

Three Simulation Models of Naval Air Defense

LTJG Baris Ozkan, Neil C. Rowe, LT Sharif H. Calfee, and John E. Hiles
Institute for Modeling, Virtual Environments, and Simulation (MOVES)
and Computer Science Department

U.S. Naval Postgraduate School
833 Dyer Road, Monterey, CA 93943

begemenozkan@hotmail.com, ncrowe@nps.edu, scalfee@hotmail.com, jehiles@nps.edu

ABSTRACT

Naval air defense is a critical facility for ship survivability and has been subject of a number of studies. We investigate here three different approaches to modeling it with computer software. One approach focuses on the problems of information processing and communication for the air-defense team, and is good for analyzing its efficiency. Another approach focuses on the inference of the nature of observed tracks, and is based on the novel psychological theory of conceptual blending. A third approach uses an expert-systems approach that can learn from experience and be more substantially automated than the other two approaches. Each approach has its own advantages: The first provides insights for organizing and managing air-defense personnel, the second provides insights into cognitive biases in analysis that should be examined during training, and the third suggests a way to mostly automate the air-defense process to save money. This work suggests the value of multiple simulations of the same process when that process is important to understand.

BACKGROUND

Naval air warfare is rapid and serious, and is a major focus of the operations of ships (Mairorano, Carr, and Bender, 1996). Air-defense teams of "watchstanders" must train extensively in search, detection, and classification of all aircraft and surface vessels within the operational area of a ship's "battle group" (coordinated group of ships). Actions taken are constrained by strict "rules of engagement". However, continued advances in speed, maneuverability, and accuracy of anti-ship missiles mean watchstanders can get confused about the identity of tracks or be unable to communicate and react quickly and correctly. Two incidents in the 1980s highlighted this: The USS *Stark* was attacked by anti-ship missiles, and the USS *Vincennes* mistakenly shot down a civilian airliner. Air defense does require human judgment because of widely varying geographical, environmental, and tactical conditions. However, it can certainly be aided by automated decision aids that summarize information (Noh & Gmytrasiewicz, 1998). And good simulations help in training air-defense personnel since exercises with real ships, aircraft, and weapons are very expensive.

In a U.S. Navy battle group, the Air-Defense Commander is responsible for the coordination of the force's ships and aircraft. This includes surveillance, detection, identification, intercept, and engagement of aircraft within the operational area ("contacts") with the primary objective of defending the group's high-value unit such as an aircraft carrier. Air defense is done in the Combat Information Center (CIC), which contains consoles for activating weapon systems, configuring sensor systems, displaying contact tracks, and communicating with other ships and aircraft.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE JUN 2005	2. REPORT TYPE	3. DATES COVERED 00-00-2005 to 00-00-2005	
4. TITLE AND SUBTITLE Three Simulation Models of Naval Air Defense		5a. CONTRACT NUMBER	
6. AUTHOR(S)		5b. GRANT NUMBER	
		5c. PROGRAM ELEMENT NUMBER	
		5d. PROJECT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School, 833 Dyer Road, Monterey, CA, 93943		5e. TASK NUMBER	
		5f. WORK UNIT NUMBER	
		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)	
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited			
13. SUPPLEMENTARY NOTES The original document contains color images.			
14. ABSTRACT			
15. SUBJECT TERMS			
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified	
			18. NUMBER OF PAGES 44
			19a. NAME OF RESPONSIBLE PERSON

RELATED WORK

A number of computer games such as the Harpoon series have simulated naval air defense, and some are quite realistic in depicting the environment and the effects of weapons. However, their primary concern is entertainment, and their psychological modeling is simple.

The Tactical Decision-Making under Stress study explored the causes of the USS *Vincennes* incident (Morrison et al, 1996). Some problems were identified with the short-term memory limitations, such as forgetting and confusing track numbers, forgetting and confusing kinematic data, and confusing tracks of contacts. Other problems were related to decision bias, such as carrying initial threat assessment throughout the scenario regardless of new information, and assessing a contact from past experiences. This work also suggested how to improve command center display consoles.

Other work examined the cognitive aspects of the threat-assessment process used by naval air-defense officers during battle group operations (Liebhaber and Smith 2000). This indicated that watchstanders had possible-track templates, derived from a set of twenty-two identifying factors, which they used to classify contacts and calculate threat assessments. The most important factors were observed to be signal emissions, course, speed, altitude, point of origin, Identification Friend or Foe (IFF) responses, flight profile, intelligence information, and distance from the detector. Each factor had a range of values associated with each category of Friendly, Neutral, and Hostile. This research was very helpful in developing all three of our simulations.

(Liebhaber & Feher, 2002) built a cognitive model of "cues" in air defense. This work showed that users definitely created templates to define which cues will be evaluated and the permissible range of data for each cue. Cues were evaluated in a fairly consistent order, weighted, and processed in sets reflecting their weights. Air-defense threat evaluators relied only the data associated with cues in their active template, did not change templates in the face of conflicting data, and were influenced by conflicting data in specific cues rather than in the overall pattern. Situational awareness in particular was identified as a primary concern during task analysis for the Joint Maritime Command Information System (Eddy, Kribs, and Cohen, 1999). It was affected by (1) capabilities, (2) training and experience, (3) preconceptions and objectives, (4) and ongoing task workload. This work suggested that watchstanders lost site of the "big picture" as task workload increased.

(Amori, 1992) presented a plan recognition system for airborne threats that does three-dimensional spatial and temporal reasoning. It exploited physical data and changes to this data, known air tactics and behaviors, and likely primary and secondary goals. One module analyzes the data associated with each track while another module analyzes coordinated activity between tracks. Reasoning is done both "backward" (about observed tracks) and "forward" (predicting the future). (Delaney, 2001) developed a system for more effective coordination of air-defense planning and execution for multi-service (i.e. Army, Air Force, Navy, and Marines) and international operations. It develops and executes a theater-wide air-defense plan providing an integrated view of the battlespace. Its focus is threats, not the personnel responding to them. Other work on classification of contacts in air defense (Barcio et al, 1995; Bloeman and Witberg, 2000; Choi and Wijesekera, 2000) explored a variety of mathematical techniques.

(Osga et al, 2001) built specialized watchstation consoles with improved human-computer interface designs. Research conducted extensive interviews and console evaluations with air-defense subject matter experts. The consoles corrected interface problems in the current AEGIS consoles which caused errors, information overload, and loss of situational awareness. These consoles reduced the needed size of the air-defense team by two or three people while increasing performance.

As for air-defense training, the Battle Force Tactical Trainer System (Federation of American Scientists, 2003) was designed for the fleet-wide training of naval units by providing each ship with a system using the existing CIC console architecture. High-fidelity scenarios can inject actual signal information into the

ship combat systems to simulate reality. The system can simulate an entire fleet of ships and their staffs and support war-gaming exercises.

Air defense also can be viewed as a problem in data fusion, the collecting of data from multiple sources and combining them to achieve more accurate results than could be achieved from a single sensor alone. Techniques that have applied to data fusion problems include expert systems and blackboard-based distributed systems (Hall & Llinas, 1997). Data fusion can involve raw data, features extracted from data, or decisions made separately on each piece of data. The first is the most accurate but requires the most processing bandwidth; the last requires the least bandwidth but is the least accurate. Our simulations have generally used raw data because the total information available in air defense is not especially large.

THE ADC SIMULATION

Our first simulation was the AEGIS Cruiser Air-Defense (ADC) Simulation, discussed in more detail in (Calfee & Rowe, 2004). It models the operations of CIC watchstanders for a U.S. Navy battle group, using multi-agent system technology (Ferber, 1999). Conceived to assist training and doctrine formulation, the simulation provides insight into the factors (skills, experience, fatigue, aircraft numbers, weather, etc.) that influence performance, especially under intense or stressful situations. It simulates air contacts as well as the actions and mental processes of the watchstanders. All simulated events are logged to permit performance analysis and reconstruction for training.

The watchstanders modeled by the ADC Simulation were:

- The Force Tactical Action Officer who controls air defense for the battle group and is responsible for major decisions such as contact classifications and weapons releases.
- The Force Anti-Air Warfare Coordinator who coordinates the movement and assignment of friendly aircraft and orders the weapons employment by ships in a battle group.
- The Ship Anti-Air Warfare Coordinator who directs aircraft detection and classification for a single ship in support of the air-defense effort and manages the identification process.
- The Ship Tactical Action Officer who leads the CIC watch-team air-defense effort for a single ship.
- The Missile Systems Supervisor who fires (under orders) the ship's surface-to-air missiles and the self-defense Close-In Weapon System.
- The Red Crown watchstander who monitors friendly aircraft for the battle group.
- The Electronic Warfare Control Officer who is responsible for the operation of the electronic-emissions detection equipment.
- The Identification Supervisor who does Identification Friend or Foe (IFF) challenges on unknown aircraft and, when directed, initiates query or warning procedures for contacts.
- The Radar Systems Controller who operates the SPY-1A/B radar systems, the primary means by which aircraft are detected and tracked.
- The Tactical Information Coordinator who operates the Tactical Digital Information Links which communicate data among the ships and aircraft in the battle group.
- The Combat Systems Coordinator who monitors the status of the combat systems that support the CIC and repairs them as necessary.

To build the simulation, interviews were conducted with five air-defense subject-matter experts from the AEGIS Training and Readiness Center Detachment in San Diego, California, and the Fleet Technical Support Center Pacific. They had five to fifteen years of naval air-defense experience. The materials of the simulation (agents, objects, and attributes) were determined from these interviews, relationships between agents and between agents and objects were explicitly defined, and tasks and actions for each agent and object were defined. The simulation was then built from this information, debugged, and some minor adjustments were done.

The ADC Simulation is implemented in Java. The center of the screen (exemplified by Figure 1) shows the locations and statuses of the contacts; the Persian Gulf in 2002 was used for the prototype. Contacts are

CONTACTS AND AGENTS

Friendly aircraft are generated by orders of the Force TAO or Force AAWC agents; others are generated at random based on the attributes of Contact Density and Hostile Contact Level. Surface-to-air missile and anti-ship missile contacts are created by orders from either the CIC or hostile aircraft. Neutral aircraft fly directly between two points. The Friendly aircraft depart and return from Friendly bases and carriers, and conduct visual intercept, identification, and possible engagement of other aircraft. The Hostile aircraft start in either Iran or Iraq, and have varied flight profiles including reconnaissance, low-altitude behavior, and approach and attack profiles.

Aircraft and missile contacts have a track number, point of origin, course, speed, altitude, radar cross-section, electronic signal emissions (based on radar type), and IFF (Identify Friend or Foe) mode. Simulated contacts are created by the ADC simulation as it runs. All contacts fly piecewise-linear paths with landing and descent behavior at the ends. They may respond to queries and warnings from the CIC, and may retreat or alter their courses accordingly. They may also experience loss of their radar or IFF systems with a certain probability.

The simulated watchstanders were implemented as agents with personal attributes and ways to communicate with one another to coordinate tasks. Methodology for modeling teams of people has become increasingly sophisticated in recent years (Weaver et al 1995), and our approach exploited this. Each action has a probability of success based on the skill attribute of the watchstander: Basic (zero to six months experience), Experienced (six months to a year), and Expert (more than a year). Skill and experience are separate: Skill affects the probability of success of an action, and experience affects the time to complete it and the degree of confidence in its results. The experience levels were Newly Qualified, Experienced (10% faster), and Expert (20% faster). For watchstanders that evaluated contacts, an Evaluation Confidence attribute is increased for each contact. The initial value is 30 and is increased at a different rate for Newly Qualified, Experienced, and Expert levels (by 2, 4, and 6 respectively), to a maximum of 95 for Newly Qualified, 90 for Experienced, and 85 for Expert.

The fatigue attribute controls the readiness of the watchstander (Burr, Palinkas, & Banta, 1993). We used values of Fully Rested (having had a minimum of five hours of rest without having performed heavy physical labor or stood any watch), Tired (a minimum of three hours of sleep or at most six hours without rest in a fairly demanding environment), and Exhausted (less than three hours of sleep, or having performed heavy physical labor, or having performed duties over six hours in a demanding environment). Fatigue decreased the success probability for an action and increased the length of time to do it. Many of the interviewed experts also argued for a "decision-maker" attribute to reflect the differences among the watchstanders in how long they took to reach a decision. The values proposed were Cautious (for a maximum of 30 seconds), Balanced (20 seconds), and Aggressive (10 seconds), with a uniform distribution of times up to these maxima.

