is ascending or descending is critical for fuel purposes.
8.	1	There is no information about how far other agents are.	Add a number next to the symbol of the agent to indicate how far it is. Alternatively, use some other symbol to indicate distance on a pre-defined scale (i.e., near, very near etc.)
9.	2, 4, 7, 8, 9,	Problems with radar lines, e.g., they don't move and they don't resize with scale changes.	The radar range should rotate (we think that there's a glitch in the program that prevents this) and also resize with changes of the scale to clearly indicate how far the range goes. Also, fill the radar range with a transparent texture to make it easier for users to assess whether something falls within the radar or not.
10.	10, 11, 12	Radar range is not explicit.	Put numbers on the display to indicate radar range.
11.	10	The scale drop-down box is not presented at a convenient location.	Put it above the Agent Awareness Display. This box is used very frequently by operators and it should be at the most intuitive location in the panel.
12.	12	The scale drop-down box should be a slider instead.	Make it a slider with buttons for most frequent settings (just like how simulation speed is set on an agent's window).
13.	9, 10, 11, 12	The heading measurement does not make sense. What is the meaning of a negative value for heading?	Make heading measurement realistic (i.e., implement it the way is done in real cockpits). Maybe measure it from the "nose" of the aircraft and label it "bearing".

14.	11	Displayed measurement do not correlate with parameters in real cockpits, e.g., F14's used in DCA do not have scan azimuth of 5.	Make sure that measurements are realistic.
15.	4, 6, 7,9, 10, 11, 12	There are no environmental or other ground features that are important for the mission in the awareness of the agent. No important military features either.	Add some important ground features in the SAP e.g., relevant ground threats when the agent is flying in their firing range, friendly ground units, places to land in case of emergency etc.
16.	4, 5	No information about the pilot and his/her wingman are represented.	Add information (preferably in visual form) that indicates whether the pilot/wingman is stressed, scared, fatigued and his/her level of expertise (e.g., mission history).
17.	3	No information about logistics is in the awareness of the agent.	Add some important logistics information e.g., how much gas is left and when and where the plane can be refueled.
18.	9, 10	No information about weapon status.	Indicate type and quantity of weapons available.
19.	6	The rules of engagement are not present in the agent's awareness.	Add such information e.g., agent has clearance to fire?
20.	6	The legend for the status of the agents in the Agent Awareness Display is located at an intrusive place in the panel.	Put the legend on the left bottom corner of the panel.
21.	7	Differentiate the information that falls within the focus of attention of the agent from the rest of the information that is present in the awareness of the agent.	Implement attention/working memory as a spotlight that highlights some information as more accessible than the rest.
22.	11	Differentiate awareness from reality.	Implement a "true" mode that shows what information is present in the world so that it can be contrasted with the awareness of the agent.

23.	5	There is no indication of how much of the available cognitive resources are consumed at a current point in time.	Accompany the above implementation of attention/working memory, with a percentage to indicate the proportion of engaged cognitive resources.
24.	5	Information is not presented in the modality in which it is presented in real combat, e.g., contacts in JSAF are seen instead of heard.	If audio can be added to future implementations of JSAF, then present contacts auditorily than visually to better match real-combat situations.
25.	10	Cockpit measurements are not presented the way pilot see them.	Operators with flying experience are used to see cockpit measurements as data boxes. Use target, system, and weapon-status data boxes to present measurements.
26.	11	No mission parameters available.	Present some mission parameters e.g., what the cappoint is, how long the agent has remained on station, fuel state, weapon status. Present these in data boxes.
27.	4	The label “milestones” assumes that events are preplanned.	Change the “milestones” text to “events”.
28.	2	The “Dismiss” button is not at a very intuitive location in the panel.	Put the “Dismiss” button on the bottom right corner of the SAP. A rearrangement of buttons and panels within the SAP needs to take place in order to minimize empty space.
29.	4	The label “milestones” assumes that events are preplanned.	Change the “milestones” text to “events”.
30.	10	Too many milestone events.	Present only the last 3-4 events.
31.	1, 12	More information should be provided for each goal in the Active Goal Display.	List the means for achieving the goal when the user clicks on the goal or puts the mouse on it for a few seconds (e.g., as a tooltip).

32.	10, 11, 12	Information provided in the Active Goal Display is not useful for operators. The A6, A21 etc. signs that are used are not meaningful.	Use the code signs for displaying goals. Change the content of the goals to include information such as the mission, the Barcap point, the altitude the agent is trying to achieve etc.
33.	7	No information about the shooting range of the plane is provided.	Add on the Aircraft Status Display information about the weapons carried (type and quantity), their range and maybe the agent's rules of thumb for using them.
34.	7	Cannot move from the SAP of one agent to the SAP of some other agent without bringing up other windows	Add shortcut buttons for easy user-navigation among all SAP's.
35.	11, 12	Some types of information are only relevant to certain users. (i.e., operators do not need that active goal display or the milestone events).	Implement different SAP modes (e.g., operator mode, programmer mode) in which the information presented is tailored to the needs of the user. Design a Global SAP in which operators can inspect all SAP's at once and select the ones they wish to examine in greater detail.

Appendix B. The new Situation Awareness Panel



Appendix C. Feedback on the new Situation Awareness Panel

	Subjects who identified problem	Problem	Suggestion for dealing with problem
1.	9	Not clear how the scale is measured. Is it from the center of the display?	Give more information for elements on the SAP. Maybe in the form of tooltips.
2.	9	What is the meaning of a negative heading?	Same as above. Provide more information on how measurements are made.
3.	12	Not clear what the line connecting the two points is. Is it the route?	Same as above.
4.	9, 10	Goal stack is not needed. It is distracting and it does not contain useful information for operators.	Provide the option to hide the goal stack.
5.	9	A target still carries the "ukn" label even after it has been identified as enemy.	Change the label when an unknown target is identified (e.g., to MIG).
6.	11	Miles and kilometers on the scale don't match; a 60-mile range should be about equal with a 100-km range.	Fix it.
7.	11	Some critical information is not included e.g., fuel state, barcap point, the route that the agent is flying.	Provide some more information maybe in the form of data boxes than can be turned on and off (to avoid cluttering the display)
8.	12	Time remaining before landing is not provided.	Provide this or the fuel state instead.
9.	12	The scale cannot be adjusted easily (it is hidden in the menus).	Put a slider to provide ease with the change of scale.

10.	12	Not clear what the reference frame of the SAP is. Is North represented up all the time? Is it a god's eye-view? Would it better to show awareness from the agent's egocentric frame?	Provide options for selecting a reference frame (i.e., change between allocentric and egocentric).
11.	12	No information about other agents (including the enemy).	Provide information about other agent's speed, heading, altitude etc. Such information becomes available to pilots in combat.