# Understanding Adaptability: A Prerequisite for Effective Performance within Complex Environments

Visualization Tools to Adapt to Complex Military Environments

Mike Barnes, John Warner, David Hillis, Liana Suantak, Jerzy Rozenblit, Patricia McDermott

# VISUALIZATION TOOLS TO ADAPT TO COMPLEX MILITARY ENVIRONMENTS

Mike Barnes, John Warner, David Hillis, Liana Suantak, Jerzy Rozenblit and Patricia McDermott



## ABSTRACT

*This chapter addresses adaptation to dynamic, novel and uncertain military environments. These environments require a shift from a maneuver warfare paradigm to an asymmetric world where shifting alliances, questionable civilian loyalties, opaque cultures, and the requirement to maintain peace one day and combat the next makes for a particularly confusing situation. This new warfare paradigm requires adaptation to an uncertain, complex environment.*

*The initial section discusses a general cognitive model of visualization called RAVENS and its importance for adaptation developed specifically to address complex military environments. RAVENS posits that humans are inherently flexible decision makers and situation awareness depends on the ability of humans to create narrative visualizations that capture the overall context of complex military environments. Using the framework as a guideline, we will examine two important visualization research*

**Understanding Adaptability: A Prerequisite for Effective Performance within Complex Environments**
**Advances in Human Performance and Cognitive Engineering Research, Volume 6, 73–113**

**ISSN: 1479-3601/doi:10.1016/S1479-3601(05)06003-0**

*programs whose purpose is to allow military operators to rapidly adapt to volatile situations. The first program investigates cognitive effects such as the framing bias and their possible interactions with a variety of display concepts during a series of missile defense simulations. The experimenters presented risk as a spatial representation of uncertainty and target value that emphasized either expected population lost or expected population saved. The second program investigated the feasibility of using visualizations generated from Scheherazade (a coevolutionary algorithm) to aid MI analysts in predicting emergent tactics of terrorist groups during urban operations. Finally, we discuss the value of these approaches for providing coherent narrative understanding as called for in the RAVENS model.*



## 1. INTRODUCTION

MG Robert Scales has published a sobering article on the nature of future conflicts entitled the *Adaptive Enemy* (Scales, 2000). What makes Scales' article particularly disturbing is that it was written before September 11th. He points out that the enemy does not need to defeat us on the battlefield. The enemy simply can outwait us, and take advantage of our vulnerabilities. The technology we possess is a two-edged sword; intelligent adversaries can use our requirements for information dominance to create deception and chaos because they can operate in smaller units and often attack unpredictably making our intelligence efforts difficult at best. Because they do not depend on a well-defined doctrine and order of battle, the simpler nature of their tactics and planning cycles makes their actions particularly difficult to predict.

This chapter addresses adaptation to this new military environment. The authors will cover a variety of visualization concepts and tools whose purpose is to improve the soldier's ability to understand and adjust rapidly to volatile, often unpredictable future military missions (Barnes, 2003; Warner, 2003). These environments require a shift from a maneuver warfare paradigm to an asymmetric world where shifting alliances, questionable civilian loyalties, opaque cultures, and the requirement to maintain peace one day and combat the next makes for a particularly confusing situation. Interestingly, this shift in paradigm parallels a shift in the scientific paradigm from mechanistic and deterministic perspectives to an emphasis on uncertain and complex processes.

This new warfare paradigm requires adaptation to an uncertain, complex environment. Over a long period of time, humans can adapt to the harshest and most inhospitable circumstances. However, the types of situations we will discuss are those that change so rapidly that any one strategy is bound to fail, especially in situations where there is no strictly predictable outcome. The term uncertainty indicates that there are *n* options with varying degrees of certitude whereas complexity refers to situations where the number of possible options is unknown or the situation is so volatile that there is no exact probability structure. It is the latter two paradigms that this chapter is particularly interested in addressing since emergent tactics and behaviors that occur in terrorist environments are precisely the events that pose the greatest challenge in the current world situation (Bar-Yam, 1997; Waldrop, 1993).

The initial section discusses a general cognitive model of visualization and its importance for adaptation developed specifically to address complex military environments (Warner, 2003). Visualization in this sense implies more than being able to image a particular situation. It also suggests a creative process where various options are considered and their consequences and likelihood are considered as well (Barnes, 2003). Experts can often visualize a correct solution to a familiar problem set quite rapidly. Unfortunately, there are many situations where such a process is not sufficient (Klein, 1998). The next section addresses cognitive biases and their impact on the human ability to visualize uncertain environments. The purpose of this ongoing research is to develop better visualization techniques for humans in uncertain – high-risk environments wherein the decision cycle is too rapid to even consider traditional decision theoretic analyses. The final section discusses Scheherazade, a visualization and simulation tool developed to aid intelligence analysts in anti-terrorist planning. Scheherazade (Barnes & Hillis, 2003) uses the complexity inherent in coevolutionary algorithms to allow the intelligence analysts to visualize unusual, difficult-to-predict behaviors that emerge out of the adaptation process. It represents a new generation of visualization tools whose purpose is not to predict but to impart insight and a better understanding of the complex cycle of adversarial behaviors.

Thus, in the course of this chapter, we will move from an overarching theoretical framework to specific problems the framework might address to a very specific (and detailed) tool that tries to develop in the analyst those skills encouraged by the framework. The central theme to the whole chapter is the idea that adaptation requires a coherent representation of key relationships in an otherwise complex and uncertain environment. People, we will claim, have a natural sense-making ability that helps them when these key relationships are identified; therefore, the important task for visualization is to help provide what the soldier needs to identify these key relationships and to provide richer mental models than cognitive limitations might otherwise make available.

## 2. RAVENS: A GENERAL COGNITIVE MODEL

In the modern battle space, the commanders and their staff face critical decisions in dynamic and uncertain environments characterized by a wide possible range of threats. Further, each operation may be quite different from other operations, even those with similar goals, making expertise hard to grow. These new environments often involve unfamiliar cultures in which some may be allies and some may be foes, low-density languages, and rapid deployment. This is a difficult combination for developing reliable intelligence. These environments are what we will call ''data-rich but knowledge poor'' in that there is a lot of complex information that can be gathered, but building useful knowledge out of all that data is not so simple. The military is getting very good at collecting and distributing large amounts of information. The more critical task of achieving situational awareness (SA), making sense out of that data for the decision maker, is much more challenging.

The goal of this section is to outline a framework for approaching the problems raised by SA for missions, such as the military now faces, that require both rapid knowledge building and previously uncommon levels of adaptiveness. An important aspect of this framework is to propose that the best sort of visualization will help decision makers maintain dynamic and adaptive models of the situation. For this reason, there will be an emphasis on storytelling (reflecting narrative cognition in the human) as a powerful human sense-making ability that needs to be cultivated in the human decision maker and supported by information automation.

The framework is referred to as the RAVENS, which stands for ''Rapid Adaptive Visualization of Emergent and Novel Situations.'' The acronym itself refers to the goal of the framework as much as the application of it. It should be noted that ''visualization'' should be interpreted broadly as any means that helps one clearly represent and think about the battle space. Visualization in this sense encompasses any or all perceptual modalities or even semantic cognitive processes. Unfortunately, while the term ''visualization'' seems tied to the visual modality, there are no really good alternate terms. The name RAVENS is also symbolic, given the military intelligence flavor of this work, in that RAVENS is associated with military intelligence and with understanding the situation. In Norse mythology, Odin sends out two ravens, named Thought and Memory, to gather information from all corners of the world and report to him each day.

The framework being described here is based on a theory of problem solving for the intelligence domain developed by Warner and Burnstein (1996) that proposes there are four modes of thinking each of which results in different transformations to information to turn it into aggregated information, meaning, and pragmatics. These four modes are constructive, diagnostic, reactive, and explanatory. This framework stands on its own without reading Warner and Burnstein. However, the framework is based on the explanatory mode, which becomes very important in sense-making.

### *2.1. Adaptiveness and Situational Awareness*

From the perspective of RAVENS, an adaptive commander or other planner is someone who can respond flexibly not only to changes in the dynamic situation, but can also adapt flexibly to knowledge about the situation. This necessarily entails more than having a handful of loosely associated facts, it requires being able to understand complex relationships in the situation and to be able to revise one's overall model of the battlespace when needed. Perception and understanding of dynamic and uncertain environments cannot be based on template solutions or fixed expectations. This calls for more than a picture of the situation. SA requires something like a dynamic, even evolving, model of the situation. These ideas have important implications for "adaptive visualization." The visualizations being used to help the decision maker must either be model-driven themselves or, at the very least, support the human in being able to think in terms of some kind of dynamic mental model. What is visualized must help the overall situation make sense to the decision maker so that the decision maker knows what aspects of the situation are relevant and most important. This point will be illustrated in the sections that follow. For example, experiments with visualizations that help soldiers overcome cognitive biases in assessing risk will be described. Still later, Scheherazade, a simulation-based visualization tool that allows the analyst to explore emergent consequences of complex interactions that might be difficult for the analyst to represent on their own will be described.

You will notice above the recurrence of terms that allude to "situational awareness." Endsley (1995, 2002) defined SA as:

> … the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.

Notice that this definition encompasses not just awareness of what is going on in the battlespace right now, answering questions like ‘‘who?’’, ‘‘what?’’, ‘‘how many?’’, ‘‘where?’’, and so on. SA includes the requisite comprehension of the big picture to make projections about the future, answering questions like ‘‘why’’, ‘‘what next?’’, ‘‘what if?’’, and other longer range planning questions. This scope is important because the key to being adaptive is to achieve coherent yet flexible understanding of the whole situation. What underlies achieving SA in novel environments is sense-making (Weick, 1995). While sense-making, figuring out how elements of the situation fit together so that the overall situation is coherent, is a uniquely human ability, information visualization can aid this activity.

### *2.2. Adaptive Sense-Making: The Value of a Good Story*

The notion of ‘‘adaptive visualization’’ turns on the concept of helping the decision maker adapt as the situation changes. At a minimum, this requires that the support technology makes the decision maker aware that the situation has changed. However, the goal would be to support those cognitive processes in the human that drive the sense-making ability. For that reason, it is important to look at the cognitive processes, which RAVENS assumes underlie SA and sense-making.

The human decision maker must make sense out of any situation, especially when uncertainty is high. As humans, we are compelled to engage in ‘‘sense-making,’’ to take any surprise, anomaly, or unexpected behavior, and make it fit smoothly with what we know. In traditional war, this is easier because the expectations and set of alternatives are much narrower. However, when key relationships are not always clear, particularly causal relationships, we resort to a cognitive process referred to here as *storytelling* or sometimes, more formally, as *narrative cognition*.