EQUIPMENT, ENVIRONMENTAL, AND DOCTRINE ATTRIBUTES

The simulation modeled seven key items of equipment (though simplifying their characteristics to avoid the use of classified information). Equipment had four readiness levels with associated probabilities of successful operation: Fully Operational (1.0), Partially Degraded (0.75), Highly Degraded (0.50), and Non-operational (0.0). In addition, if the user has activated the Scenario Equipment Failure Option, any of the systems could randomly fail during the scenario, requiring the watchstander agents to troubleshoot them until successful.

To model the SPY-1B radar system, receiver operating characteristics from Swerling II statistics were used (Alvarez-Vaquero 1996). Data was obtained from the AEGIS SPY-1B Radar Sphere Calibration Test Procedure of the Naval Sea Systems Command. The SLQ-32 radar detects and classifies electronic signals emitted by aircraft and shipboard radar systems. The Identification Friend or Foe (IFF) System recognizes friendly and neutral aircraft in five categories or modes. Link 11 and Link 16 rapidly disseminate

information about aircraft and ship contacts in the operational area, and maintain situational awareness about the battlespace.

In the simulation, surface-to-air missiles were simulated with a 0.70 probability of intercepting their target and a range of eighty nautical miles. To maintain realism, only two missiles can be launched against a target; if they fail, two additional missiles can be fired. The Close-In Weapons System is a twenty-millimeter shipboard self-defense system that contains its own radar and fire control system. In the simulation, it has a range of one nautical mile and a 0.50 probability of hitting its target.

The simulation has three options for the Environment attribute: clear weather, heavy rain, and heavy clutter. Their primary effect is on detection and communications systems. The Contact Density attribute controls the number of contacts (low, medium, or high). The Scenario Threat Level attribute (white, yellow, or red) affects classification of aircraft contacts: The higher the threat level, the more likely the team is to classify aircraft as Suspect or Hostile, and the more numerous and aggressive are the hostile contacts.

An AEGIS Doctrine defines additional procedures and situational parameters for the CIC. The simulation implemented just the Auto-Special Doctrine, a weapons doctrine used to reduce reaction time and human errors when a fast-moving anti-ship cruise missile is detected very close to the ship. It directs that the ship's combat systems will automatically engage the hostile missile with surface-to-air missiles.

WATCHSTANDER PROCEDURES

Air defense has three phases: (1) contact detection and reporting, (2) contact classification, and (3) action response. Each has assigned watchstanders and a defined flow of information (Figure 2). The simulation agents follow this plan of action. To model human processing limitations, there are input-message reception queues, message-priority processors, priority queues, action processors, and output-message transmission queues. Watchstander agents place order/request messages into other watchstander's input-message queues. 15 kinds of reports and 6 kinds of orders are handled in the implementation.

Watchstanders must prioritize contacts. Their criteria are newness of the contact, closeness to the ship, whether it is approaching, and duration since the last examination. But not all watchstander agents in the simulation rate these criteria the same way. Furthermore, watchstanders must periodically reevaluate the same contacts. To model the cognitive and decision-making aspects of contact classification, linear models were used that take a weighted sum of six numeric factors: closing course, speed, altitude, signal, origin, and mode. Different weights are used for each Scenario Threat Level (White, Yellow, and Red). Four thresholds are used to distinguish the conclusions Hostile, Suspect, Neutral, Unknown, and Friendly based on the advice of the experts as well as their actual usage by operating naval forces (see Table 1). Initial contacts are usually Unknown.

LOGGING

Five simulation logs are kept: the Scenario Events Log, the Watchstander Decision History Log, the CIC Equipment Status Log, the Watchstander Performance Log, and the Parse/Analyzer Log. The Scenario Events Log maintains a high-level record of all events. From the logs, values are calculated for average initial detection time of aircraft, average initial classification time of aircraft, and average correct classification time of aircraft. For the individual watchstanders, values are calculated for the number of errors, number of total actions attempted, percentage of errors in attempted actions, average action durations, and average communications time.

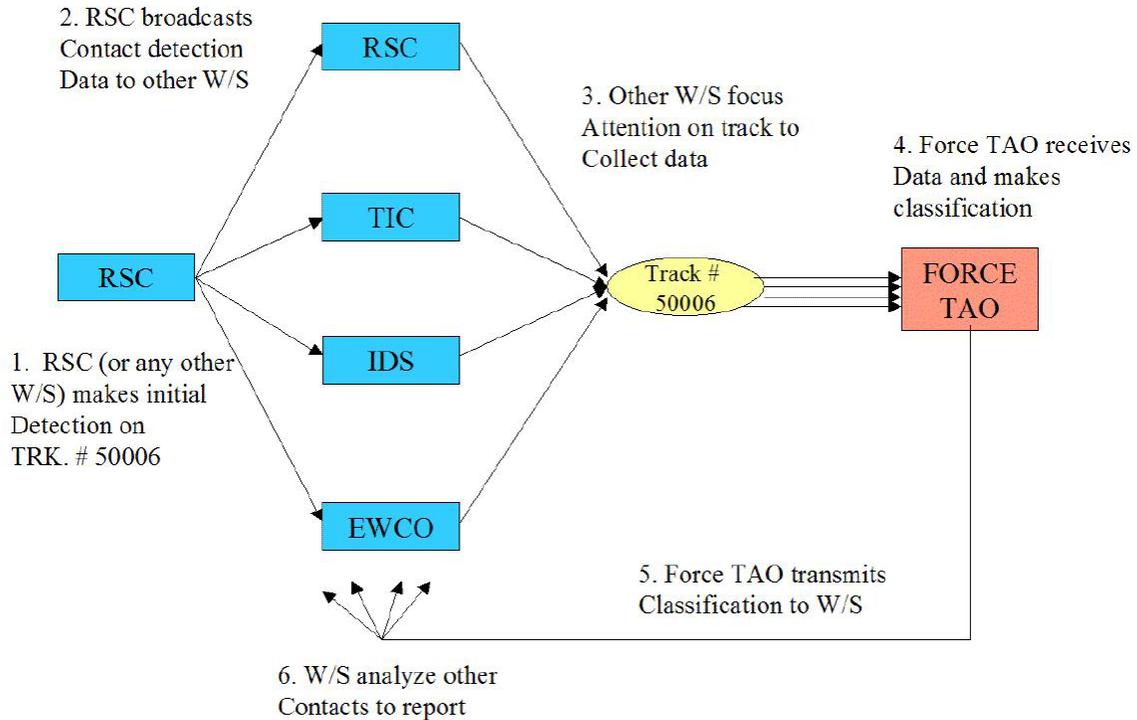


Figure 2: Communications between the watchstander agents.

Table 1: Default classification thresholds.

Contact Classification	Threat Level White Thresholds	Threat Level Yellow Thresholds	Threat Level Red Thresholds
Hostile	≥ 600	≥ 500	≥ 450
Suspect	500 to 599	450 to 499	400 to 449
Neutral	400 to 499	300 to 449	200 to 399
Unknown	-399 to 399	-399 to 301	-399 to 199
Friendly	≤ -400	≤ -400	≤ -400

EVALUATION OF THE ADC SIMULATION

We asked nine air-defense experts at the ATRC Detachment in San Diego to evaluate the simulation. On a scale of 1 to 5, where 1 meant "strongly disagree" and 5 meant "strongly agree", the experts were reasonably satisfied with the user interface (Table 2). They were also queried as to the realism of the simulation's variation in performance with key parameters on a scale of 1 to 7 (Table 3). For instance, one question (the upper left of the table) asked whether it was realistic for the Radar Systems Controller that performance time improved and the number of errors decreased when the Experience level increased. These survey results were also encouraging.

As for the predictive power of the simulation, five factors were selected as the focus of experiments: team skill, experience, fatigue, and the operational status of the SPY-1B radar. For each factor value, ten scenario runs were conducted, for 170 individual runs. The experts suggested that the most useful performance metrics were the duration of the actions of the watchstanders and their error rate. Unless otherwise indicated, tests assumed the watchstanders were Experienced in both the skill and experience factors, were fully rested, were balanced decision-makers, and had fully functional equipment. It was assumed that the external environment had medium contact density and threat level White; the hostile

contact number was low; and it was clear weather. To simplify comparisons between tests, kinematic attributes of the contacts were constant, starting locations were constant, destination points were the same for the same starting point, no new contacts were created, defensive measures were disallowed, and IFF was always present for Neutrals. Table 4 shows the average percentage change in performance when varying the four test parameters from one extreme to the other; the error rate is significantly more affected than the other metrics.

Table 2: Mean results for survey questions on the interface.

Property	Survey result mean	Property	Survey result mean
Agent pop-up menu	4.2	Contact pop-up menu	4.2
Simulation logs menu	4.0	Doctrine setup menu	4.0
Scenario external attribute menu	4.2	Equipment setup menu	4.0
Watchstander attribute menu	4.0	File menu	4.4
Submenu items logically organized	3.8	Menus logically located by functional area	3.8
Menus easy to understand	3.8	Pop-up menus arranged logically	4.2
Menus arranged logically	4.0	Menus are intuitive	3.8
Methods for tasks are reasonable	3.8	Tasks are understandable	4.0

Table 3: Mean results for survey questions on the simulation realism.

Issue as to realism	Survey result mean	Issue as to realism	Survey result mean
RSC skill change	6.11	RSC experience change	6.00
RSC fatigue change	5.33	RSC SPY radar change	5.00
Team interaction with RSC performance	5.00	EWCO skill change	6.22
EWCO experience change	5.89	EWCO fatigue change	5.44
EWCO SLQ-32 change	5.44	Team interaction with EWCO performance	5.44
FORCE TAO skill change	6.00	FORCE TAO experience change	5.78
FORCE TAO fatigue change	5.56	FORCE TAO decision-maker change	4.33
Team interaction with FORCE TAO performance	5.22	Team interaction with FORCE TAO decision-maker type	4.44
Realism of Trial #1	5.00	Realism of Trial #2	4.67

Table 4: Average percentage change in task time, communications time, and error rate of three watchstanders, when varying key parameters over their range.

	Skill	Experience	Fatigue	SPY Radar Status
RSC (Radar Systems Controller)	3, 27, 47	5, 6, 14	5, 6, 14	7, 16, 11
EWCO (Electronic Warfare Control Officer)	5, 2, 35	12, 9, 34	5, 4, 41	2, 10, 15
FORCE TAO (Force Tactical Action Officer)	2, 11, 63	2, 2, 23	4, 5, 67	7, 17, 15

A variety of additional experiments were conducted. One interesting one compared a scenario where the FORCE TAO's skill and experience were Expert while the fatigue attribute was Exhausted (Trial #1) with a scenario in which the FORCE TAO's skill and experience were Basic and Newly Qualified, respectively, while the fatigue attribute was Well Rested (Trial #2). For the other watchstanders in Trial #1, the skill and experience attributes were Basic and Newly Qualified, respectively, while their fatigue attributes were Fully Rested; in Trial #2, their skill and experiences attributes were Expert while their fatigue attributes were Exhausted. Trial #1 showed better performance than Trial #2 except in the initial radar detection time, suggesting that the status of the team is more important than the status of their commander.

THE ADL SIMULATION

Our second simulation was the Air Defense Laboratory (ADL) Simulation (Ozkan, 2004). It focused on modeling how the Anti-Air Warfare Coordinator (AAWC) reasons during threat assessment for a frigate performing air defense. It also provides a simulation environment that allows users to create realistic air-defense scenarios and examine automated reasoning about the tracks. It is written in Java and uses multi-agent system technology to model the components of reasoning (as opposed to the agents in the previous ADC Simulation, each of which modeled a person). Agents include both track-generator agents that control aircraft activities based on the type of the aircraft, and track-predictor agents that receive data about the aircraft and generate predictions about their identity and possible intent. The simulation program uses a multi-threaded environment: More than 100 threads run in a five-track scenario. XML files are used to store data logs.

CONCEPTUAL BLENDING

The ADL Simulation differs from previous research in that it uses "conceptual blending theory" for the cognitive model. This model should permit better modeling of the way watchstanders reason about contacts. This model is implemented using the CMAS Library, a set of utilities for implementing conceptual blending, developed by the IAGO (Integrated Asymmetric Goal Organization) project at our school.

Conceptual blending theory is a general psychological theory of reasoning (Fauconnier & Turner, 2004) focused on reasoning by analogy in "mental spaces". In a simplest blend operation, there are input, generic, and blend mental spaces. Generic space contains the common input elements of the input spaces as well as the general rules and templates for the inputs. Blend space is the place where the emergent structure occurs. The projected elements from each input space and generic space create a new structure in the blend space. Such blend operations can be cascaded in an "integration network" (Turner & Fauconnier, 1998).

Blending involves composition, completion, and elaboration. Composition involves relating an element of one input space to another with "vital relations"; this matching generally occurs under a "frame". Completion is pattern completion in which generic space is involved in the blending operation. If the elements from both input spaces match the information stored in the generic space, reasoning by analogy is done. Elaboration is an operation that creates a new "emergent" structure in the blend space after composition and completion. Not all elements of the input spaces are projected into blend space, since processes use "selective projection". For example for air contacts, one input space holds aircraft parameters and one holds the air-defense concept; the blend space need not include the color of the aircraft.

Finding the relations between spaces becomes the most important issue in constructing new concepts in integration networks. Turner and Fauconnier propose change, identity, time, space, cause-effect, part-whole, representation, role, analogy, disanalogy, property, similarity, intentionality, and uniqueness as relations. They also propose four kinds of topology for integration networks: simplex networks, mirror networks, single-scope networks, and double-scope networks. Simplex networks have an "organizing frame" in one input space and relevant data in the other; mirror networks share the same organizing frame. The ADL Simulation uses only the simplex and mirror networks; simplex networks are used to assign properties to air tracks, and mirror networks are used to recognize coordinated activities between tracks.

Implementation of the ADL Simulation

Like the ADC simulation, the ADL simulation has a user interface with menu options, an output panel, a toolbar, and a tactical display (Figure 3). The user can specify the environment, create and delete airbases, air routes, and joint points, create user-derived aircraft, load a predefined scenario, save a prepared scenario, and create an Air Tasking Order message for friendly activity in the environment. An output panel shows data of the selected track and the ratings of the competing identities for the track (Figure 4). Logging options can also be specified.

A user can specify the number of randomly generated tracks and the percentages of the types of aircraft. They can also specify particular aircraft and track parameters for testing purposes. They can specifically create snoopers (hostile aircraft collecting intelligence while circling a location), coordinated detachment attacks (tracks that split when two closely adjacent aircraft diverge), and missile firings (tracks that diverge from a firing aircraft).

Our implementation of conceptual blending uses “connectors” and “tickets” (Hiles, 2002). Connectors are like procedure calls in a programming language, and are a way to communicate and coordinate agents. Tickets are like conditional rules in a programming language, and trigger actions when a set of data matches some predefined criteria. Tickets can be complete or incomplete, sequential or non-sequential.

TRACK-GENERATOR AGENTS

Track-generator agents control simulated aircraft including their waypoints, turning, speed changes, altitude changes, radar and IFF transponder usage, and attack tactics. There are seven types. Their implementation was based on our experience, tactical-procedure publications, and air-warfare game documents (Prima Games, 2004).

Track agents find routes by searching across waypoints by using a combination of A* search and depth-first search. At each step in calculating a route, the track agent measures the distance to the next waypoint and determines the course to reach it. When the aircraft comes to the turning point, it turns to a new waypoint with its "turning angle" (the amount the aircraft turns in one tenth of a second). Based on the speed and course, new location points are calculated and track position is set to these points. On the last waypoint the aircraft starts decreasing altitude and speed, and subsequently finishes its mission by landing at the destination airport. This modeling was more accurate and sophisticated than that of the ADC Simulation.

Typical parameter values for Civilian, Friendly, and Hostile aircraft are shown in Table 5. The Civilian aircraft take off from an airport and follow air routes to the destination airport. Takeoff and destination airports are picked randomly. Civilian aircraft fly exclusively between a set of predefined waypoints. Besides this behavior, Friendly and Hostile aircraft also have specialized assigned tasks. Friendly aircraft in our simulation can do offensive counter-air missions, offensive counter-air sweeping to shoot down enemy aircraft, reconnaissance, close-air support, barrier combat air patrol, deep missions, battle damage assessment, suppression of enemy air defense, defensive counter-air missions, and escort missions. Friendly aircraft originate from a friendly country and end in a friendly country, and their IFF-4 values are true with very high probability. Hostile aircraft in our simulation can do high-altitude dive bomb attacks, regular dive-bomb attacks, popup attacks, and masquerading as a civilian aircraft. All hostile aircraft take off from the "threat expected" sector except for the aircraft that have chosen a masquerade attack mission. These aircraft use randomly picked IFF-1, 2, and 3 settings; the IFF-4 value is false for Enemy aircraft with very high probability.

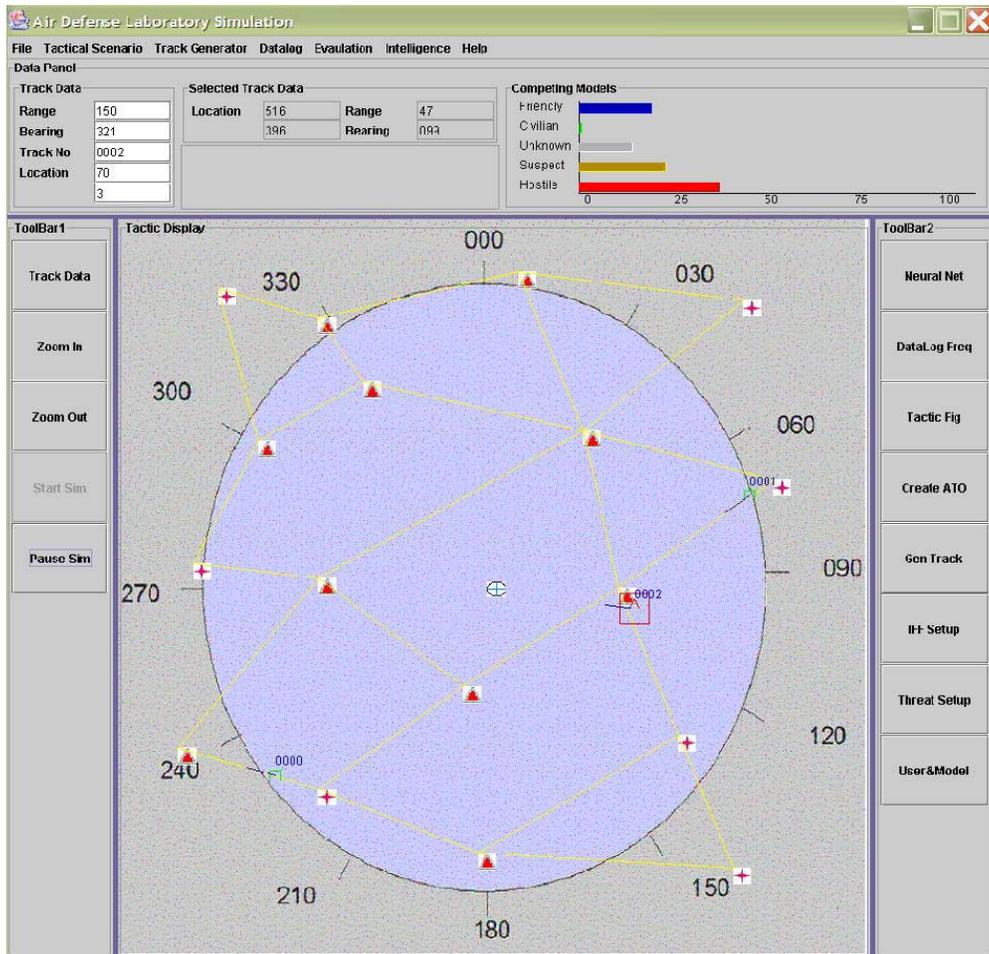


Figure 3: ADL Simulation interface.

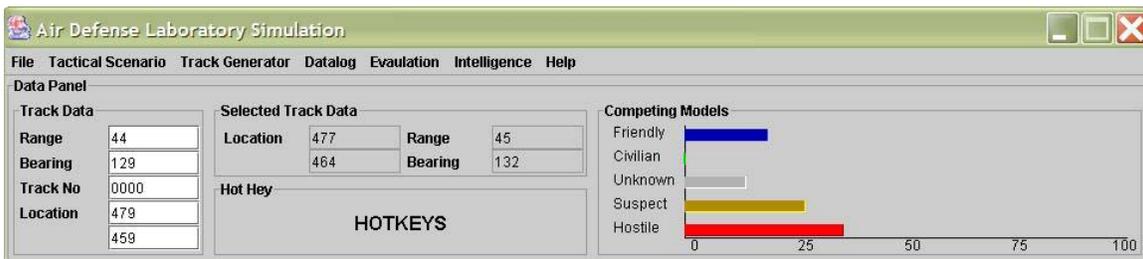


Figure 4: Output panel for the ADL simulation.

Besides these main track-generator agents, there are several others. A Coordinated Detachment Attack track agent at first shows only one aircraft; at about 30 nm from the ship, the track splits and two aircraft fly 20° away from the previous course in different directions. The Missile Attack track agent's mission is to lead the aircraft toward the ship and release its missile about 40 nm away from the ship. Then they turn away and leave the area. Finally, there are missile track agents of two kinds. The sea-skimming missile flies just over the sea with an 80ft altitude; the pop-up missile increases altitude suddenly at close range to the ship and then dives into ship.

Table 5: Aircraft attributes.

Attribute	Civilian	Friendly	Hostile
<i>Takeoff and destination airbases</i>	Randomly chosen	Randomly chosen	Randomly chosen
<i>Initial speed</i>	100	100	100
<i>Maximum speed</i>	400 + random number (1-100)	500 + random number (1-100)	500 + random number (1-100)
<i>Max Acceleration</i>	10-12	10-17	10-17
<i>Maximum altitude</i>	Around 30000	Around 20000 but varies with mission	Around 20000 but varies with mission
<i>IFF-2</i>	0	Determined by ATO	Randomly chosen
<i>IFF-2</i>	0	Determined by ATO	Randomly chosen
<i>IFF-3</i>	Random number 1-9999	Determined by ATO	Randomly chosen
<i>IFF-4</i>	False	True	False
<i>Radar Status</i>	On	On	On
<i>Radar Emission</i>	Civilian type	Military type	Military type
<i>Turning angle</i>	2 degrees	4 degrees	4 degrees
<i>IFF Transponder Status</i>	On	On	On
<i>Origin</i>	Neutral sector	Safe sector	Threat-expected sector

REACTIVE TRACK-PREDICTOR AGENTS

For each factor we defined as being important for air defense, we implemented a reactive agent to monitor its relevant data and inform higher-level “reasoning” agents of changes. We identified 17 reactive agents, all implemented as individual threads. We used the factors in (Liebhaber and Smith, 2000) plus a few additional ones.

The Airplane agent continuously compares the aircraft’s location to airspaces, standard commercial routes that civilian aircraft must follow; being in an airplane increases the probability that an aircraft will be civilian. The Origin agent compares the origin of the aircraft with the threat-expected sectors; friendly aircraft originate with a very high probability from the friendly-country sectors while hostile aircraft originate from the threat-expected sectors.

Naval ships have Electromagnetic Support Measurement (ESM) equipment to detect radar emissions. Civilian and military aircraft have specific navigation radars; military fire-control radars and missile-seeker radars are higher frequencies than navigation and surveillance radars. (Liebhaber & Smith, 2000), cites radar detection as a major identification clue. We defined five different ESM reports in the simulation: civilian, military, military fire-control, military seeker, and no radar emission. The IFF agent reports Identification Friend or Foe information; IFF Mode 1 shows the mission of a military aircraft, Mode 2 the squadron, Mode 3 the air-traffic control controlling the aircraft (also used by civilian aircraft), Mode 4 encrypted information friendly forces, and Mode C the altimeter value of the aircraft. Similarly, a radar status agent follows radar emissions.