This framework takes a constructivist approach to SA. Constructivist approaches to psychology assume that we do not simply sense and perceive photographically veridical exemplars of ‘‘what is out there’’ and store them in memory, like photos in a filing cabinet. Rather, we extract essential features of reality and use them as a ‘‘good enough’’ version of the original (Fiske, 1993; Kintsch, 1998; Klein, 1998). This abstracted structure (sometimes referred to as a *schema*) is then imposed on reality where it acts as a filter for both perception and memory. In our attempts to maintain coherence, either perception or memory may be changed to get the ‘‘best fit’’ to the overall structure. Thus, both perception and memory, the things we think about, are constructed in real time, not just stored and retrieved. Storytelling is, in fact, the ultimate constructive process as the result is not just a representation, but also a constructive simulation. The advantage of a constructive cognitive system is that it is efficient – a lot can be done, given sufficient experience, with expectations about the world without verifying every pixel of reality.

Narrative cognition is a process by which humans flexibly and adaptively build knowledge about and make sense of the world, to transfer this understanding to others, and, in so doing, allow others to do a better job of reading our minds (that is, inferring our intentions). Storytelling can be a tool for sense-making in the face of less-than-perfect and incomplete data because it does not depend entirely on truth and logic. Rather, it fits the data to what you know and have experienced and even to your sense of self in the world. This can have some drawbacks, but it is not a flaw as some have suggested (most notably, Kahneman & Tversky, 1973). Brains are not limited to syllogistic and Boolean rationality as digital computational devices are and, as a result, people do not make their decisions on the basis of well-formed if–then statements and calculated probabilities. This, we would argue, is mostly (but not always) a good thing. It really depends on what your objective is and what data or information you have to work with. However, as you will see in Section 3, the brain will tend to fill gaps based on expectations and even biases that, though efficient, may lead us to counterproductive analyses.

The idea that human storytelling is a powerful and even preferred means of understanding situations and learning about the world is not new. It has been taken quite seriously by the humanities, including history, anthropology, law, education, management, artificial intelligence, human–computer interface design, and constructionist-based psychotherapy (Bruner, 1990, 2002; Harre & Gillett, 1994; Hirokawa, DeGooyer, & Valde, 2000; Klein, 1998; Murray, 1995; Pennington & Hastie, 1993; Rossiter, 1999; Shank, 1990; Snowden, 2000; Weick, 1995).

The RAVENS framework considers ‘‘story’’ or narrative to be a cognitive structure, like a schema, but model-like so that relationships between people and events are captured holistically. Stories, as knowledge structures, are rooted to a particular situation; they reflect time, place, intent, behavior, and goals. They also provide a point-of-view (POV) or experiencing self, which is as important to gathering intelligence as it is to any situated work of fiction. This is what makes storytelling pragmatic (Fiske, 1993; Warner & Burnstein, 1996). We will thus define this story representation as a thematically organized and coherent representation or model of actors, roles, behaviors, goals and their relationships to one's self or POV, and one's previous knowledge of the world.

We will first elaborate a few of the terms used in this definition so as to be as clear and concrete as possible. *Thematically* refers to an essential property of stories, that they have a central subject or "thing in the world" that they are modeling. It is this theme that gives the relations, particularly causal relationships, between the essential elements (actors, their roles, their actions in those roles) not only meaning but consequences in the model. *Coherent* refers to the result of what has come to be called sense-making. Sense-making is the ability to take unexpected events and normalize them with our already existing knowledge and experience. This may be done by altering the model or by changing our expectations. Finally, the terms *self* or *POV* (these are interchangeable) refer to the "narrative center of gravity" (Bruner, 1990) that drives the model, provides the emotion, defines the consequences as well as the goals, and provides this knowledge "structure" with its flexibility and dynamism.

Coherence or sense-making relies on this "best-fit" idea implied by constructive theories. As long as what is experienced fits within a tolerable range of variability to our expectations, it will be incorporated into our model of the situation. This means the situation will continue to be coherent and will not require a lot of extra cognitive resources to maintain coherence. However, in the types of missions currently being seen, where expectations are so different from outcomes, they can no longer be accommodated by the same model. It is here that storytelling is invoked to generate either new relationships or a new model. This is what allows the human to be adaptive.

Storytelling is a tool for and invoked by dealing with unexpected outcomes. It is no accident that stories as entertainment turn on conflict and difficulty. Narrative structure, because it is richly semantic (carries meaning) and pragmatic (relates that meaning to one's goals, POV, and emotions), allows one to make projections about the future (plans) and decisions. That is, narrative is an ideal way to discover and represent causal chains. When something unexpected occurs, one of two things can happen, depending on factors such as the importance of the event (pragmatically), how great the mismatch is with expectations, the consequences of the difference for achieving the goal, and other factors. One possibility is that the narrative structure imposes itself on the unexpected outcome or event and forces a fit in the model. This is always the least costly solution, in terms of cognitive resources, and thus is probably always the brain's default solution. The second possibility is that the narrative schema shifts dramatically. This is much harder – we usually have to sit down for a while after this happens and may even experience it with some amount of stress and alarm.

Stories are easily shared, if imperfectly. That is, to the degree that our models of the world overlap and are mutually coherent, stories can be passed from one person to another. The advantages of this are enormous. It means that one person can pass not just information but his/her understanding of that information to another. This is no small accomplishment. Because stories are constructed not only from actual perceptual information, but also from one's personal experience and POV, each person's story is somewhat idiosyncratic. This range of individual interpretation is somewhat bounded by culture and domain, but not as much as one might think. We often misunderstand each other, but it is in fact much more surprising how often misunderstandings do not happen. If someone narrates a situation for another and they share enough experience, culture, and domain in common, the person receiving that story should be able to go away with a "good enough" understanding of the situation as the narrator understands it and may even be able to "read" the narrator's mind. That is, the listener may be able to understand the intent of a request from the narrator even though the request lacks sufficient detail in its surface form (Klein, 1998). This is very hard to do with almost any other kind of information presentation.

However, this story sharing and mind reading is a double-edged sword for adaptability. We said that this depends on overlap between our models of the world and, thus, overlap in our perceptions and experience. If, however, we are involved with another culture where there is much less overlap in perception and experience but we assume they should see things the same way, we are not going to be adaptive at all. That is why the mental model needs to be constructed from data, not from experience alone.

### *2.3. Sense-Making: Building Understanding in the Decision Maker*

Above, the case was made for storytelling as a unique human ability for the essential task of making sense of the world in order to achieve SA. It was suggested that passed-on culture and domain stories make it easier to accomplish this sense-making task. Nowhere is sense-making more vital than in military intelligence. However, this is the task that has become much harder when operations and their contexts change from mission to mission. Can decision support technology and human engineering be used to support storytelling in order to provide more rapid coherence and SA?

One fruitful area to consider is the idea of narrative information visualization (Gershon & Page, 2001). Most information in a database, which is where the visualized information is drawn from, tends to be typically organized in an object-oriented manner. This hierarchical inheritance structure is useful for the kinds of knowledge that can be agreed upon, objects and ideas that are well defined within the culture or domain, such as general knowledge and domain knowledge. However, it is not the structure that will help to construct a story. What is needed is a way of organizing information thematically. What is the current situation about, who are the major players, and what roles do they seem to be acting out given their behaviors?

The key to aiding the human is to provide him or her with visualizations that encourage building narrative out of the data, or at the very least providing key relationships in a way that the human will understand that they are important and build on them. This suggests that presenting data in a model-driven simulation, perhaps with game-play interaction, would be a beneficial way to represent the data one needs to incorporate the situation. These simulations would themselves be organized by assumptions about theme and roles of different actors and might be able to provide collaborative story comparison among staff or different levels of echelon. This kind of story-like external representation and simulation tool or tool suite could prove invaluable to the human decision maker, particularly for aiding generations of multiple hypotheses.

Simulation tools, which have recently become more of a focus for training and system design, are equally important as tools in the planning process in the battlespace, helping to make sense of the unfolding situation, and actually testing coherence. In planning a simulation during scenario design, generating stories helps one consider all the possible interactions between agents and actions. This has been realized for some time in user interface design with a movement sometimes referred to as *scenario-based design* (Carroll, 1999). For example, if one is trying to design a new information interface, he or she might consider the role the human is going to play, what information he or she, or some subject matter expert (SME), think is relevant and what use they are going to put it to. What kinds of things might actually happen when the user tries to use this interface? What other thing is the user trying to do or think about at the same time? Is trying to do certain things at certain times going to end up frustrating the user because of the different ways tasks are set up – that is, how does the user feel and what is the user's POV?

The same thing applies to a complex simulation. Simulations of human systems in dynamic environments are not always captured well by parameter lists and algorithmic weightings. They are also not captured in static scenarios. Simulations need stories as well, not unlike the writers' conference that might be done before creating a television series and its progress over the next season. In this case, a mixture of computer programmers, soldier SMEs, and even storytellers collaborate to evaluate and elaborate the nature of the thematic organization of this simulation. Simulations, like stories, need to be imbedded in a domain and a context (Snowden, 2000). General simulations are not likely to capture the kinds of interactions that we see as the context of each operation changes. When things change, the simulation has to adapt, just as the decision maker's thinking has to adapt. The simulation will never be able to provide the decisions, but it must be able to portray the changes those decisions are going to be based on.

However, there still seems to be some limitations to the dynamics of these solutions. To accomplish narrative visualization of an unfolding situation, you would like the model to evolve just as the real situation does and adapt to those changes. Evolutionary algorithms seem to show a great deal of promise here. These algorithms (see Section 4 for an example) might be utilized to model self-organizing social networks and dynamic probability spaces, creating much more adaptability than current network approaches that still rely too heavily on pre-determined rule sets.

Immersive simulation is a technique that has been explored for training situations for some time. Immersive simulations are simulations that utilize all the sensory modalities to create as realistic an experience as possible. Movies on a large screen in a dark theater with surround sound are an immersive simulation. In fact, the military has turned to movie studios, entertainment engineers, and video game developers to help them explore ways of building experience. This allows soldiers to explore possible outcomes and relative risk while experiencing the more pragmatic relationships in a way that would be hard to capture without that degree of immersion. Because of the inherent "storytelling" involved in setting up an immersive simulation, gaps in knowledge might be quickly identified.