Other agents monitor the contact heading change (since military aircraft turn more), maximum acceleration (since military interceptors and fighter aircraft can accelerate more than civilian aircraft), altitude (military aircraft fly at varied altitudes), maximum altitude, maximum speed (military aircraft have larger maximum speed than civilian aircraft), and speed variation (civilian aircraft maintain a steadier speed than military aircraft). A Snooper-Detector agent identifies snooping activities by tracking reported locations in a two-dimensional array. If the number of locations in a given time within a sufficiently small bounding polygon exceeds a certain number, the agent identifies this as a snooper.

Random errors are added by the agents into the parameters they report to make the simulation more realistic. Errors increase with the range of the contact because of free-space loss, attenuation, distortion, fading, and multi-path propagations. We used a formula for the error percentage which was the natural exponent of the range times a constant, for which we used 0.033 as a reasonable guess (the actual formula is classified). This means that at a 128 nm range, the kinematics values received from the air track are wrong 71% of the time in the simulation with a random variation.

REASONING TRACK-PREDICTOR AGENTS

Each contact has an associated higher-level track-predictor agent which infers the identity and the potential intention of the aircraft. These “reasoning” agents use messages from the reactive track-predictor agents described above. Each reasoning agent has two kinds of tickets, identity and independent. We define an identity ticket for each aircraft identity; each identity has a weight and the weights are changed from the information transmitted with connectors from the reactive agents. Initially, the Unknown has a 0.5 weight and all other models have 0.0001.

- A Civilian identity ticket contains six data frames: a ESM frame, an altitude frame, a speed frame, an airplane frame, an IFF evaluation frame, and an origin frame
- A Friendly identity ticket has an ESM frame, an IFF evaluation frame, an origin frame, and an ATO frame.
- The Hostile identity ticket has an ESM frame, a range frame, an altitude frame, an airplane frame, a CPA frame, an origin frame, an IFF evaluation frame, a speed frame, a maximum-speed frame, and a combination of a altitude, range and CPA frames.
- A Suspect identity ticket has an IFF evaluation frame, an altitude frame, an origin frame, a speed frame, and a maximum-speed frame.
- An Unknown identity ticket has an IFF evaluation frame and an ESM frame.

The Predictor track agent calculates the weights of each identity ticket by adding or subtracting contributions from the reactive agents in support of each identity. Two weighting methods were explored: analog, where the weight change is a continuous variable, and digital, where the weight change is constant once the reactive-agent result is over a threshold. For instance, for altitude change and the Civilian ticket, the analog approach adds $5e^{0.0001A}$ where A is the altitude to the weight, while the digital approach adds 100 to the weight if the aircraft is between 25,000 feet and 35,000 feet. Similarly, we add $100\sin(1.81(S - 400))$ where S is the speed with the analog approach, and 100 with the digital approach if the speed of the aircraft is between 0.76 and 0.89 mach.

The track-predictor agents also have some independent tickets:

- The ATO Evaluation ticket has frames for IFF-1, IFF-2, IFF-3, heading, location, and time.
- The CPA calculator ticket has a frame for heading change.
- The IFF Evaluation ticket has frames for IFF-1, IFF-2, IFF-3, and IFF-4.
- The Split-Identity Detector ticket has two frames for location.
- The Combination independent ticket has three frames for altitude, CPA, and range. If aircraft is inbound, its range is close, and its altitude is low, a combination ticket sets the combination frame of the hostile ticket to true to provide extra weight to the hostile identity.

The ATO ticket sets the ATO frame of the Friendly identity ticket by comparing the IFF, location, and heading of the contact to that of known air-tasking orders. The CPA ticket calculates the closest point that the air track will pass by the ship. The IFF Evaluation ticket combines data from four IFF frames. IFF Mod I and IFF Mod II may be set or not set; if they are set, they may be right or wrong. That makes total of three possibilities for each IFF Mod I and IFF Mod II. IFF Mod III and Mod IV have two possibilities each. Thus there are 36 different combinations for IFF Evaluation ticket. We defined a table for these combinations giving a weight value for each identity ticket for each combination, based on standard procedures for air-defense teams. Finally, the split-activity detector ticket has two location frames, one set by another track-predictor agent’s location connector when that track is first created, the other is set by

track-location data. If the other track location is found close to the first track location, a split connector is extended.

Besides these, there are “regional” agents which combine information from tracks. They look for snooper activity, coordinated detachment activity, and merge activities. When a snooper is detected, the threat level is increased to yellow. The coordinated detachment detector ticket finds two correlated contracts that are close to the home ship. The merge agent looks for nearby tracks at a similar altitude. Figure 5 shows how merge detection can be viewed as form of the “mirror network” form of conceptual blending.

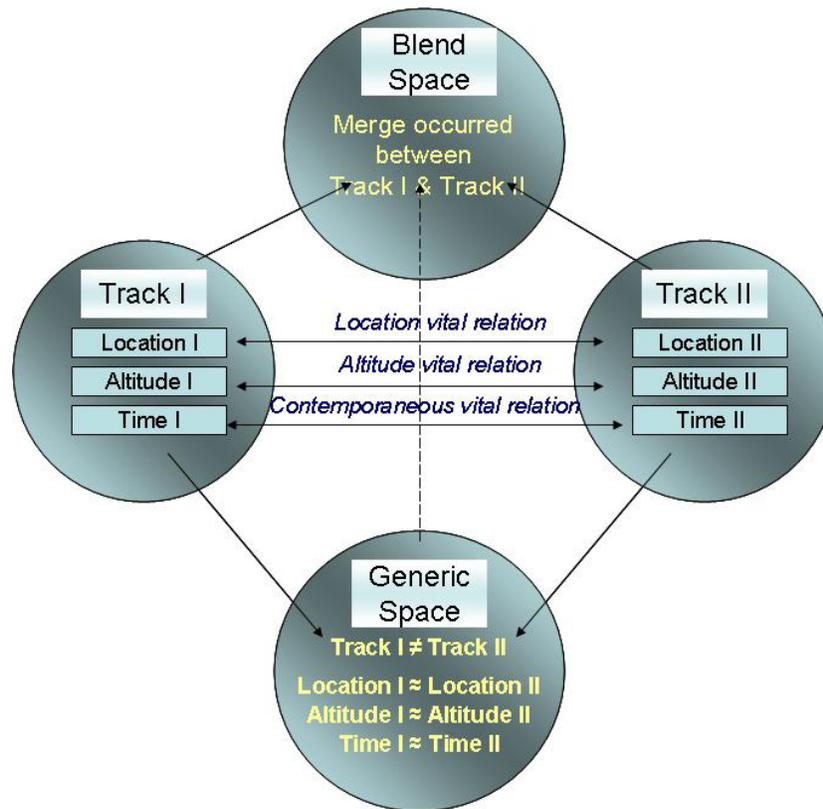


Figure 5: The merge detector blending operation.

EVALUATION OF THE ADL SIMULATION

Our tests used a range threshold of 25 nm, a speed threshold of 500 knots, and a CPA threshold of 15 nm. We ran the simulation 10 times for each test, which resulted in 190 runs. 10 runs tested the level of reality of the ADL Simulation, 90 runs tested analog decision-making, and 90 runs tested digital decisionmaking. We limited each scenario time period to 5-6 minutes. We used a uniform distribution when selecting random numbers. The scenarios tested are listed in Table 6.

The ADL Simulation was tested by two air-warfare officers (AAWOs), two principal warfare officers (PWOs), and three Air Force pilots. We ran the simulation ten times with different scenarios for each subject. We asked the PWOs and one AAWO to assess the closeness of the decisions of the model to decisions of the real air-warfare personnel. We asked them to talk continuously while they made decisions to catch the factors affecting the decisionmaking process, and recorded their voices on a tape recorder in addition to logging the runs they viewed. In general the experts confirmed the realism of the simulation, but:

Table 6: Test scenarios.

Scenario	Scenario
1	5 civilian aircraft with/without threat intelligence
2	3 civilian aircraft and 1 friendly aircraft with/without threat intelligence
3	3 civilian, 1 friendly, and 1 hostile aircraft with/without threat intelligence
4	2 civilian, 1 hostile, 1 snoopers, and 1 friendly aircraft with/without threat intelligence
5	3 civilian aircraft and a coordinated detachment attack with/without threat intelligence
6	3 civilian aircraft and a missile attack with/without threat intelligence
7	3 civilian aircraft and a terrorist attack with/without threat intelligence
8	3 civilian aircraft, 1 missile, and a coordinated attack with/without threat intelligence
9	3 civilian aircraft and a terrorist attack with/without threat intelligence

- One expert criticized the lack of issued warnings to air contacts.
- Two experts stated that it would be more realistic if the ship had movement capability.
- Two experts criticized the lack of tasking for air-defense missions. We restricted the simulation to only one ship and ignored Link services, but real-world air defense involves coordination between ships.
- Three experts criticized the Cartesian coordinate system used in the simulation.
- Four experts declared that the civilian, snoopers, coordinated detachment attack, and missile-attack agents behaved as they should, but that the behaviors of the friendly and hostile aircraft could be made more realistic. Some of this is due to our use of necessarily unclassified models from game technologies.

We recorded the actual identities of the aircraft in each simulation and compared them with the track-predictor identities. We ran the simulation ten times for each of nine different scenarios. Tables 7, 8, and 9 show the average time in seconds for the model to identify the contacts correctly. It can be seen that analog evidence fusion performed better than digital fusion.

Table 7: Average identification time for civilian contacts.

Time to Identify Civilian	Analog process without Threat Intelligence	Analog Process with Threat Intelligence	Digital Process without Threat Intelligence	Digital Process with Threat Intelligence
Mean	17.33	18.38	17.98	20.479
Standard Deviation (s)	5.68	7.315	4.665	5.157
Variance (s ²)	32.269	53.519	21.77	26.597

Table 8: Average identification time for friendly contacts.

Time to Identify Friendly	Analog process without Threat Intelligence	Analog Process with Threat Intelligence	Digital Process without Threat Intelligence	Digital Process with Threat Intelligence
Mean	11.139	10.74	13.09	12.599
Standard Deviation (s)	5.324	4.135	4.722	4.606
Variance (s ²)	28.347	17.1	22.301	21.222

The results also confirmed that the factors used by the subjects were a subset of the factors defined in the ADL Simulation. However, the ADL Simulation was ten times faster than human decisionmaking on the average. No differences were found between the final identifications made by the simulation and experts.

Table 9: Average identification time for hostile contacts.

6 Minute Scenario 1 Hostile Track		Analog process without Threat Intelligence	Analog Process with Threat Intelligence	Digital Process without Threat Intelligence	Digital Process with Threat Intelligence
Time to Identify Hostile (sec)	Mean	11.3	11.52	13.3	11.059
	Standard Deviation (s)	0.82	1.541	1.232	0.482
	Variance (s ²)	0.674	2.376	1.519	0.233
Total Time aircraft is identified as Hostile (sec)	Mean	180.66 (51.21%)	274.66 (76.3%)	195.959 (56%)	114.16 (31.7%)
	Standard Deviation (s)	17.093	89.879	61.245	57.047
	Variance (s ²)	313.069	8078.37	3751.05	3254.452
Total Time aircraft is identified as Suspect (sec)	Mean	157.836 (44.74%)	63.44 (17.62%)	135.94 (38%)	227.16 (63.1%)
	Standard Deviation (s)	16.807	89.882	59.459	55.257
	Variance (s ²)	282.501	8078.89	3535.427	3053.343

A BAYESIAN SIMULATION

Our third approach was to build a relatively simple Bayesian model of the reasoning in naval air defense, a model that could learn from experience. The analysis of the factors used by experts in (Liebhaber & Smith, 2000) is surprisingly simple, since the air-defense environment has few clues and the way the clues are combined is not complex. So rather than studying human performance by hand-coding a detailed model as the previous two simulations did, it might make more sense to code a relatively simple model and have it adjust a few parameters from training using the classic techniques of Bayesian analysis. Another advantage could be that people have difficult trusting a model they cannot follow, and Bayesian methods often provide understandable simple models.

The factors identified in (Liebhaber & Smith, 2000) fall into two categories: Those affecting the time a contact could take to reach the home ship, and the intrinsic suspiciousness of the contact. The first functions like a negative cost, and the second like a probability for a decision-theoretic model. It thus makes sense to segregate the factors affecting these two metrics and then combine their influences. For time to reach the ship, speed, maximum speed, heading, turn radius, and bearing to the ship are the key factors. As a first approximation, we can add the time for an aircraft to turn in a circle to a bearing facing the ship and the time to reach the ship from there. Assuming the aircraft is distance D from the ship at a bearing of B from the ship and a heading of H with turn radius in the azimuth (map) plane of R , the distance it will need to travel to reach the ship by the shortest route is:

$$R \min(|H - B|, |B - H|) + \sqrt{D^2 + 2R^2(1 - \cos(H - B))} - 2DR \sin(H - D) .$$

As for the relationship between time to potential attack and perceived seriousness of the threat, it makes sense, on the average, that the more time available to counter a threat, the more options are available to counter that threat and the more effective those countermeasures will be. So we will assume that the seriousness of the threat is a decreasing sigmoid function of the time to reach the ship.

As for the intrinsic suspiciousness of the contact, (Lieberhaber & Smith, 2000) cite the following factors from their interviews with experts, most of which we implemented for our simulation:

- Low-altitude level flight
- Significant distance from a civilian airline
- Hostile or unknown airport of origin
- Sharp turn made
- High speed
- Aircraft over water
- Not heading toward a civilian airport
- Military-type electronic emissions
- Nonzero or nonexistent IFF response
- Weapons systems apparent
- Missile launches
- Coordination with other aircraft
- Air support apparent (not implemented)
- Intelligence reports indicating hostilities (not implemented)

These factors are close to independent of one another when spurious, so a Naive Bayes formula is an appropriate approximation. We use the odds form of Naive Bayes to handle better the occurrence of both positive and negative evidence, and only try to estimate the degree to which an aircraft is hostile (the most important conclusion about a track). We use:

$$o(H | (E1 \& E2 \& \dots \& En)) = [o(H | E1) / o(H)] * [o(H | E2) / o(H)] * \dots * [o(H | En) / o(H)] * o(H)$$
where o represents odds or $p/(1-p)$, H represent the hypothesis that the aircraft is hostile, " $|$ " means "given that", and $E1, E2, \dots, En$ represent pieces of evidence.

We maintain counts for each of four counts for a hostility factor: (1) the number of occurrences of the factor with hostile contacts, (2) the number of occurrences with nonhostile contacts, (3) the number of times the factor did not occur with hostile contacts, and (4) the number of times the factor did not occur with nonhostile contacts. Initial counts were assigned from intuition. Then every time a track terminates during the simulation, we add its weighted counts to the current counts. Weightings were based on the final assessed probability that the contact was hostile, not the probability when the clue was observed. For instance, if the final probability of track 17 is that it was 80% likely to be hostile, and the track was within an airline 20 times and outside 70 times, we add to the count for "hostile when in an airline" the number $20 * 0.8 = 16$, to the count for "hostile when not in an airline" the number $70 * 0.8 = 56$, the count "not hostile when in an airline" $20 * 0.2 = 4$, and the count "not hostile when not in an airline" $70 * 0.2 = 14$.

Figure 6 shows an example display from the simulation, which was also written in Java. The ship is the yellow circle; the shaded area is water; blue circles are friendly airports; green circles are neutral airports; red circles are hostile airports; and red rectangles are air contacts displayed with identification numbers and current altitude. Eight regions within each rectangle represent the presence or absence of eight main clues used in the odds multiplication: The first row shows (left to right) the occurrence of low-altitude level flight, being outside an airline at some time, hostile origin, and a significant turn at some point not near an airport; the second row shows (left to right) high speed at some time, that the aircraft was over water at some time, the aircraft is not heading toward a neutral or friendly airport, and some military-suggestive event occurred (weapons presence detected, a launch, nonzero IFF, or coordinated activity with another track). The right side of the display shows the current estimated probability that the contact is hostile and the current estimated danger level for the ship from that contact. In the display shown, the system has correctly recognized all of the hostile aircraft from their behavior, although the probabilities will be improved as more evidence accumulates. Several kinds of aircraft tracks are created randomly for scenarios, including point-to-point travel, arrival from outside the view area, snooping, coordinated attacks, and hijacked civilian aircraft. The simulation can easily track hundreds of aircraft well in excess of human capabilities and appeared to better handle large numbers of aircraft than the other two simulations.

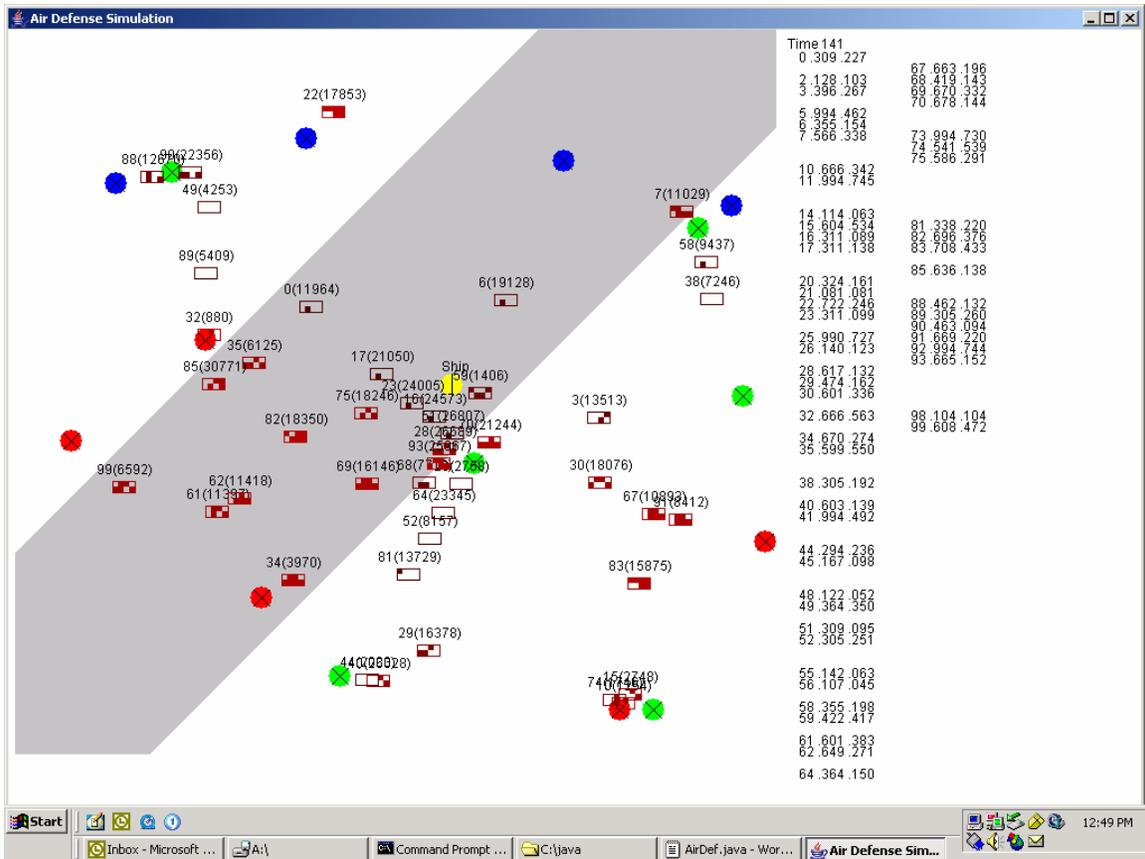


Figure 6: View of the Bayesian simulation.

We measured performance as the root-mean-square error in identification (with 1.0 being the perfect probability for a hostile contact and 0.0 for a nonhostile contact). Learning definitely improved performance in distinguishing hostile contacts. For instance, after running four times on the same scenario with 100 randomly generated aircraft tracks and a conservative value for the learning rate, performance improved from a 0.377 root-mean-squared error over the entire run to 0.369. Performance after running on four different randomly generated 100-aircraft scenarios went from 0.377 to 0.383. However, when run on a fifth different scenario of 100 tracks, performance of it was 0.398 whereas performance on the fifth scenario when trained four times on the first scenario was 0.411, demonstrating the importance of variety in training. All these scenarios involved around 12,000 error terms for 12,000 platform-time data points.

We also showed the system could learn to some extent a new idea, that hijacked civilian aircraft could be used to attack the ship. We excluded the hijacked tracks from four scenarios, and trained on those four in succession; the performance went from 0.345 to 0.341. Then performance was 0.411 on a scenario with hijacking. We then trained it on the four previous scenarios which included hijacking, and performance ended at 0.403, so it was able to readjust, and more successfully than having just trained on one scenario with hijacking, albeit less successfully than with more-restricted scenarios. In general, the system can be misled into incorrect learning by a cleverly deceptive adversary, as for instance if hostile aircraft head repeatedly toward hostile airports until one attack when they head toward the ship. But human air-defense experts would be misled in much the same way. The simulation, however, can handle much more complex situations than a human could without becoming confused. It requires only 47 kilobytes of source code.

CONCLUSIONS

All three simulations should be useful in training air-defense personnel. The ADC Simulation (1.004 megabytes in its JAR or executable file, including images) focused on the social interaction of an eleven-member CIC air defense team. It was particularly useful in analyzing the effects of watchstander skill, experience, fatigue, type of decision-maker, and environmental attributes on the performance of the individual as well as the overall CIC watch team, and showed bottlenecks and other inefficiencies. Assessment by air-defense experts suggested that it was quite realistic. However, it is tailored to the way air defense is conducted today and would need redesign to model new methods.

The ADL simulation (3.137 megabytes of source code and images including 1.987 megabytes for JAR files for the CMAS Library) focused on decision-making about contacts. It seemed to perform well in modeling the results of human air-defense assessment as judged by experts. However, it is a complex program because of its "conceptual blending" and the complexity does not appear justified by any significant improvement in performance over the Bayesian simulation which was 60 times smaller not even counting the CMAS Library. (To be fair, it was more user-friendly for training needs than the Bayesian simulation.) While the tickets and connectors of the ADL simulation should in principle be better able to model the process by which evidence is considered by air-defense experts, we could not prove it did. So Occam's Razor applies and we should prefer a simpler simulation. Since air defense is far from the usual linguistic domains to which conceptual blending has been applied, and is not especially complex, air defense either may not be an appropriate application for conceptual blending or more work on details may be needed to apply it better.

The Bayesian simulation (0.047 megabytes of source code with no images, resulting in 0.023 megabytes of Java-class code) showed that naval air defense can be learned by a relatively simple program. This suggests that eleven people are not necessary to do it, and careful design of an initial air-defense program may not be critical. Does this mean that naval air defense is "solved"? Not by any means, because there are still difficult issues in the interaction between software and air-defense personnel. But the simplicity of the Bayesian model suggests that judicious use of software support will enable air-defense tasks and their training to be simplified, to enable better human and human-system performance.

REFERENCES

- Alvarez-Vaquero, F., "Signal Processing Algorithms by Permutation Test in Radar Applications," SPIE Proceedings, Vol. 2841, Advanced Signal Processing Algorithms, Architectures, and Implementations VI, Denver Colorado, August 1996, pp. 134-140.
- Amori, R., "An Adversarial Plan Recognition System for Multi-Agent Airborne Threats," Symposium on Applied Computing, 1992.
- Barcio, B., Ramaswamy, S., MacFadzean, R., & Barber, K., "Object-Oriented Analysis, Modeling, and Simulation of a Notional Air Defense System," IEEE International Conference on Systems for the 21st Century, Vancouver, BC, Canada, Vol. 5, October 1995, pp. 22-25.
- Bloeman, A., & Witberg, R., "Anti-Air Warfare Research for Naval Forces," *Naval Forces*, Vol. 21, No. 5, 2000, pp. 20-24.
- Burr, R., Palinkas, L., and Banta, G., "Psychological Effects of Sustained Shipboard Operations on U.S. Navy Personnel," *Current Psychology: Developmental, Learning, Personality, Social*, Vol. 12, 1993, pp. 113-129.
- Calfee, S., & Rowe, N., "Multi-Agent Simulation of Human Behavior in Naval Air Defense," *Naval Engineers Journal*, Vol. 116, No. 4, Fall 2004, pp. 53-64.
- Choi, S., & Wijesekera, D., "The DADSim Air Defense Simulation Environment," Proceedings of the Fifth IEEE International Symposium on High Assurance Systems Engineering, 2000, pp. 75-82.
- Delaney, M., "AADC (Area Air Defense Commander): The Essential Link," *Sea Power*, Vol. 44, No. 3, pp. 30-34, March 2001.