The upshot is that there is a need to provide visualizations that can put the human in a narrative mode while planning, can provide model or story-based organizations of data, and can tell the decision maker something about what the data mean to them and to their mission. There is also a need to capture the dynamic nature of the situation as it evolves in the models. If the human and the decision aid share similar models, the human will be able to see much more easily what relationships fit and what relationships do not. That will make for an adaptive planner. The RAVENS framework provides an approach for exploring (research) and developing such tools and for further investigating how human narrative cognition is used in sense-making.

## 3. VISUALIZING RISK

In the previous section, a broad conceptual theory was presented that suggests there is a need to create coherent relationships between knowledge and information as a foundation for rapid adaptation that can reduce both uncertainty and complexity. In this section, a particular visualization problem and some empirical attempts to provide narrative coherence will be examined. The problem presented here is that of visualizing uncertainty and risk. Adapting to uncertainty requires a coherent mental representation of the situation. One of the key problems in presenting information about risk and uncertainty are the cognitive biases that decision makers bring to the task (Wickens & Hollands, 2000; McDermott et al., 2002). It is also true that experts can make effective decisions rapidly if they are familiar with their environment (Klein, 1998; Lopes, 1986). Unfortunately, as we have pointed out in the discussion of RAVENS, in part because each situation tends to be different, we will not always be able to have experts with long experience making decisions as might have been possible in the latter years of the Cold War. The question that this series of experiments addresses is how to display uncertainty and risk in order to enhance the soldier's understanding of the military situation so that the resulting decisions resemble the expert and not the biased decision maker. Doing so requires modeling complex relationships in a way that allows the soldier to understand not just what is happening now, but what is likely to happen in the future.

Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni (1999) posit that the reason for many of the observed deficiencies in decision making is an impoverished mental model of the decision space. Humans partition their decision space into a perceived number of options with weightings representing the certitude of each option. The problem according to these authors is that the decision maker often perceives fewer options than in fact exist in that space. For example, they examined the famous "Monty Hall" problem of a contestant choosing a prize behind one of the three doors. In this case, after Monty shows a contestant that the prize is not behind one of the two doors they did not choose, the contestant usually stayed with the original choice because they assumed it was equally likely that the prize was in either of the options left. However, analysis indicates that counterintuitively, there are six possible outcomes: 2/3 of which favors switching doors (the answer hinges on the number of cases in which the prize is not behind the chosen doors (see Johnson-Laird et al., 1999, p. 82). This example almost has an air of trickery about it but it does illustrate that as a probability space becomes more complex, humans tend to simplify the mental model necessary to visualize the uncertainty space. The purpose of visualization aids is to alleviate the computational and visualization requirements humans encounter when making decisions in complex environments.

The initial experiment in this study attempted to simplify the veridical visualization process for the operator while ensuring that the display was not subject to well-known decision biases. The experiment was set in a national missile defense (NMD) paradigm with a trained but non-expert subject pool (i.e., the actual system was not built and the Air Force operators had initial training in order to help design the evolving system). Based on the frequency visualization literature and discussion with engineers, the probabilities were recast as number of possible "leakers." This allowed the operators to visualize the probability of an enemy missile not being intercepted (leakers) as a concrete frequency (i.e., 20 possible leakers per 1,000 represented a 2% expected leaker rate). Unfortunately, this measure also involved the possibility of the operators being influenced by a well-known decision bias – the framing effect (Tversky & Kahneman, 1981). Framing theory predicted that the operator decision strategy would change depending on whether outcomes were presented in terms of possible losses as opposed to possible gains. We investigated the possibility of a framing effect and the efficacy of giving the operators leakers as opposed to probability of success representations by testing 16 operators during a game-like simulation of an ongoing missile attack. The operators made missile allocation decisions and answered SA probes in a time-constrained setting. The results did not indicate a framing effect on decision strategy; however, representing the efficacy of the allocation plan in terms of expected frequency of leakers did improve the operator's SA score. Presumably, operators were not influenced in their decision making by how the uncertainty information was represented (frequency of possible leakers versus probability of success display) but they were able to answer probe questions more accurately and more rapidly if they were presented as leakers as opposed to probability information (Barnes, Wickens, & Smith, 2000).

The follow-on experiments investigated format variables and in particular how they might interact with cognitive biases such as framing (Barnes, McDermott, Hutchins, Gillan, & Rothrock, 2003; McDermott, et al., 2003). In general, risk information is better displayed with a graphical format (Smith & Wickens, 1999). However, there are a number of experiments that attest to superiority of textual formats. The crucial factor seems to be whether the presented information captures the dynamics of the process being represented (Meyer, Shamo, & Gopher, 1999; Wickens & Hollands, 2000). Thus, the efficacy of the format depends on the specific decision environment and on task demands as well as type of format. The NMD situation is both an important military problem and a case in which the physics of the sensors and interceptors can be used to make fairly accurate predictions about threat probabilities, value of the intended targets, and the probability of successful interception. A simulation environment in which the operator's task is to ensure proper allocation of defensive missiles was created. The simulated display indicates both the state of uncertainty and the value of the defended targets (U.S. cities) during an ongoing missile attack. The problem space is volatile because probabilities change as events unfold and interceptors succeed or fail and additional threats are discovered by the defensive systems. The experimental constraints required the operator to respond within 10–15 s of receiving a probe emulating the rapid decision cycle of many military situations.

### *3.1. Simulation Tool*

Before conducting the simulation experiment, the type of graphical format that would be most effective for this type of environment needed to be determined. Two types of presentation format, square or circular, and two presentation modes, risk information being separated or integrated (Gillan & Hutchins, 2002) were examined. The integrated displays were also configural in the sense that while the separable displays showed two dimensions: probability of interception and population of defended city, the combined displays area showed the emergent dimension of overall risk (expected population lost). The square display in particular was similar conceptually to the configural displays evaluated by Bennett, Toms, and Woods (1993) for nuclear power plants.

A pilot study was conducted with 24 college students to determine how the graphics should be presented. The goal was to determine which display elements were most easily understood and best supported SA. The pilot was a complete $2 \times 2$ within-subjects factorial design consisting of four trials. Each trial consisted of various presentations of the same style of display. Every display had six graphs of the same type, with each graph representing a city, A through F, respectively. Trial order was counterbalanced. Subjects were shown various city displays that varied in overall risk, population, and probability of being successfully attacked and were asked to respond to printed probes querying the subject on various attributes of the cities. Subjects responded by clicking an ''answer'' button on the display. Response times were recorded from the moment the ''ready'' button was clicked until the moment the "answer" button was clicked. Accuracy of the typed responses was also recorded. This procedure was repeated for each display combination. The results indicated that the combined (i.e., integral) displays resulted in both more accurate and faster performance. This was a speed–accuracy trade-off with square type of format. The square format resulted in slower, but more accurate performance. Because the difference in latency was fairly small and accuracy was considered the more important criterion, the square format was used to evaluate configural representations in the simulation experiment.

The purpose of the simulation experiment was to understand the interaction of cognitive and perceptual display factors on both the operators' missile allocation decisions and SA performance. The Barnes et al. (2000) experiment failed to show the well-documented framing effect (cf., Mayhorn, Fisk, & Whittle, 2002). One major difference between the two paradigms is that the students in the Tversky and Kahneman (1981) study made decisions based on overall risk whereas our study was concerned only with how uncertainty was represented. In military environments, overall risk is the parameter the commander uses to make decisions (i.e., the military value of making the correct or incorrect choice weighted by its probability). In the present experiment, overall risk was shown in two presentation modes, integral or separable, and two formats, alphanumeric or graphical. Also, risk was framed in two conditions: possible population saved or possible population loss in terms of the current allocation plan. Fig. 1 shows risk information in a separable display with gain information – values of a successful decision. The value of the target was defined in terms of the

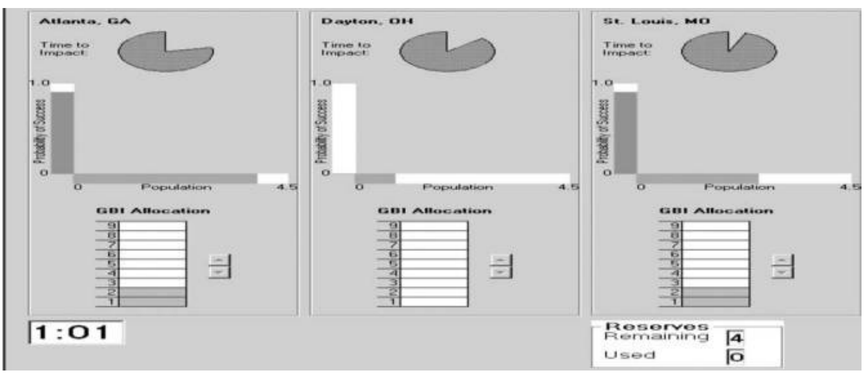


*Fig. 1.* Visualization Aid Showing Separable Information in Terms of Probability of Successful Interception and Value of the Target in Terms of Population.

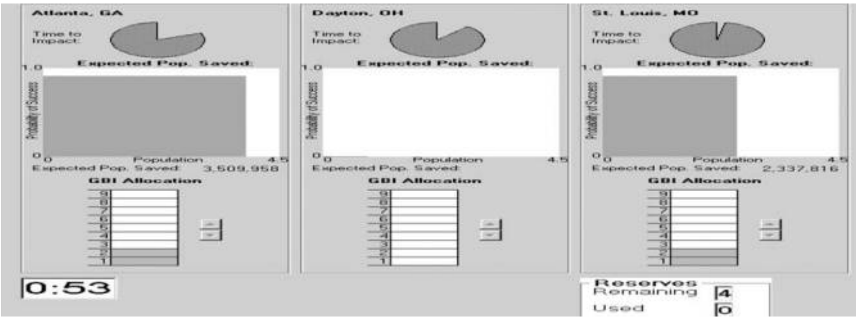


*Fig. 2*. Display Showing Overall Risk Area as Expected Number of Civilians Saved (Gain Information).

population of the city being targeted. The probability represented the probability that one of our missiles would intercept the incoming missile successfully. The ground-based interceptors (GBI) were inventory on hand and the remaining portion of the circle was time left to make an allocation decision.

Fig. 2 shows similar information; however, the expected value of this particular strategy is shown as the combined area of probability of success and population (overall risk). (Note that both displays are somewhat distorted to fit on the page.) In terms of experimental conditions, this display is graphic, integral (combined), and framed in terms of expected population saved.

The other framing condition shows the same information in terms of possible losses. The *Y* dimension is the probability that the GBI will not intercept the incoming enemy missile whereas the *X* information is still the population of the city being targeted. For the integral display, the emergent area is now the expected number of civilian casualties. The emergent areas in the loss displays are shown in red so that the subjects would not be confused between the two displays (green – number saved versus red – number of casualties). It is important to note that loss information is simply one minus the probability of success; the same information is imparted with a loss or gain display. However, framing theory suggests that operators should respond differently to the two displays. The prediction important for this paradigm (which differs from the original Tversky & Kahneman (1981) – in that there are no sure gain decisions) is that the operator will be biased toward avoiding losses versus the same decision framed to improve potential gains. This prediction follows from the steepness of the loss curve compared to the gain curves in Prospect theory as well as loss avoidance behaviors noted in other decision-making paradigms (Tversky & Kahneman, 1981; Shafir & Tversky, 1995). One prediction is straightforward when the information is given in terms of expected losses; the operator should take more missiles out of reserves than when the same information is given in terms of possible gains. The other is subtler; situations were set up where small cities were not covered with any interceptor missiles. In half of these conditions, the expected value for allocating a reserve interceptor (GBIs) to an uncovered small city was higher than to put the reserve GBI on one of the other cities. In the other half, the expected value was higher whenever the city was left uncovered (assign no missiles from reserves) and the operator used the reserve missile to protect one of the other (much larger) cities. The loss avoidance thesis would predict that the operator would leave fewer cities uncovered in both conditions when the displays emphasized possible losses versus the same decision situation when the information was given in terms of expected gains. It is important to remember that the expected utility was the same for both the loss and gain conditions and the subject's individual risk biases should not influence the results (they experienced both conditions – it was the relative decisions under the two frames that were being measured).

### *3.2. Simulation Study Methods*

Research was conducted to investigate how the following display parameters affected SA and decision making:

- Presentation mode: Information about city population and probability of success was integrated into a value for expected population saved or kept separate.
- Format: Information presented in graphical or alphanumeric format.
- Frame: Information was conveyed in terms of gains (probability of success and expected population saved) or losses (probability of failure and expected population lost).

The experiment was a $2.2 \times 2$ mixed factorial design with presentation mode as a between variable. Format and frame were within variables and frame was blocked. The levels of format, presentation mode, and frame were combined in all possible ways to create eight displays. The within-subject variables were counterbalanced to avoid sequence effects.

Forty-eight students were recruited from introductory psychology classes at the New Mexico State University. Subjects participated on a volunteer basis. Participants sat in front of the computer monitor and entered their responses with a mouse. The NMD scenario software was developed at Micro Analysis & Design. Instructions were delivered via a self-paced PowerPoint® presentation. Some information and its format were common across the eight displays. The number of reserve GBIs used, the number of reserve GBIs remaining, and time elapsed in the scenario were numerically represented. Each target city had its own area of the display. Within that area, the following were presented (from top to bottom): time until the incoming missile impacts, probability of successfully intercepting the incoming missile, population of the target city, and GBI missile allocation along with up and down arrow buttons that allowed participants to increase or decrease the number of reserve GBIs allocated. The way the remaining information was represented depended on the specific display. The graphic displays have been described above.

*Integral alphanumeric gain.* Information about the GBIs allocated, time to impact, probability of success, and value as well as the value of expected population saved were numerical. The color green indicated the gain condition.

*Integral alphanumeric loss.* Information about the GBIs allocated, time to impact, probability of failure, and value as well as the value of expected population lost were numerical. The color red indicated the loss condition.

Sixteen scenarios were used in the experimental trials. There were four types of scenarios that dictated the number of cities attacked, the size of the cities, and the number of GBIs allocated. Scenarios were randomly matched to displays so that each display had four associated scenarios, one of each type. This was done for the separate (between) condition and repeated for the combined (between) condition.

The scenarios had several factors in common. Between three and five cities were attacked, three or four reserve GBIs were available, and between zero and three initial GBIs were assigned to each city. Each scenario was 2 min long and contained four SA probes and one reserve missile allocation probe. The SA probes were intended to engage the participants, increase workload, and determine how well the displays were understood. The scenario was frozen during SA probes and participants had 10 s to respond by clicking on the appropriate answer. Failure to respond was recorded as a "no answer." The SA probes asked about relationships such as, "Which city has the highest probability of success?" The information on the screen was occluded during the probe so participants had to rely on memory. The reserve allocation probes always occurred at the end of the scenario and lasted for 15 s. Participants were prompted to allocate missiles and the up and down arrows for each city became active. As participants allocated missiles, the information on the display (i.e., probability of success, population saved) was updated dynamically to show the impact of allocations. Participants could continue to change allocations for the full 15 s, which was displayed via a timer countdown.

The session began with a self-paced training presentation that described the participant's role, the displays, and the scenarios. The participant had the opportunity to ask questions. The order of practice and experimental trials was as follows: four practice scenarios (one of each display type), two practice scenarios in the frame that they were about to be tested in (i.e., gain or loss), eight experimental trials in the first frame, two practice scenarios in the second frame, and eight experimental scenarios in the second frame. The entire procedure took approximately 90 min. The practice and experimental trials were presented in random order blocked by frame. At the conclusion of each scenario, the participant clicked on the "OK" button to advance to the next scenario.

### *3.3. Results of Simulation Study*

The dependent measures recorded by the simulation software were: reserve GBI allocation, SA probe accuracy, and SA probe response time. Decision making was assessed via GBI allocations. SA was measured by SA probe accuracy and response time. In addition, risk behavior was analyzed in relation to whether GBIs were allocated to small cities. Repeated measures ANOVA were performed on the decision score (i.e., reserve GBI allocation), accuracy, response time, and small city coverage.

*Decision score*. A decision score was used to measure deviation from normative GBI allocation. This score depended on how the participant allocated his or her reserves compared to an optimal solution.

The decision score equaled the participant's expected population loss minus the optimal expected value where the participant's expected population loss was the sum of the probability of failure multiplied by the population for every city in the scenario including the possibility of a future attack. The same information could have been given from the inverse (expected lives saved), but a common reference measure was used for all conditions. There was a significant interaction of frame by content, $F(1, 46) = 12.484$, $p \leqslant 0.0001$. In the integral displays, performance was better (fewer expected casualties) with the loss (200,000 versus 260,000) condition but in the separable displays, performance was better in the gain condition with no significant main effects.

*SA accuracy*. The SA accuracy measure was an average percent correct for the probe conditions. If the participant did not respond this was treated as an incorrect response. There was a significant interaction of frame by format, $F(1, 46) = 7.163$, $p \leqslant 0.010$. In the graphical displays there was almost no difference between gain and loss conditions but in the alphanumeric displays, SA accuracy was higher in the gain condition than in the loss condition. There was a main effect of frame, $F(1, 46) = 16.642$, $p \leqslant 0.0001$ and a main effect of format, $F(1, 46) = 9.371$, $p \leqslant 0.004$. SA accuracy was better in gain than loss (0.859 versus 0.807) and better in graphical than alphanumeric (0.854 versus 0.813).

*SA response time*. Response time was recorded in seconds. There were four instances in which a participant did not respond and these were coded as the maximum time (10 s). There was a significant interaction of frame by format, $F(1, 46) = 4.594$, $p < 0.037$. In the alphanumeric displays, response time on both gain and loss displays was roughly equivalent. However, in the graphical displays, performance on the gain and loss displays differed with response time being higher in the loss condition than in the gain condition.

*Small cities analysis*. In order to see if participants were risk seeking or risk averse in relation to small cities (i.e., did they have a bias to not leave any city uncovered and avoid a sure loss) and how this was impacted by display type, the coverage of small cities was analyzed. The experiment was designed so that in some of the scenarios the normative allocation involved leaving one small city uncovered. Since the distribution of "uncovered" small cities was not balanced across conditions, the scenarios were split into two types and analyzed separately: those in which the small cities should be covered and those in which a small city should be uncovered.

For the small cities that should have been covered (i.e., defended), there were three effects. There was a two-way interaction of frame by format, $F(1, 46) = 60.926$, $p < 0.0001$. The difference between gain and loss displays was relatively large in the graphical condition compared to the difference between gain and loss displays in the alphanumeric condition. There was a main effect of frame, $F(1, 46) = 124.305$, $p < 0.0001$ such that more small cities were appropriately covered in the loss displays than the gain displays (0.569 versus 0.369). There was also a main effect of format, $F(1, 46) = 24.306$, $p < 0.0001$ such that more small cities were appropriately covered in the alphanumeric displays than the graphical displays (0.507 versus 0.431).

For the cities that should have been left uncovered by the expected value rules, the same effects were found but the strengths of the effects were weaker.

There was a two-way interaction of frame by format, $F(1, 46) = 5.736$, $p < 0.021$. In the graphical condition, there were more cities appropriately left uncovered (0.20) for gain displays than for loss displays (0.10) while there was virtually no difference between gain and loss displays in the alphanumeric condition. There was a main effect of frame, $F(1, 46) = 5.618$, $p < 0.022$, with a higher percentage of subjects correctly leaving cities uncovered in the gain frame than in the loss frame (0.158 versus 0.108). There was a main effect of format, $F(1, 46) = 7.108$, $p < 0.011$, with a higher percentage of subjects correctly leaving cities uncovered in the graphical condition than in the alphanumeric condition (0.156 versus 0.109).

### *3.4. Discussion*

The principal purpose of the study was to understand the relationship between visualization and the framing decision bias. Although the framing model makes specific predictions about risk seeking and risk aversion in particular situations, both processes are a reflection of the well-documented loss avoidance bias. Humans attempt to avoid sure losses even when the alternative decision would lead to a higher expected value for lives saved. Specifically, the psychological loss functions are steeper than gain functions and it is worse to lose money than to fail to show a gain of equal monetary value (Shafir & Tversky, 1995).

The latter inequality implies that if the same situation is framed in terms of losses, the results will be different from that framed in terms of possible gains. In the present experiment, participants were expected to take more missiles out of reserves to avoid losses compared to the same situation wherein the results were framed in terms of possible lives saved. This behavior was predicted even though the importance of keeping some missiles in reserve was emphasized in the instructions (i.e., it was not optimal to take too many missiles out of reserves). Although the framing trend was in the predicted direction, the prediction on reserve decisions was not significant because the participants tended to take out almost all their reserves in both framing conditions resulting in sub-optimal decisions. The average value for missiles left in reserve was less than one for both cases. This general trend agrees with previous research and is probably an extension of the loss avoidance principle. In the above-reported first experiment, NMD operators reported being more concerned about the present attack than about possible future attacks to the extent that they tended to underweigh the probabilities of the future attack (Barnes et al., 2000). Again, this makes sense if they are biased toward reducing losses in the present compared to possible future losses.