Eddy, M., Kribs, H., & Cohen, M., "Cognitive and Behavioral Task Implications for Three Dimensional Displays Used in Combat Information/Direction Centers," Technical Report 1792, SPAWAR, San Diego, March 1999.

Fauconnier, G., & Turner, M., *The Way We Think*. New York: Basic Books, 2002.

Federation of American Scientists, "AN/USQ-T46(V) Battle Force Tactical Training System," retrieved from www.fas.org/man/dod/-101/sys/ship/weaps/an-usq-t46.htm, May 2003.

Ferber, J., *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*, Addison-Wesley, 1999.

Hall, D., & Llinas, H., "An Introduction to Multisensor Data Fusion," *Proceedings of the IEEE*, 85, 1997.

Hiles, J., "StoryEngine: Dynamic Story Production Using Software Agents That Discover Plans," unpublished technical paper, Naval Postgraduate School, 2002.

Liebhaber, M. J., & Feher, B., "Air Threat Assessment: Research, Model, and Display Guidelines", Pacific Science and Engineering Group (2002).

Liebhaber, M. J., & Smith, P., "Naval Air Defense Threat Assessment: Cognitive Factors and Model," Command and Control Research and Technology Symposium, Monterey, CA, June 2000.

Maiorano, A., Carr, N., & Bender, T., "Primer on Naval Theater Air Defense," *Joint Forces Quarterly*, No. 11, Spring 1996, pp. 22-28.

Morrison, J., Kelly, R., Moore, R., & Hutchins, S., "Implications of Decision-Making Research for Decision Support and Displays," in Cannon-Bowers, J.A., and Salas, E. (eds.), *Making Decisions under Stress: Implications for Training and Simulation*, Washington DC: APA Press, 1998, pp. 375-406.

Noh, S., and Gmytrasiewicz, P., "Rational Communicative Behavior in Anti-Air Defense," Proceedings of International Conference on Multi Agent Systems, Paris, France, July 1998, pp. 214-221.

Osga, G., Van Orden, K., Campbell, N., Kellmeyer, D., & Lulue, D., "Design and Evaluation of Warfighter Task Support Methods in a Multi-Modal Watch Station," Technical Report 1864, Space and Naval Warfare Systems Center, San Diego, June 2001.

Ozkan, B. E., "Autonomous Agent-Based Simulation of a Model Simulating the Human Air-Threat Assessment Process", M.S. thesis, U.S. Naval Postgraduate School, March 2004.

Prima Games, "Prima Fast Track Guides," www.primagames.com/strateg/guide/90, 1 March 2004.

Turner, M., & Fauconnier, G., "Conceptual Integration Networks," *Cognitive Science*, Vol. 22, No. 1, 1998, pp. 133-187.

Weaver, J.L., Bowers, C., Salas, E., & Cannon-Bowers, J., "Networked Simulations: New Paradigms for Team Performance Research," *Behavior Research Methods, Instruments, and Computers*, Vol. 27, No. 1, 1995, pp. 12-24.

Acknowledgements: This work was supported in part by the Chief of Naval Operations, U.S. Navy, N61F22. The Space and Naval Warfare Systems Center, San Diego, California funded LT Calfee's research through the SPAWAR Research Fellowship Program. The views expressed in this paper are those of the authors and do not reflect policy of the U.S. Navy or the U.S. Department of Defense.

Three Simulation Models of Naval Air Defense

*LTJG Baris Ozkan, Neil C. Rowe,
LT Sharif H. Calfee, and John E. Hiles*
U.S. Naval Postgraduate School
Contact: ncrowe@nps.edu

Why isn't one simulation enough?

- ❑ Different simulations can focus on different features of a problem.
- ❑ Combining all features of a problem into one simulation may be too confusing to understand.
- ❑ Simulations can (1) simulate ordinary people, (2) simulate experts, (3) simulate good learners.
- ❑ We built three agent-based simulations of naval air defense of those kinds.
- ❑ Primary architects were Calfee for (1), Ozkan for (2), and Rowe for (3).

Naval air defense

- ❑ Goal is to protect a naval ship from air attack.
- ❑ Inputs are locations of “contacts” (platforms) obtained from radar and their observed properties.
- ❑ Outputs are orders for defensive measures.
- ❑ For U.S. Navy, air defense is done in the Combat Information Center (CIC) by 11 or more people.
- ❑ Thorough training is important, so all simulations logged results and replay of scenarios for use in a training tool.
- ❑ (Liebhaber & Smith, 2000) gives an excellent list of clues for evaluating contacts as reported by human experts from the U.S. Navy. So we used that.

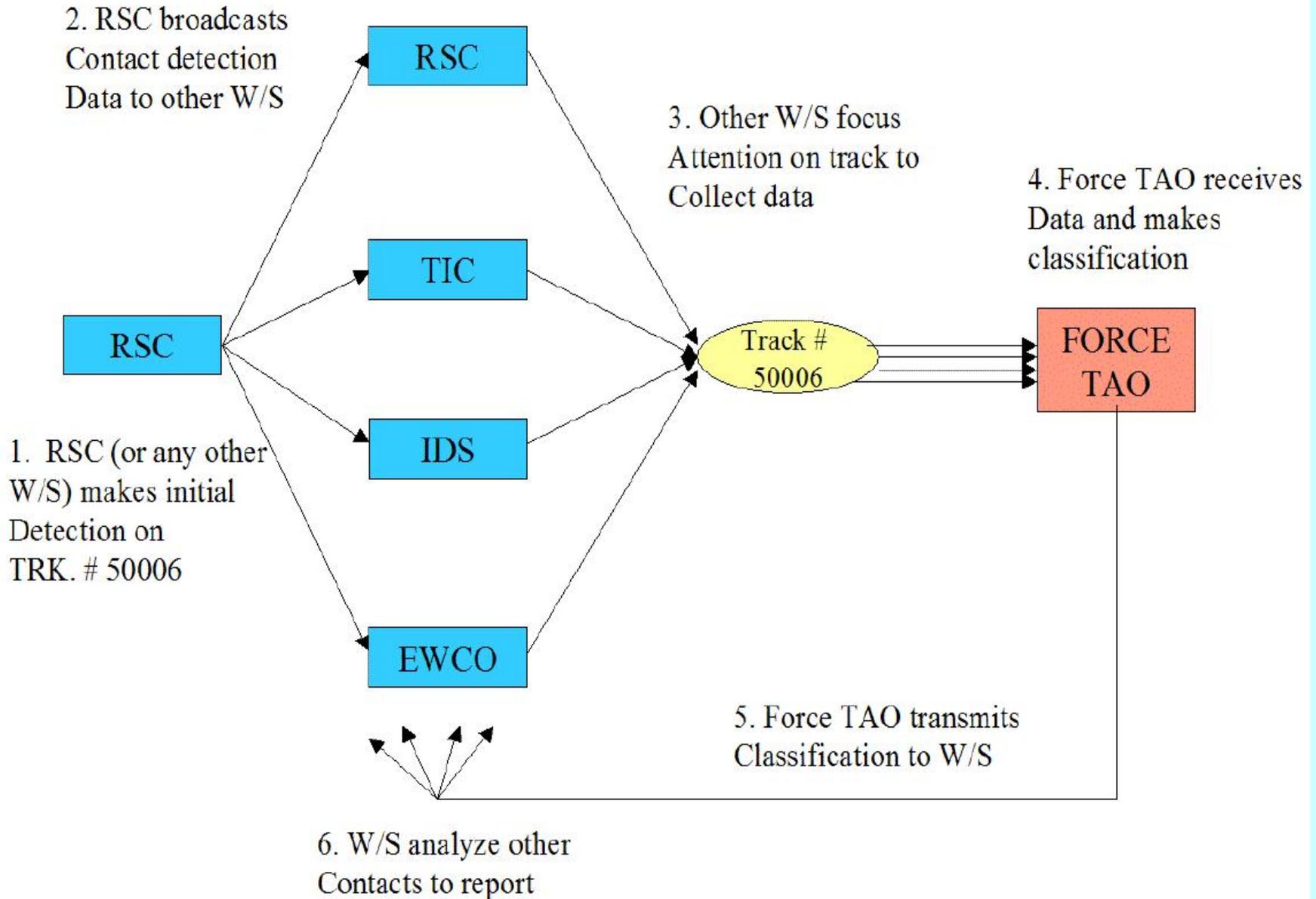
Track factors cited in (Liebhaber & Smith, 2000)

- ❑ Low-altitude level flight
- ❑ Significant distance from civilian airplane
- ❑ Hostile or unknown airport of origin
- ❑ High speed
- ❑ Sharp turn
- ❑ Aircraft over water
- ❑ Not heading to civilian airport
- ❑ Military-type electronic emissions
- ❑ Nonzero or nonexistent IFF response
- ❑ Weapons system apparent
- ❑ Missile launch
- ❑ Coordination with other aircraft
- ❑ Air support
- ❑ Intelligence reports suggesting hostilities

The ADC (Calfee) simulation interface



Information flow between simulated people



More about the Calfee simulation

- ❑ Modeled human imperfections too: Workload limits, lack of training, fatigue, equipment failure, situation criticality, weather.
- ❑ A neural network aggregated clues to classify contacts as Hostile, Suspect, Unknown, Neutral, and Friendly (at different thresholds of a single metric).
- ❑ Details were obtained from interviews with air-defense personnel.
- ❑ Subsequent tests at SPAWAR confirmed the realism of the simulation.
- ❑ The simulation is excellent for answering “what if” questions about the effect of factors like fatigue and training.

The ADL (Ozkan) simulation

- ❑ Focused on modeling the reasoning about air contacts done by the Anti-Air Warfare Commander (AAWC).
- ❑ Used “conceptual blending” to model these inferences, a form of reasoning by analogy.
- ❑ Used agents to represent the pieces of reasoning, not people.
- ❑ Modeling was done by interviews and documents.
- ❑ Experts confirmed the accuracy of the simulation.

The ADL simulation interface

The screenshot displays the Air Defense Laboratory Simulation (ADL) interface. The main window title is "Air Defense Laboratory Simulation". The menu bar includes "File", "Tactical Scenario", "Track Generator", "DataLog", "Evaluation", "Intelligence", and "Help".

Data Panel:

- Track Data:** Range: 150, Bearing: 321, Track No: 0002, Location: 70, 3.
- Selected Track Data:** Location: 518, 396; Range: 47; Bearing: 093.
- Competing Models:** A horizontal bar chart showing the distribution of models: Friendly (blue), Civilian (green), Unknown (grey), Suspect (yellow), and Hostile (red). The Hostile model is the most prominent, extending to approximately 35 on a scale of 0 to 100.

Tactic Display: A circular radar-like display with a blue background and a yellow grid. The display is labeled with bearings: 000, 030, 060, 090, 120, 150, 180, 210, 240, 270, 300, and 330. Several tracks are visible, represented by red triangles and yellow lines. Track 0002 is highlighted with a red box. Other tracks are labeled 0000, 0001, and 0002. A central symbol resembling a circle with a cross is also present.

ToolBar1 (Left): Track Data, Zoom In, Zoom Out, Start Sim, Pause Sim.