Conditions were also created where the expected value of covering all the target cities with GBIs would result in more expected losses than if the GBIs were allocated to protect the larger cities and some of the smaller cities were not protected (uncovered). This was an attempt to replicate the risk-seeking behavior that Tversky and Kahneman (1981) found in the original study (replicated recently by Mayhorn et al., 2002). The logic in the present experiment was that test subjects would avoid leaving a small city uncovered (because it is a sure loss) to a greater extent for loss display conditions than for gain conditions even in cases where larger cities were not given adequate protection based on expected value models. This prediction was supported by the data. Participants in the loss presentation condition covered more small cities under conditions where it was both appropriate (higher expected value by doing so) and inappropriate (lower expected value by doing so). Thus, subjects presented with loss displays were risk seeking in the sense of Tversky and Kahneman; they would rather risk a possible higher expected loss for larger cities than accept a sure loss for a smaller city. In general, both where it was appropriate and where it was inappropriate, information in terms of losses made the subjects more sensitive to protecting against the sure losses of smaller cities. An unexpected finding was that the loss avoidance effects were more pronounced for the graphical displays than for alphanumeric displays. This suggests that the graphical representations emphasized the impact of the loss of small cities indicating an important interaction between the visualization format and the framing bias.

Predictions were also made that improved visualization would lessen the loss avoidance biases. Observers' visualization parameters were varied in two manners. In the risk content conditions, probability ($p$) and population value ($v$) was presented for each city as either separate indices or as a combined expected value. This was important theoretically because the Tversky and Kahneman (1981) experiment presented this information as separate values. Thus, it is possible that their effects were due to the computational overhead that computing an expected value entailed. In the presentation conditions, information was presented as either combined information (expected risk) or as separate values (population and probability). For format, the graphical displays showed this as a configural area or as two bar graphs. We argued that the configural representations would make the implications of expected value solutions more obvious to the observer ameliorating their loss avoidance biases. The only significant effect was an interaction between frame and presentation. As predicted, combining information as an expected value improved decision score performance for the loss framing condition; however, the gain condition actually showed degraded performance when it was presented as a combined expected value. The formatting of the display had no significant effect suggesting it was not a perceptual artifact of the displays themselves. Thus, the framing bias was evinced by the type of reserve decisions operators made but framing showed no main effect on decision-making effectiveness. It appears that gain and loss information resulted in non-optimal decisions that canceled each other out in terms of overall effectiveness. We are in the process of investigating possible reasons for this in a follow-up study.

For NMD, operators' SA is a crucial part of their mission statement. During an actual missile attack, being aware of the parameter of the unfolding attack was not only important for missile allocation but the results were also critical to numerous other entities both for military and civil disaster planning. For that reason, an independent measure of SA was used. For probe performance, graphical configural displays were better than the numeric displays, a finding that agreed with most of the literature on the efficacy of graphical representations for uncertainty data (Smith & Wickens, 1999). More surprisingly, gain information was remembered better and responded to more rapidly than loss information. This contradicted the previous research on NMD performance, wherein the operators had better SA for "missile leakers" (based on probability of loss) than for probability of success information (Barnes et al., 2000). The present study offers a more complete picture of the unfolding NMD situation since it combines risk and uncertainty for the decision maker.

The results imply that the framing bias exists but affects decision effectiveness only tangentially. In general, the operator tended to use all reserves in both framing conditions. In follow-on experiments, we are investigating better feedback visualizations and training techniques to improve decision effectiveness. The operator must learn to use the visualization information to make more effective decisions to adapt to the changing NMD situation. It seems likely the import of reserve decisions was allusive to the operators in our experiments because they did not receive feedback on the effects of holding missiles in reserve. Future efforts will focus on the type and amount of feedback visualization necessary to make NMD operators aware of the implications of their allocation decisions. If our emerging hypothesis is correct, the utility of visualizations will be their ability to signal to the operator (based on past experience) the consequences of the risk factors on future as well as current missile attacks. The framing effect and its implications on SA are the basis of the study we completed recently and are in the process of analyzing (Rothrock et al., 2003). The purpose of that study was to understand more fully the somewhat contradictory results we obtained between the first and third studies.

In summary, this series of studies has shown important interactions between cognitive biases and visualization methods. The ongoing research is an attempt to mitigate these effects in order to develop visualization and training methods to improve decision making and SA under risk in volatile environments. Success would seem to depend on integrating coherent risk and uncertainty information into a richer mental model that provides SA for both the present situation and for future possibilities.



## 4. COEVOLUTIONARY ALGORITHMS FOR PEACEKEEPING SIMULATIONS: VISUALIZING EMERGENT SOLUTIONS

In the previous section, a series of empirical studies was used to examine the application of coherent visualizations to a specific problem set, decision making about risk in the face of present and future uncertainty. In this section, an example of a visualization tool that might be used to support narrative thinking about complex and uncertain situations is presented. Here we are not focused on empirical results or types of visualization. Rather, we are looking at the kinds of decision aids that might help the soldier maintain narrative SA. That is, what sort of tool can visualize key relationships and make them coherent, which the theory of Section 2 says is important, and also help decision makers include in their solution space a richer number of possible solutions, which was seen as important in the decision-making experiments of Section 3.

This is a specific research tool, a prototype, intended to allow an analyst to be able to see and explore emergent relationships that might not normally be explored. The tool visualizes their interactions as a narrative (3D animation), but then provides a rich data set for filling in one's mental model where experience (or lack of it) and cognitive biases might leave gaps. As a prototype, the specific visualizations used are not being presented as the most optimal solutions, but rather are presenting the conceptual tool as an available approach that can help the analyst both understand and adapt to complex and uncertain situations. It does so by pushing the analyst into a scenario-driven narrative mode for making complex patterns and relationships coherent. Thus, this section should show how some of the more conceptual goals of the theory in Section 2 might be actually implemented in decision support architecture.

The domain of this simulation-based analysis system is the less-conventional set of military missions classified under the label of stability and support operations (SASO), including peacekeeping, disaster relief, anti-terrorist security operations, and other non-force-on-force combat scenarios. The military defines the purpose of stability operations as being "to promote and sustain regional and global stability" and the primary role of stability operations is "to meet the immediate needs of designated groups, for a limited time, until civil authorities can accomplish these tasks without military assistance" (Department of the Army, 2003). SASO operations may also have the goals of keeping armed conflicts contained and quieting domestic disturbances. The resulting scenarios are very similar to situations that troops in Iraq face daily.

This section introduces a system that provides multi-sided coevolution for military peacekeeping missions. Because of the complex environment, the interactions of sides are captured in a simulation in which all sides compete (i.e., adapt) to the changing situations with analytical visualization methods being used to display results in a format intuitive to the analyst.

Many applications currently exist to support commanders' decision making in conventional warfare scenarios. FOX (Schlabach, Hayes, & Goldberg, 1999) notably used genetic algorithms to create conventional warfare courses-of-action (COA). One of the major drawbacks of FOX was that it only allowed evolution of the blue side (U.S. forces) against several static COAs of the red (enemy) side. Hillis and Winkler (2001) developed a coevolutionary approach to the FOX simulation that allowed the red and blue sides to adapt to each other's tactics. However, the conventionality of the tactics and the simplicity of the paradigm did not allow Military Intelligence (MI) analysts to gain insights into more complex environments such as Iraq. The principal advantage of Scheherazade visualizations is in their ability to portray emergent solutions as the different entities compete to obtain conflicting goals in complex SASO environments (Schlabach & Hillis, 2003). In this sense, it captures the dynamics of entities that adapt to a volatile environment in contrast to rule-based solutions, which have a more constrained set of outcomes.

### *4.1. Overview*

A four-part approach has been developed: creating the SASO environment, the coevolution of COAs by several agents, the SASO simulation named

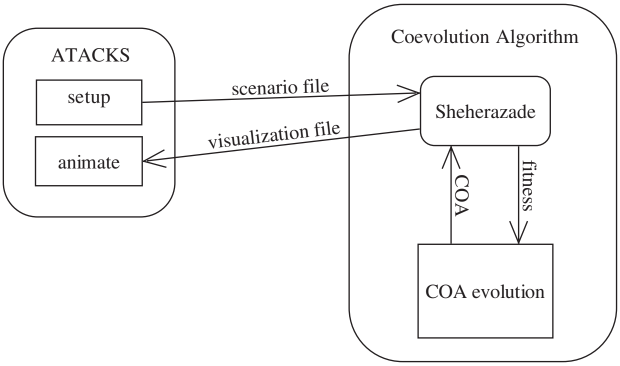


*Fig. 3.* System Overview.

Scheherazade (Schlabach & Hillis, 2003), and the analysis of the results with evolution and scenario visualizations. These parts work together as shown in Fig. 3.

### *4.2. Creating the SASO Environment*

The SASO environment was developed in a cooperative enterprise between Ft. Huachuca researchers and ARL (Schlabach & Hillis, 2003). The scenario contains entities, factions, and locales (see Fig. 4). The entities could be terrorists, refugees, media, non-government organizations, or peacekeeping forces. Each entity has relative combat, intelligence, and other strengths and belongs to an allegiance or faction. Each faction has a starting animosity or friendliness to every other faction. The locales contain percentages of the population of each faction, and each locale has a starting attitude or calmness. The military expert defines all the values for the entities, factions, and locales to start modeling a situation he or she wants to investigate by having the system generate COAs for each of the entities.

After establishing the environment, the military expert also assigns each entity to an agent. A simple assignment consists of each agent's entities belonging to one faction. The assignment of entities to agents is important in defining the fitness functions or goals of each agent. For example, if all entities of agent 3 belong to the faction named "Eastern Alliance," then the fitness function of that agent could be to cause local unrest. Another fitness

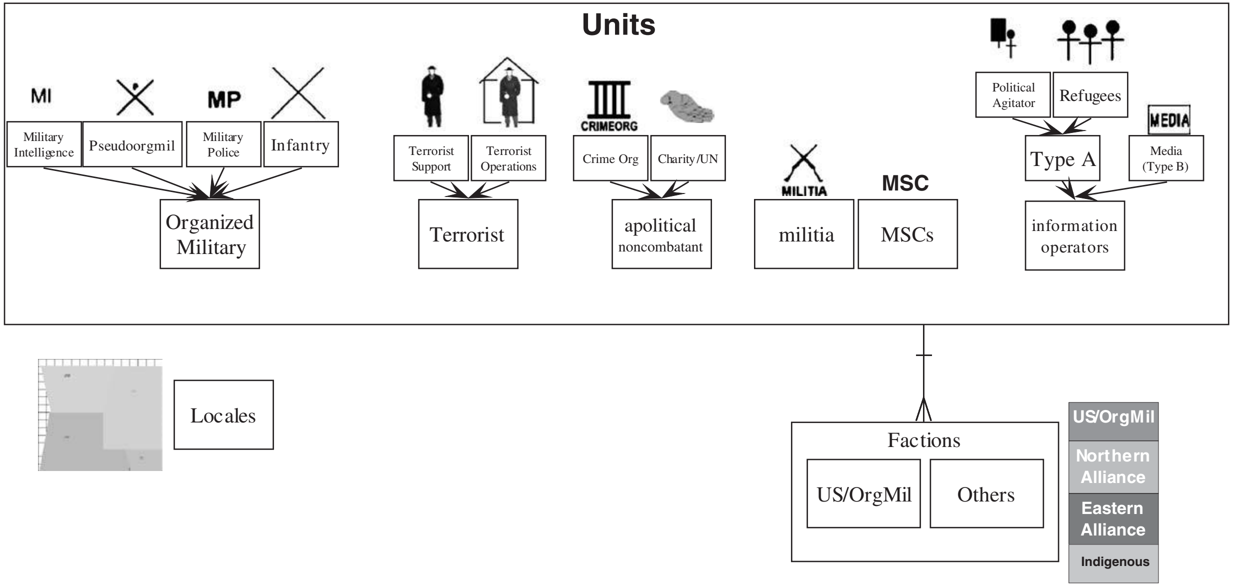


*Fig. 4.* Entities and their Relationships.

criterion could be to inflict as much damage as possible on another faction or it could be a weighted combination of several of these factors.

### 4.3. *Coevolution of COAs*

Once the scenario parameters have been defined, the system can begin evolving movements and targets for each of the agents through a genetic algorithm (GA). The GA allows each agent to take its turn evolving. As each agent changes the movements and target factions of its entities, each of the other agents tries to find a better set of movements and target factions for its own entities to reach a better fitness score.

For many entities, the chromosomes for the GA consist of a list of scheduled movements and targets, if the entity type engages in combat. For the organized military entities that make up the blue or friendly faction, the chromosomes are a set of assignments of units to military subordinate commands, which are, in turn assigned to locales.

The GA takes these chromosomes and passes them to the SASO simulation, which returns with several scores that make up the fitness function. (The scores are actually averages of many runs of the same chromosomes due to the stochastic nature of the simulation.) The GA uses the fitness function to select, crossover, and mutate new COAs, which are again played, scored, and selected. The agents take their respective turns evolving against the other agents' COAs, changing their strategies in reaction to the other agents' changes.

### 4.4. *Scheherazade: The SASO Simulation*

As mentioned previously, the coevolution algorithm uses the SASO simulation, named Scheherazade, after the famous character of "1,001 Arabian Nights," as values for its fitness function. The basic concepts of entities, factions, and locales will be introduced in Section 4.4.1. A more detailed description of the environment, entities, events, and explanation of respective COAs in the simulation algorithm follows.

#### *4.4.1. Environment*

The basic environment consists of factions and locales. Factions are deliberately vague groupings of people assumed to be nominally united by common affiliations and that tend to be thought of as a group, such as "NATO" or "the ethnic Serbs." They are not necessarily politically or organizationally united.

A faction's "animosity" toward another faction represents its "feelings" for that faction. Every provocation (incident, violation, terrorist hit, military attack, etc.) causes the victim's faction to increase its animosity level toward the perpetrator's faction.

Locales are geographical regions, which can be thought of as neighborhoods within a city or states/provinces within a nation. Locales have several properties including geographic size, population size, and which other locales are neighbors. The local population of a locale is also divided among the factions. An important property that affects many events in the simulation is the locale's "attitude."

"Attitude" for each locale influences the probability that one faction will attack a target of another faction, given the chance. The attitude score cumulatively reflects the effects of the recent incidents in the locale, emotional value of locale, population over/under density, developmental factors, etc. Attitude is also usually an important factor in the fitness functions of the agents.

### *4.4.2. Entities and COAs*

Entity types describe the options available to a player set. Different entity types have different COA mechanisms such as a list of movements, times, and targets or a division of power among locales.

For each simulation turn, or clock tick, every entity is located within one of the locales. Entities within the same locale have a probability of interacting, depending upon a number of factors, which include entity characteristics as well as the locales' properties listed earlier.

The entity types Scheherazade offers the scenario designer are organized military, militia, terrorists, information operators, and apolitical noncombatants, as shown in Fig. 4. Each entity type has different properties and behaviors some of which contribute to calming effects, while others contribute to agitating effects. There are certain properties and behaviors which all entities share. An entity's calming or agitating effects are amplified proportionally to the relative size of the local populace in its faction. Every attack and incident drives the attitude up in proportion to its severity.

### *4.4.3. Events*

In Scheherazade the entities interact with the environment to model certain real-world events such as:

Clock ticks represent temporal dimensions defined by the analyst. They regulate the order of play for agent actions. Within a clock tick, entities within each locale consult the locale conditions (current attitudes, population demographics, other entities in the locale) to determine its action for that clock tick. Incidents describe a class of interactions between entities or an entity and the local population. Incidents have an associated "severity" rating to determine the appropriate adjustment to make to attitude levels. When an incident occurs that includes combat, Scheherazade consults the weighted combat values of the contributing entities and assesses combat attrition.

### *4.5. Visualization*

Our research in this area concentrates on the visualization of the novel symbologies of the entities and their behaviors. We provide visualizations of specific actions, as well as abstracted, conceptual displays of the relationships of entities and regions. We use the Advanced Tactical Architecture for Combat Knowledge System (ATACKS), previously described in Suantak, Momen, Rozenblit, Barnes, and Fichtl (2001), to show the movements and graphs of a simulation run.

ATACKS is a three-dimensional (3-D) visualization tool that facilitates rapid, flexible development of high-level battlespace representations as well as execution and assessment of war-gaming scenarios. It expands standard battlefield symbology by providing abstract symbols on 3-D abstract battle space terrains. It extends normal spatial visualization through process-centered displays that seek to enhance the commander's understanding of the situation by presenting qualitative data in novel formats.

The visualizations of a simulation run play an important role in our development and analysis of Scheherazade. ATACKS provides a graphical user interface to set up a scenario and then provides several displays to show important events and values for a scenario.

### *4.6. SASO Simulation Analysis*

Once a military expert has used ATACKS to define a scenario, the resulting file is used to run the coevolution algorithm. As the coevolution algorithm is running, it passes COAs to Scheherazade in order to get back the fitness scores. Scheherazade produces an output visualization file of each of the best COAs for each agent per generation. These data files can then be read by ATACKS to create an animation and several displays that show the COAs, movements, incidents, attitudes, damage (attrition), and animosities for that run. Most importantly, these displays show relationships, such as the effect of movement and incidents on locale attitudes, and relative changes of animosities between factions.

A set of easily recognizable icons for each type of entity was created. Fig. 4 shows the entities and their relationships, as well as example colors for locales and factions. The entity icons were then placed on 3-D units that move between locales, as shown in Fig. 5.

Each icon is filled in with its faction color. The colored bar graph in the center of the locale indicates the percentages of population that belong to each faction in that locale. For each clock tick, entities move in and out of locales in the 3-D environment of ATACKS. However, an overall view of the movements and events has proven more helpful in understanding scenarios, as the RAVENS framework would lead one to expect.

Using consistent colors and icons, a movement graph showing the in and out movements of each entity appears in Fig. 6. The background colors

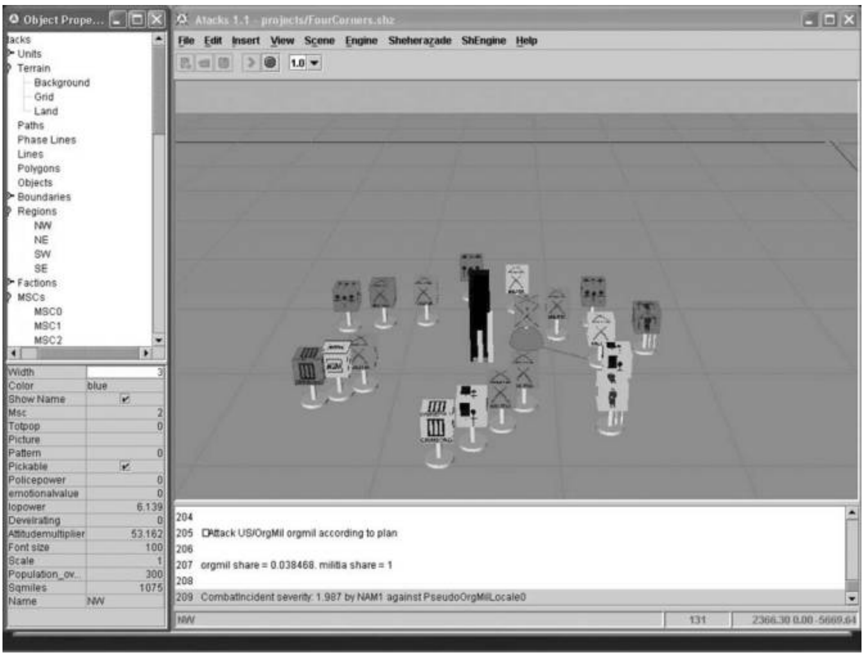


*Fig. 5.* ATACKS Snapshot Showing Sherherazade Entities.