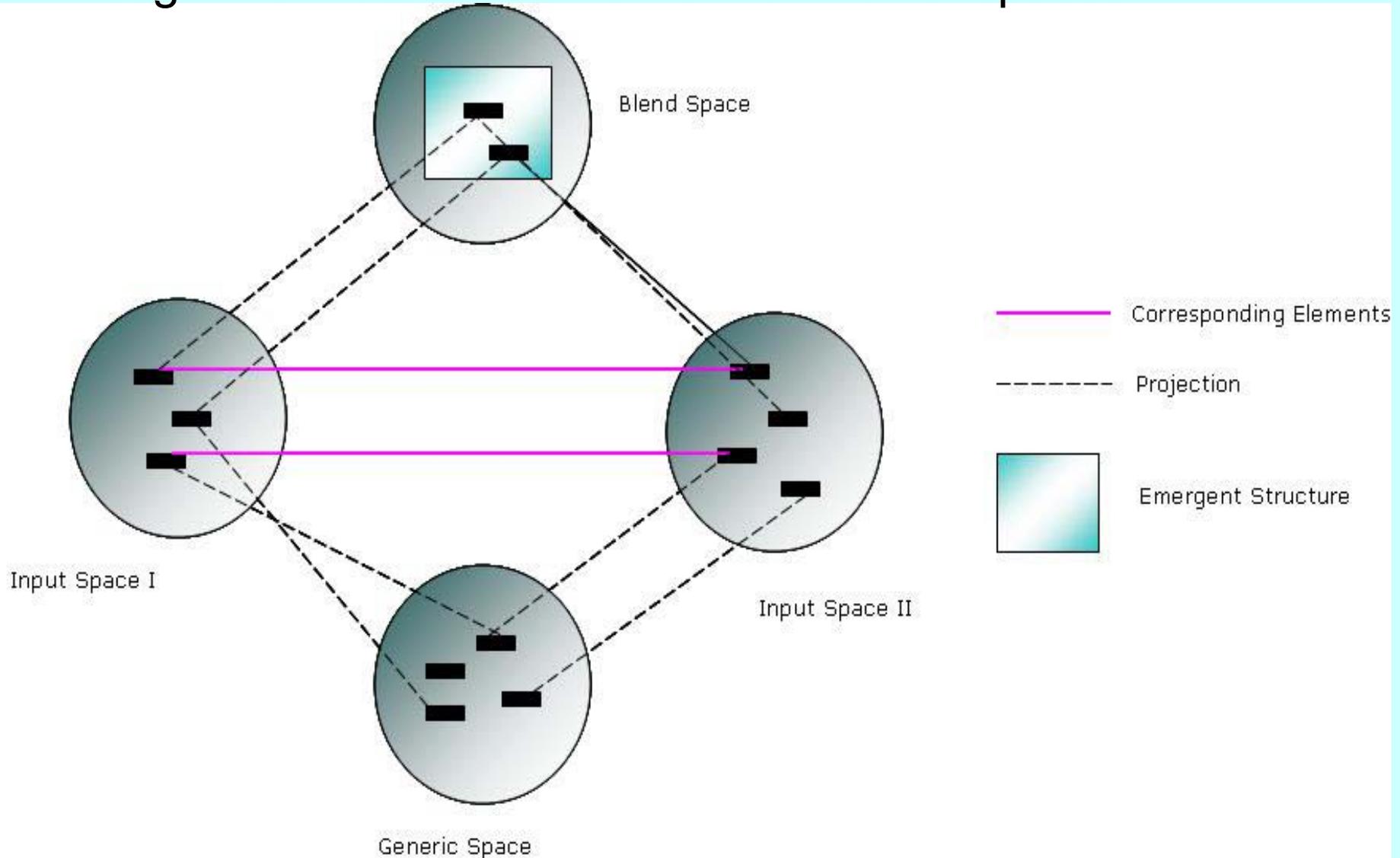
ToolBar2 (Right): Neural Net, DataLog Freq, Tactic Fig, Create ATO, Gen Track, IFF Setup, Threat Setup, User&Model.

Agents in the ADL Simulation



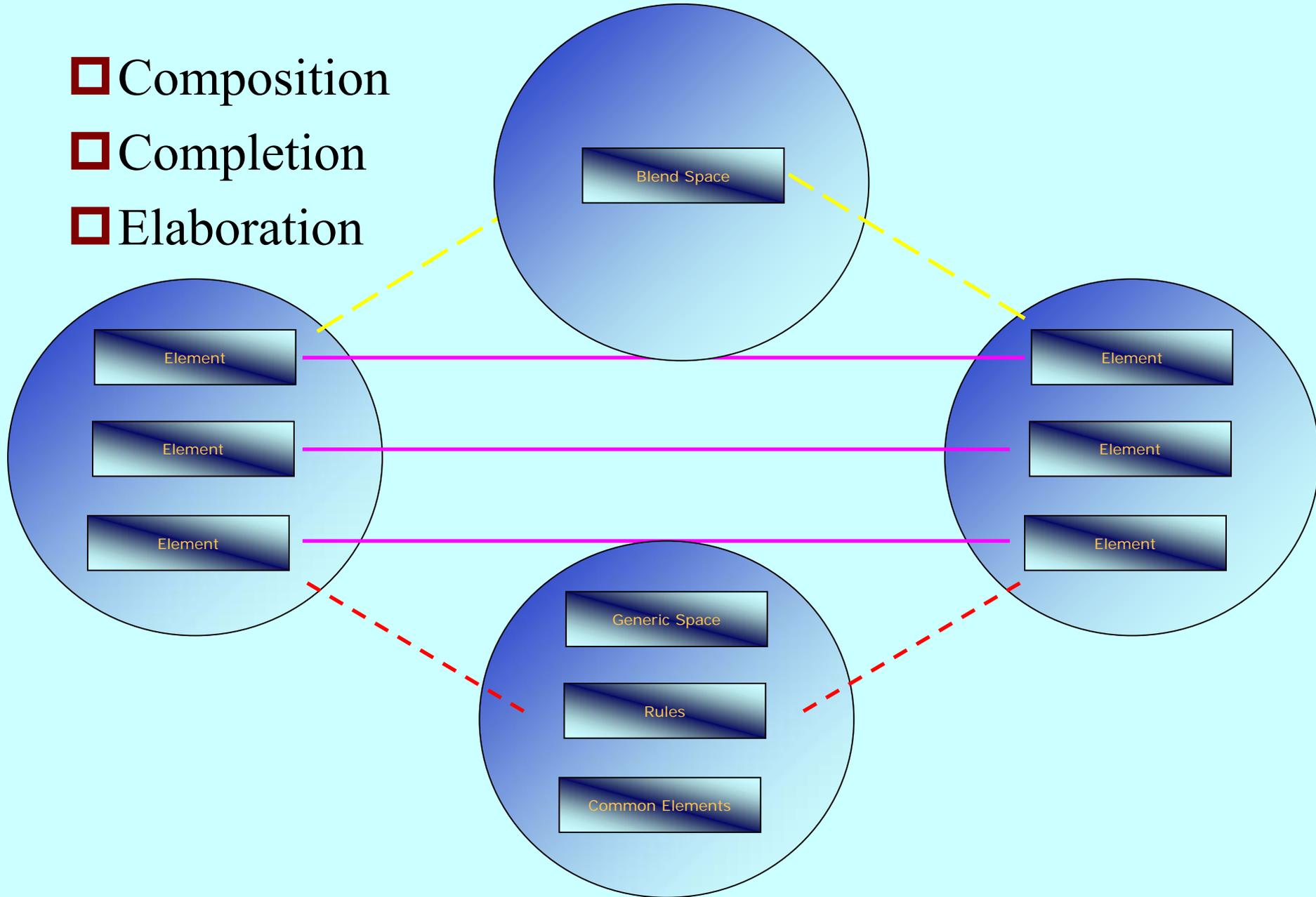
Principles of Blending

Conceptual blending is a set of operations for combining cognitive models in a network of mental spaces.

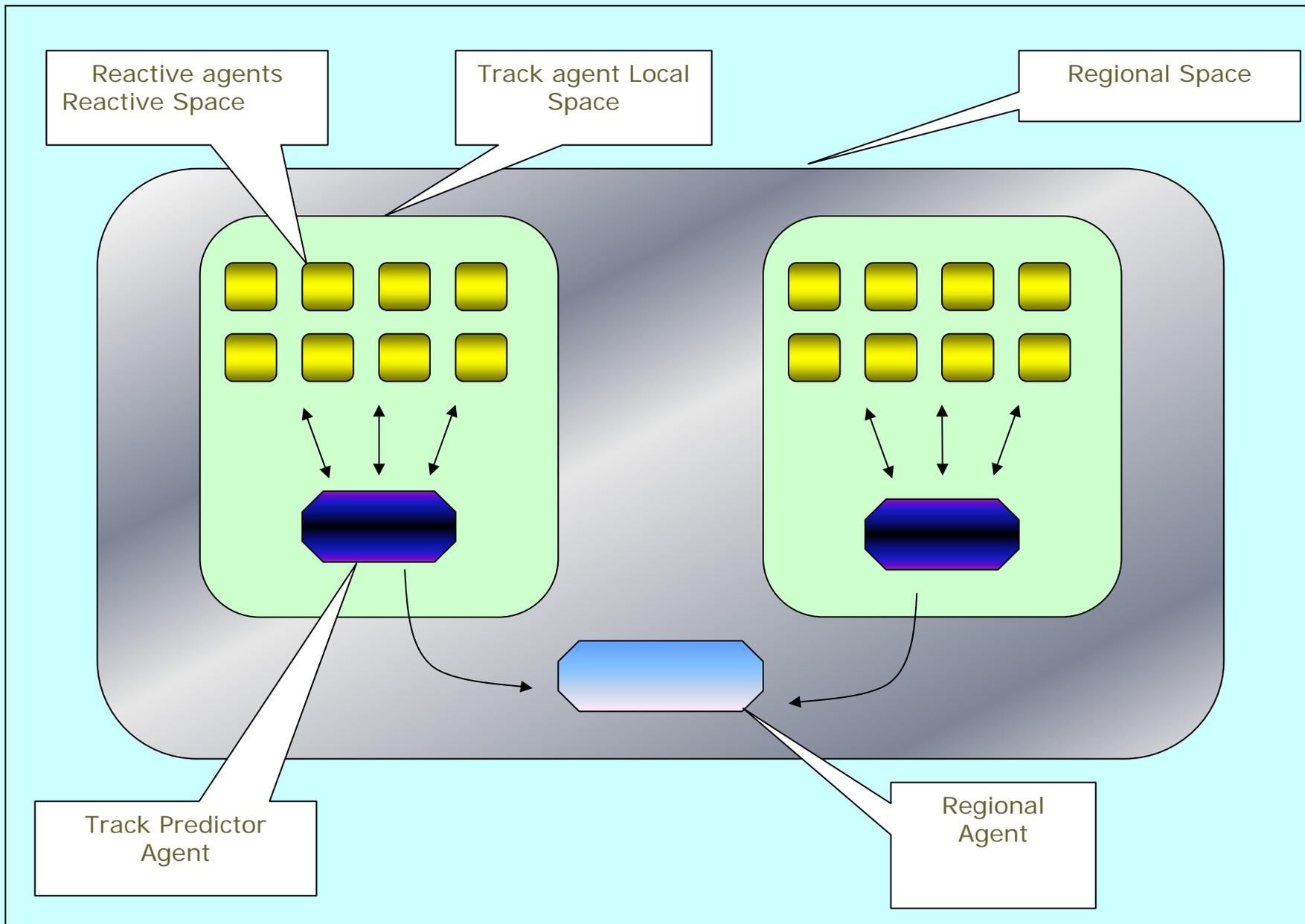


Operations of Blending

- Composition
- Completion
- Elaboration



Air Defense Laboratory Simulation



Air Defense Laboratory (ADL) Simulation

Reactive Agent Factors

```
graph TD; A[Reactive Agent Factors] --- B[Used continuously]; A --- C[Used once and then only used if changed];
```

Used continuously

- Location
- Heading
- Speed
- Altitude
- Range
- CPA

Used once and then only used if changed

- ESM
- IFF values
- IFF transponder status
- Radar status
- Intelligence
- Geopolitical situation
- Origin
- ATO
- Detachment
- Split
- Merge