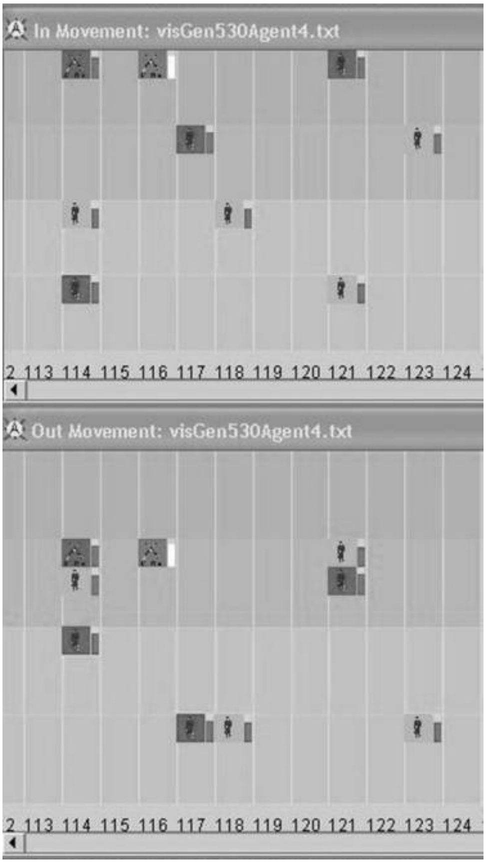


*Fig. 6.* Movement Graphs.

indicate locales; the numbers on the bottom indicate on which clock tick an entity moves. The bar on the right of each icon indicates the power of that entity, a value that significantly affects attitude (depending on the type of entity) of that locale. The color of the bar graph indicates whether the entity is calming (white) or agitating (red). Thus, this display conveys entity type, entity allegiance, clock tick moved, locale moved to (and out of), relative power, and type of power. Similar types of graphs can be used to show movement and targets within COAs. Further, these icons can be interactive. Right clicking on an icon can allow you to reassign targets within a COA on the fly.

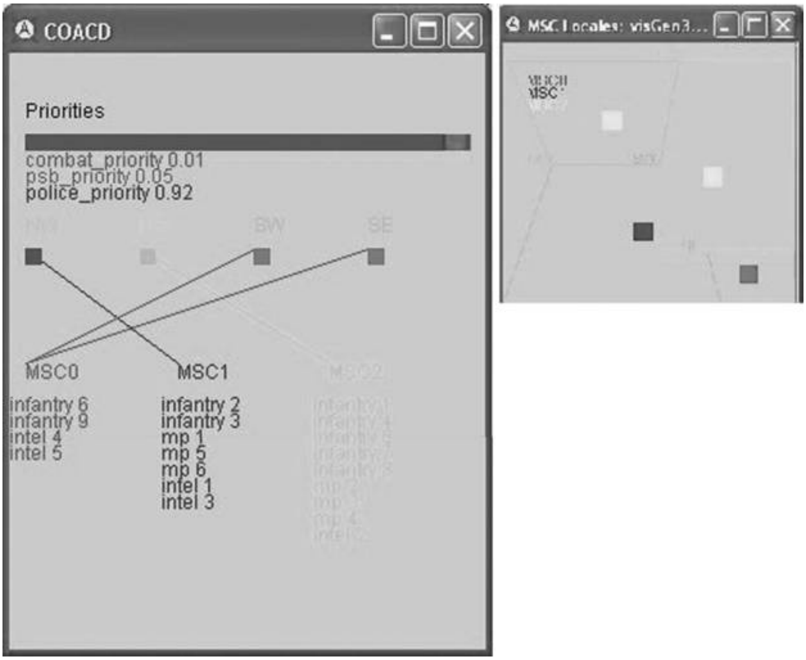


*Fig. 7.* Organized Military COA.

The COA of the organized military unit has a different structure. Therefore, it requires a different kind of visualization as seen in Fig. 7, which shows which entities have been assigned to which Military Support Command (MSC) and which MSCs are assigned to each locale. As Fig. 9 shows, an MSC can be responsible for more than one locale.

Other factors, such as faction animosities and locale attitudes are important influences on the SASO simulation that may be a part of the fitness functions. The animosities are graphed on a starplot (Wong & Bergeron, 1997). For example, the first graph in Fig. 8 shows the animosities for and against the U.S. faction. Each radial corresponds to a faction, consistent with the colors of the entities. The further away from the center, the more animosity the U.S. has for the corresponding faction. The middle of the radial denotes a neutral attitude.

Standard line graphs, with some annotation, show changes in attitudes, accumulations of damage per faction, and any other factors of interest to the user. The line graph shown in Fig. 9 graphs the attitudes of the four locales over clock ticks, color-coded again to the locale colors.

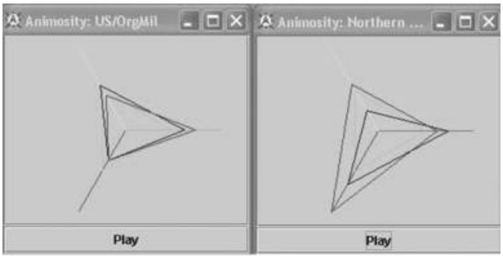


*Fig. 8.* Starplots of Animosities.

Each line for each locale is labeled with the incidents that occurred at that clock tick. The graph in Fig. 9 shows calm attitudes in the locales, and more and more incidents as time goes by. A large jump in incidents occurs in the SW locale toward the end of the scenario, agitating that locale. This graph illustrates a strategy by one of the factions to increase the attitude in that locale.

Currently, the most clear and meaningful information comes from looking at the attitude line graph. Once the analyst identifies a trend, more information about the incidents can be found on the incident graph, such as that shown in Fig. 10. This display shows the two units involved in the incident. The perpetrator is shown on top, the locale in which it occurred is the background color, the clock tick is on the $x$-axis, the type of incident is written as an abbreviation underneath the two icons, and the bar graph on the right indicates the severity of the incident. For example, a yellow militia unit attacked a blue organized military unit at clock tick 95, causing an incident of noticeable severity (severities tend to be quite low).

Therefore, to analyze an SASO simulation battle, a user can examine the various displays, from attitudes to animosities, and movements and incidents. This ability has been invaluable in trying to understand the dynamics of the system as a whole.

### *4.7. Coevolution Analysis*

Applying coevolution to this problem is a natural choice because it should be more likely to produce robust plans and also give some indication of possible dangerous counterplans from the other players. There are other, more important advantages as well. Observing the coevolutionary run as it plays out can yield surprising insights into the nature of the war-game

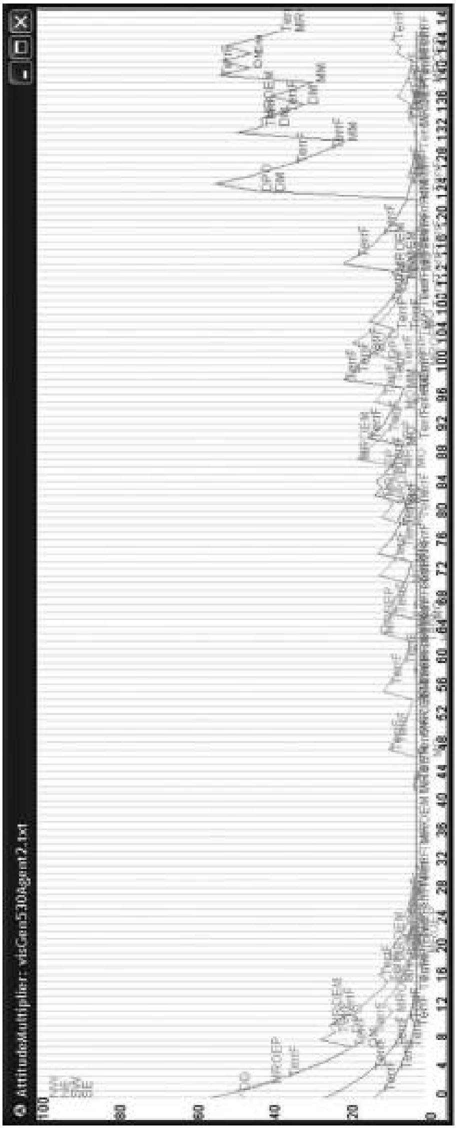


*Fig. 9.* Locale Attitudes and Incidents.

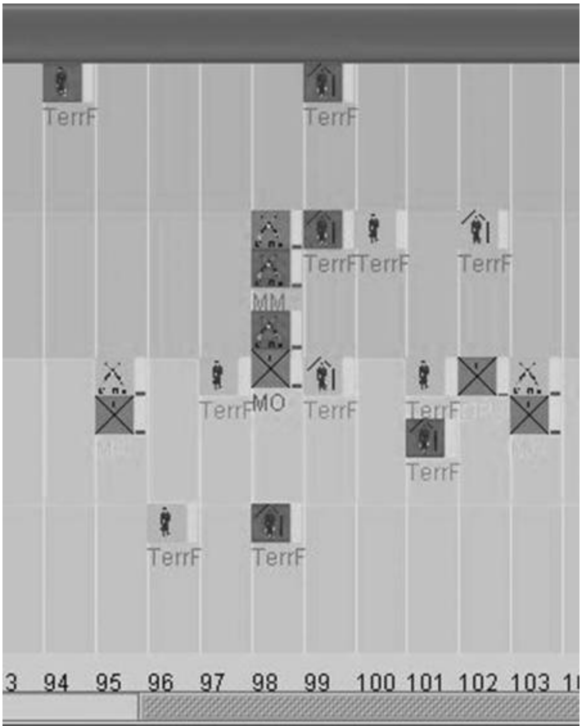


*Fig. 10*. Incidents.

simulator and potentially about the real-world situation it is intended to model.

In our experience, coevolutionary war gaming rarely leads to stable states where each player settles on one strategy. Instead, the system tends to drift between quasi-stable cycles where, like the game (scissors, paper, stone), a changed plan by one player leads to natural and obvious reactions from the other players. Obviousness, however, lies in the eyes of the beholder. The utility of the technique comes from the fact that the GA frequently exploits aspects of the system of which the human designers and operators were unaware.

Our objective is to harness this effect in order to show an operator important facts about the simulation (and hopefully about the mission itself) that might otherwise be overlooked. Currently, finding meaning from a coevolutionary run is more art than science. One watches coevolution as it unfolds, following plan and counterplan as the players continually adapt. By definition, we do not know for what we are searching until, in a Zen-like flash of enlightenment, it seems obvious. The purpose of our visualization research is to portray results in such a manner that these insights are obvious to the trained analyst (and not only to the experienced expert).

The coevolutionary process requires visualizations of the simulation in order to compare successful COAs. The GA explores the search space by finding COAs that produce better outcomes for the respective player. However, it does not explain why one COA may be better than another COA. By comparing the simulation runs, an analyst can use the visualizations to determine what actual strategy resulted in a better score and more importantly, why.

Each agent evolves against the current best set of COAs of the other agents, known as the "hill," and only gets a chance to update the hill with its own best strategy every 10 cycles. This allows all of the agents enough time to evolve more mature strategies. Therefore, at generation 1, agent 1 may obtain the best score. For the next nine generations, each of the other agents evolves against the current best score. Finally, at generation 10, agent 2 is allowed to place its best set of movements, targets, etc. as the new best score (i.e., the new king of the hill). This system has the effect that each agent's strategy can potentially change radically every 10 generations.