Predictor Agent Competing Models

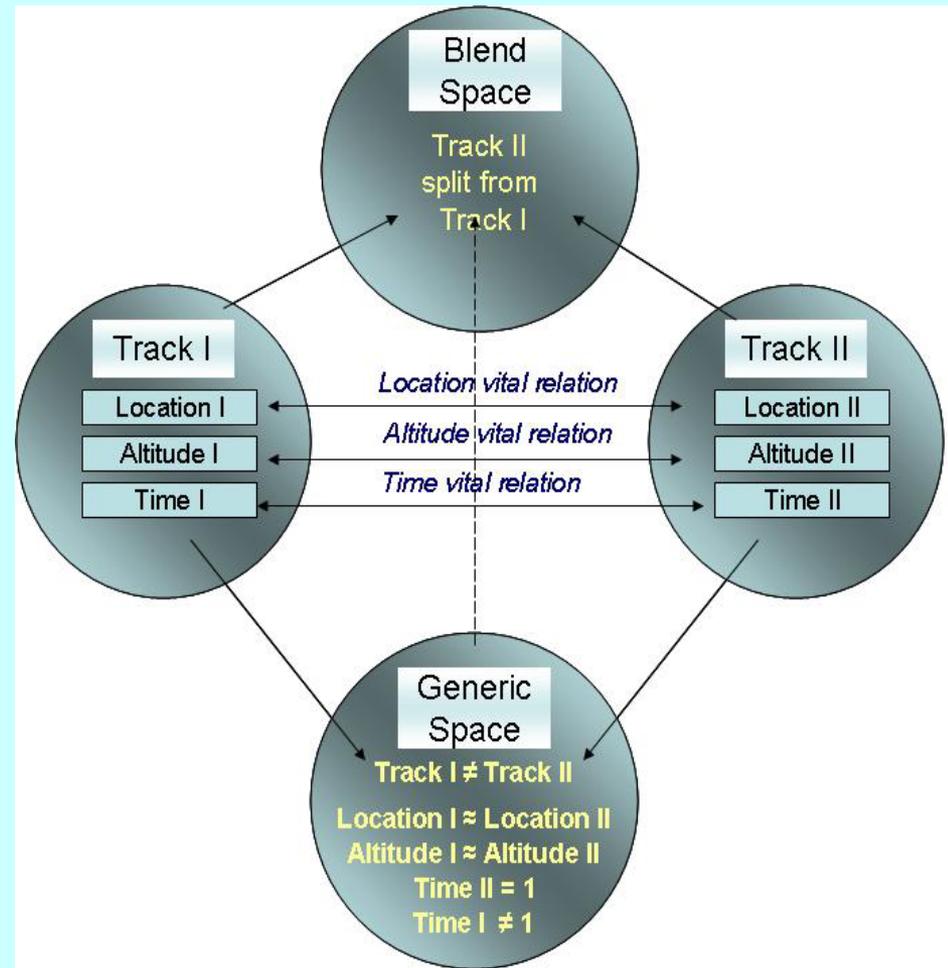
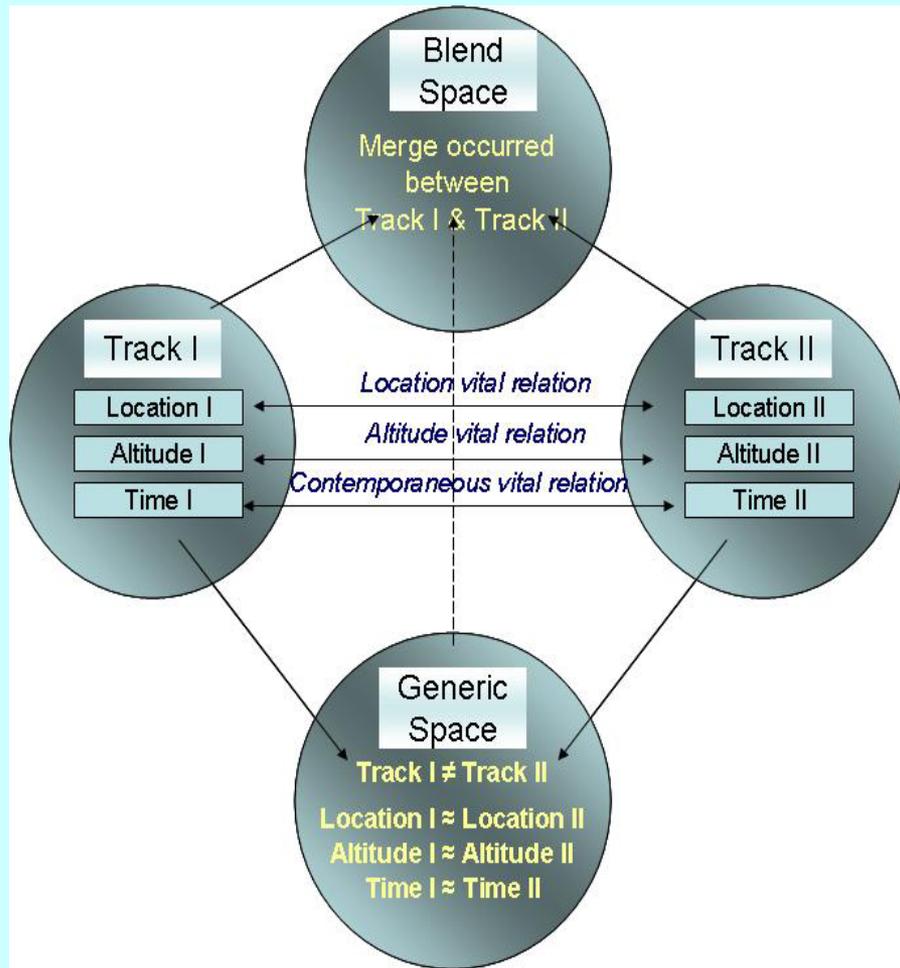
- Each identity is defined as a ticket.
- Tickets find a weight for each identity.
- Highest-weighted identity becomes the active model.
- Reactive agent connectors set the frames of these tickets.
- Also there are independent tickets: CPA calculator, IFF Evaluation, split detector, etc.



Regional Agent Activities

- ❑ Detecting merge activity
- ❑ Detecting coordinated detachment activity
- ❑ Detecting snooper coordinated attack activity
- ❑ Determining threat level

ADL Simulation and Blending



Is air defense too simple for conceptual blending?

- ❑ This is pretty far from the original use of conceptual blending to explain linguistic metaphors.
- ❑ In particular, you need to blend rather different mental spaces to get some power -- here we're blending the same kind, track data.
- ❑ Thus it's more appropriate to view this as inheritance rather than reasoning by analogy (which is what blending is).
- ❑ Inheritance: For every X and Y and for some Z ,
 $p(X, Y) \leftarrow r(X, Z), p(Z, Y)$.

The Bayesian (Rowe) simulation

- ❑ Neither of the previous simulation learn much from experience.
- ❑ A simulation could keep statistics from exercises to learn what clues signal hostile behavior. Use final assessment of a track and find what clues appeared in the course of the track.
- ❑ Such a simulation could be quite simple since it wouldn't need a lot of initial knowledge.
- ❑ Bayesian reasoning is the simplest way to implement such a learning system.

Naïve-Bayes odds calculation

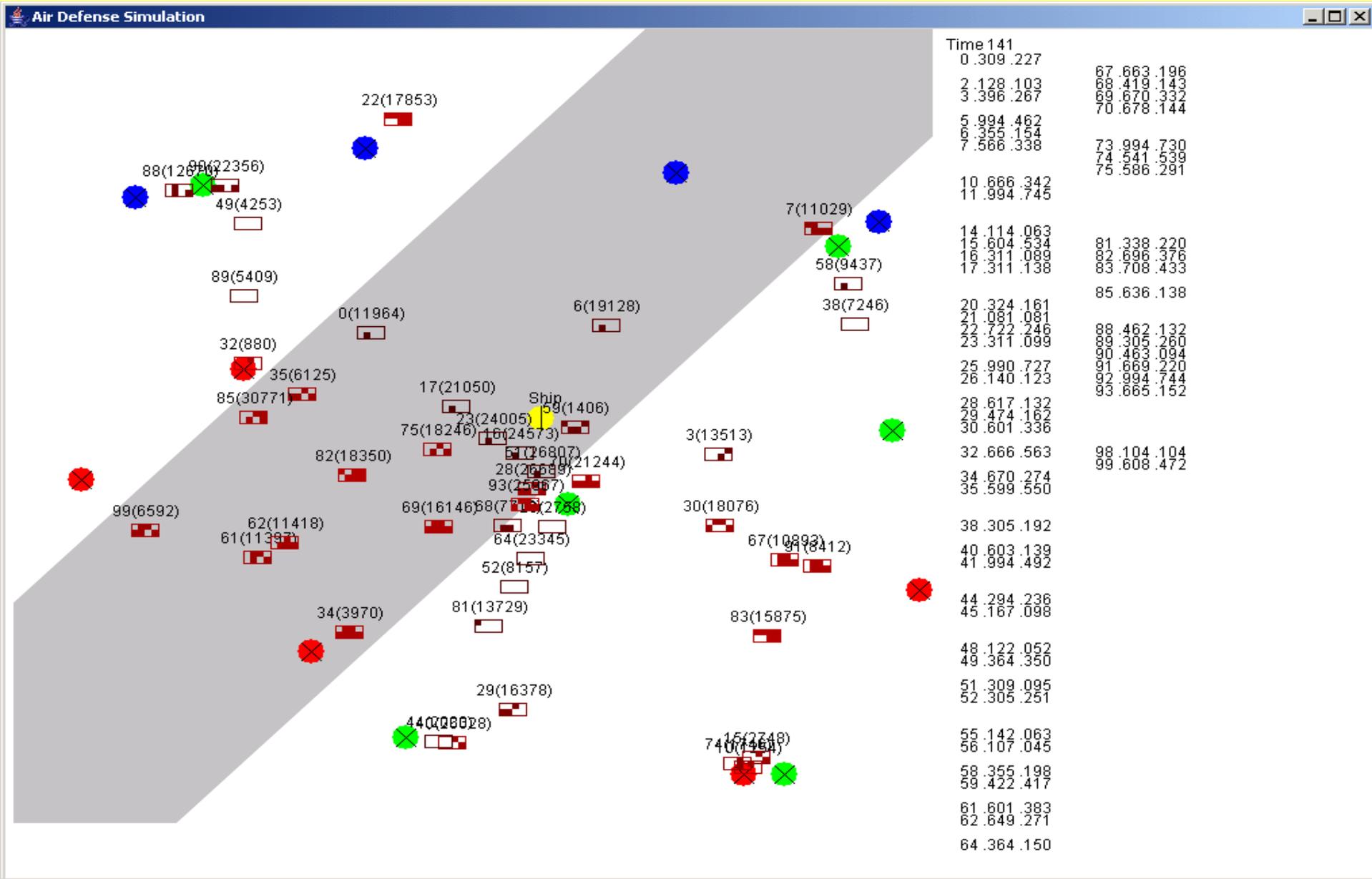
$$o(H | (E1 \& E2 \& \dots \& En)) = \\ [o(H | E1) / o(H)] * [o(H | E2) / o(H)] * \dots * [o(H | En) / o(H)] * o(H)$$

This gives a formula for the odds that an aircraft is hostile (H) given evidence E1, E2, etc.

Here o represents odds or $p/(1-p)$ and “|” means “given that”.

Improvement if there are many time steps: Take to $1/M$ power each of the bracket ratios, where M is the “time window”. This means updates don't change values as fast.

The Bayesian simulation interface



Testing of the Bayesian simulation

- ❑ We generated a variety of test scenarios involving civilian, friendly, and hostile aircraft. We can easily do > 100 aircraft at one time.
- ❑ Tracks could be scheduled flights (some originating outside radar range), military reconnaissance, "snoopers", outright attacks, and hijacked civilian aircraft.
- ❑ Results showed the system could improve with experience as it learned clues.
- ❑ Results showed it could learn how to respond to a new threat it had not seen before, the hijacked civilian aircraft, when first trained on scenarios without it.

Conclusions

- ❑ Occam's Razor applies: Bayesian simulation seems to do almost everything the Ozkan simulation does, in 60 times less code. Thus we should prefer the former to automate air defense (but not to study it).
- ❑ Air defense may be too far from linguistics, the original domain for conceptual blending.
- ❑ The Calfee simulation addresses a different problem, of modeling personnel. But Bayesian simulation suggests automating much of what those 11 people do.
- ❑ Bayesian simulation requires good training of program, which may be hard to set up.
- ❑ Bayesian simulation can be fooled by deliberately deceptive enemies, but so can people.

Automation reduces the need for training

- ❑ If we can significantly automate parts of air defense, it simplifies the tasks of the remaining personnel.
- ❑ That means less training is required.
- ❑ Thus the goal of Mike Zyda's USC-ISI group is self-contradictory: If we can develop wonderful virtual environments for training, we can usually automate the tasks taught and have no need for training.
- ❑ Exceptions would be skills requiring human judgment -- human vanity exaggerates their extent.
- ❑ But human judgment shows many suboptimal biases.
- ❑ The USS Stark and USS Vincennes incidents illustrate that people can have poor decision-making in air defense -- a computer might manage it better.