These kinds of analyses have led to many insights into our system. Using visualization to show trends in this complex environment has led to a much better understanding of how the separate rules and the coevolution interact. There is still research to be done before the system is going to be useful for an analyst. In conjunction with military experts, we are currently developing a user-friendly interface that will permit the analyst to define Scheherazade's parameters. Eventually, the interface will allow the operator to play the role of any of the *n* factions in the gaming environment and visualize the adaptations that other factions use to improve their scores.

Developing the optimal mix of set rules and coevolution will require interacting with actual analysts during field exercises or even actual operations to define the proper type of visualizations and the underlying intelligent software. More rigorous evaluations and in situ testing will be required before its military utility is established. However, the concept is promising because it is designed to impart insight using a narrative format. Scheherazade does not so much predict the future as it highlights which of the "1,001 possible stories" are not only most likely but also those that are most potentially dangerous and counterintuitive. In effect, the analyst is able to visualize adaptation in process.

# 5. CONCLUSION

This chapter was based on the RAVENS framework that elucidated the cognitive aspects of battlefield visualization. RAVENS posits that humans are inherently flexible decision makers and SA depends on the ability of humans to create narrative visualizations that capture the overall context of complex military environments. Using the framework as a guideline, we discussed two important visualization research programs whose purpose is to allow military operators to rapidly adapt to volatile situations. The first program investigated cognitive effects such as the framing bias and their possible interactions with a variety of display concepts during a series of missile defense simulations. The experimenters presented risk as a spatial representation of uncertainty and target value that emphasized either expected population lost or expected population saved. The second program investigated the feasibility of using visualizations generated from Scheherazade (a coevolutionary algorithm) to aid MI analysts in predicting emergent tactics of terrorist groups during urban operations. The underpinnings of Scheherazade consist of a user interface to define a peacekeeping environment, a simple simulation tool, a coevolutionary algorithm for iterative adaptations, and a visualization module. We conclude, based on our analysis of future combat environments, research results, and the reviewed literature:

- Adaptations to future military missions will be particularly difficult because of the complexity and uncertainty inherent in asymmetric and anti-terrorist warfare.
- Visualization aids must promote insight and flexibility rather than doctrinal rule-based solutions suggested by past decision aiding research.
- RAVENS is a cognitive framework for military adaptation that posits that the best tool the human has for understanding dynamic and uncertain situations, and to make pragmatic decisions, is their own narrative cognitive abilities.
- Risk management and uncertainty judgments are involved in most important military decisions in adaptive environments. Visualizations need to be developed that impart SA but the designer must consider human cognitive biases as well as the perceptual characteristics of the display. These visualizations need to provide a rich and coherent mental model of possible outcomes.
- The initial research indicates that developing risk displays for missile defense is not straightforward. Our results show that humans process loss

and gain information differently but the framing biases tend to cancel each other out in terms of making optimal decisions.

- For SA, positive (gain) information was processed more effectively than negative (loss) information in most but not in all experimental conditions. Graphical formats that showed risk as an emergent property of area were in general more effective than other displays examined for SA.
- The general conclusion was that the operators did not necessarily understand the consequences of their missile defense decisions. Future research will examine the interaction of visualization cues with feedback and training parameters.
- A narrative model of visualization is inherently flexible in that it allows a common motif to be shared among multiple players with multiple variations possible as circumstances change.
- Scheherazade is an example of a new generation of visualization tools that use narrative formats and adaptive algorithms to interact with intelligence analysts.
- The goal of Scheherazade is to provide a means of representing and modeling dynamic and uncertain environments richly enough to give an analyst a coherent understanding of the consequences of possible tactical adaptations made by themselves and their adversaries.

# REFERENCES

Barnes, M. J. (2003). *The human dimension of battlespace visualization: Research and design issues (ARL-TR-2855)*. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Barnes, M. J., & Hillis, D. (2003). *Dynamic visualization of co-evolving adversaries in small-scale contingencies*. ARL Compendium FY02 DRI Final Reports (pp. 43–57). U.S. Army Research Laboratory, Adelphi, MD.

Barnes, M., McDermott, P., Hutchins, S., Gillan, D., & Rothrock, L. (2003). *The presentation of risk and uncertainty in the context of National Missile Defense simulations*. Poster session presented at the proceedings of Collaborative Technologies Alliance (CTA) Symposium, College Park, MD.

Barnes, M. J., Wickens, C. D., & Smith, M. (2000). Visualizing uncertainty in an automated National Missile Defense simulation environment. *Proceedings of the 4th annual FedLab symposium: Advanced displays and interactive displays*, U.S. Army Research Laboratory, Adelphi, MD (pp. 107–111).

Bar-Yam, Y. (1997). *Dynamics of complex systems*. Reading, MA: Perseus Books.

Bennett, K. B., Toms, M. L., & Woods, D. D. (1993). Emergent features and graphical elements: Designing more effective configural displays. *Human Factors*, *35*(1), 71–97.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Bruner, J. S. (2002). *Making stories: Law, literature, life*. New York, NY: Farrar, Straus, & Giroux.

Carroll, J. (1999). Five reasons for scenario-based design. *Proceedings of the 32nd Hawaii international conference on system sciences – 1999*, IEEE.

Department of the Army. (2003). *Stability and support operations* (FM 3–07). Washington, DC.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32–64.

Endsley, M. R. (2002). Theoretical underpinnings of situation awareness: A critical review. In: M. R. Endsley & D. J. Garland (Eds), *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fiske, S. T. (1993). Social cognition and social perception. *Annual Review of Psychology*, *44*, 155–194.

Gershon, N., & Page, W. (2001). What storytelling can do for information visualization. *Communications of the ACM*, *44*, 31–37.

Gillan, D., & Hutchins, S. (2002). *Display situation awareness*: *A study into information integration and presentation*. Unpublished manuscript. New Mexico State University, Las Cruces, NM.

Harre, R., & Gillett, G. (1994). *The discursive mind*. Thousand Oaks, CA: Sage.

Hillis, D., & Winkler R. (2001). The fox and the hare: A co-evolutionary approach to course of action generation. *Proceedings of the 5th annual Fed Lab symposium*: *Advanced displays and interactive displays consortium*, College Park, Maryland, USA.

Hirokawa, R., DeGooyer, D., & Valde, K. (2000). Using narratives to study task group effectiveness. *Small Group Research*, *31*, 573–591.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J. (1999). Naïve probability: A mental model of extensional reasoning. *Psychological Review*, *106*(1), 62–88.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Klein, G. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Lopes, L. (1986). Aesthetics and decision science. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-16*(3), 434–437.

Mayhorn, C. B., Fisk, A. D., & Whittle, J. D. (2002). Decisions, decisions: Analysis of age, cohort and time of testing of risky decisions options. *Human Factors*, *44*(4), 515–521.

McDermott, P., Hutchins, S., Barnes, M., Könecke, C., Gillan, D., & Rothrock, L. (2002). The presentation of risk and uncertainty in the context of national missile defense simulations. *Proceedings of the human factors and ergonomics society 46th annual meeting*, Baltimore, Maryland, USA.

Meyer, J., Shamo, M. K., & Gopher, D. (1999). Information structure and the relative efficacy of tables and graphs. *Human Factors*, *41*, 570–587.

Murray, K. (1995). Narrative partitioning: The ins and outs of identity construction. Posted at http://home.mira.net/~kmurray/psych/in&out.html, abridged from a published book chapter.

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition*, *49*, 123–163.

Rossiter, M. (1999). A narrative approach to development: Implications for adult education. *Adult Education Quarterly*, *50*, 56–71.

Rothrock, L., Sungsoon, P., Barnes, M. J., McDermott, P., Hutchins, S., & Gillan, D. (2003). Systematic analysis of risk visualization strategies for homeland defense. *Proceedings of the 2003 IEEE international conference on systems*, *man* & *cybernetics*, Washington, DC.

Scales, R. J. (2000). Adaptive enemy: Achieving victory by avoiding defeat. *Joint Force Quarterly (JFQ)*, 7–14.

Schlabach, J. L., Hayes, C. C., & Goldberg, D. E. (1999). FOX-GA: A genetic algorithm for generating and analyzing battlefield courses of action. *Evolutionary Computation*, *7*(1), 45–68.

Schlabach, J., & Hillis, D. (2003). *Sheherazade: A research platform for decision support in military stability and support operations*. BCBL-H Tech Report 2003–01. Battle Command Battle Lab, Ft. Huachuca, AZ.

Shafir, E., & Tversky, A. (1995). Decision making. In: E. E. Smith & E. E. D. Osherson (Eds), *Thinking* (pp. 77–100). Cambridge, MA: MIT.

Shank, R. (1990). *Tell me a story*. Evanston, IL: Northwestern University Press.

Smith, M., & Wickens, C. D. (1999). *The effects of highlighting and event history on operator decision making in a national missile defense system application*. Tech. Rep. No. ARL-99–4. Aviation Research Laboratory, University of Illinois, Savoy, IL.

Snowden, D. (2000). New wine in old wineskins: From organic to complex knowledge management through the use of story. *Emergence*, *2*, 50–64.

Suantak, L., Momen, F., Rozenblit, J., Barnes, M., & Fichtl, T. (2001). Intelligent decision support of support and stability operations (SASO) through symbolic visualization. *IEEE International Conference on Systems, Man, and Cybernetics*, *5*, 2927–2931.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.

Waldrop, W. M. (1993). *Complexity: The emerging science at the edge of order and chaos*. New York: Simon Schuster.

Warner, J. (2003). RAVENS: A framework for rapid adaptive visualization of emergent and novel situations. *Proceedings of the RTO Systems Concepts and Integration* (*SCI-129*) *Panel*, Prague, Czech Republic.

Warner, J., & Burnstein, D. (1996). *Situation, domain and coherence: Towards a pragmatic psychology of understanding (ARL-CR-306)*. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Weick, K. (1995). *Sense making in organizations*. Thousand Oaks, CA: Sage Publications.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance*. Upper Saddle River, NJ: Prentice-Hall.

Wong, P. C., & Bergeron, R. D. (1997). 30 Years of multidimensional multivariate visualization. In: G. M. Nielson, H. Hagen & H. Müller (Eds), *Scientific visualization – overview, methodologies, techniques* (pp. 3–29). Los Alamitos, CA: IEEE Computer Society Press.