Coactive Emergence as a Sensemaking Strategy for Cyber Operations

Jeffrey M. Bradshaw[1,*], Marco Carvalho,[2] Larry Bunch[1], Tom Eskridge[1], Paul J. Feltovich[1], Chris Forsythe[3], Robert R. Hoffman[1], Matt Johnson[1], Dan Kidwell[4] and David D. Woods[5]

[1]Florida Institute for Human and Machine Cognition (IHMC), 40 South Alcaniz Street, Pensacola, FL 32502
[2]Department of Computer Science, Florida Institute of Technology (FIT), Melbourne, FL
[3]Sandia National Laboratories, Albuquerque, NM
[4]U.S. Department of Defense, U.S. Government
[5]The Ohio State University, Columbus, OH

Abstract

In this article we describe how we apply the concept of *coactive emergence* as a phenomenon of complexity that has implications for the design of sensemaking support tools involving a combination of human analysts and software agents. We apply this concept in the design of work methods for distributed sensemaking in cyber operations. Sensemaking is a motivated, continuous effort to understand, anticipate, and act upon complex situations. We discuss selected results of a macrocognitive work analysis that informed our focus for design and development of support tools. In that analysis, we identified seven target topics that would be the focus of our research: engaging automation as a full partner, reducing the volume of uncorrelated events, continuous knowledge discovery, more effective visualizations, collaboration and sharing, minimizing tedious work, and architecting scalability and resilience. In addressing the first target topic, we show how coactive emergence inspires an agent-supported threat understanding process that is consistent with Klein's Data/Frame theory of sensemaking. In subsequent sections, we describe our efforts to address the remaining six target topics as part of design and development of a cyber operations framework called Sol. Specifically, we describe the use of agents, policies, and visualization to enable coactive emergence for taskwork and teamwork. We also show how policy-governed agents working collaboratively with people can help in additional ways. We introduce the primary implementation frameworks that provide the core capabilities of our Sol cyber framework: the *Luna Software Agent Framework,* and the *KAoS Policy Services Framework*. We describe areas for future development of Sol, including the incorporation of the *VIA Cross-Layer Communications Substrate*. Finally, we describe recent results and current plans for empirical studies addressing some of the issues raised in this article.

Keywords: cyber defense, cyber operations, cyber security, teamwork, software agents, policy management, organic resilience, coactive emergence, sensemaking

*Corresponding author. Email:jbradshaw@ihmc.us

1. Introduction

In broad terms, the work of the cybersecurity professional, on behalf of their organization, is to formulate answers and undertake actions in response to questions such as the following:

- *What is the nature and purpose of current attacks and what is their origin?*
- *What are the attackers doing now and what might they do next?*
- *How do the attacks affect my mission now and how might they affect it in the future?*
- *What options do I have to defend against these attacks?*
- *How effective will a given option be against these attacks and what effect will exercising it have on my mission and how is it likely to affect the future actions of allies and adversaries?*
- *How do I prevent or mitigate the impact of such attacks in the future?*

Analysts working in large-scale Network Operations Centers (NOCs) are a vital part of cyber defense as they monitor, detect, understand, and respond to attacks or other conditions (e.g., power failures) that might impact mission performance. Typically working in close proximity within large rooms filled with individual workstations and a video wall at the front intended to keep everyone aware of important developments that may affect their work, they are organized into hierarchical groups with different duties or spans of responsibility. Some analysts are more focused on ongoing monitoring of events at the moment-to-moment level, while others are responsible for strategic direction or in-depth analysis of serious incidents.

Despite the significant attention being given to the critical challenges of cyber operations within large-scale NOCs, the ability to keep up with the increasing volume

and sophistication of network attacks is seriously lagging. Cyber defense, by its very nature, is asymmetrically disadvantaged in its efforts to fend off attackers and the perception by most of the experienced analysts we have encountered is that the imbalance is worsening. While attackers can strike at their leisure and can profit from the careless exposure of virtually any vulnerability, defenders must be continually vigilant and responsive—both proactively and reactively—to potential threats relating to any aspect of their systems.

Merely throwing more computing horsepower at fundamentally limited visualization and analytic approaches will not advance our aims. Extensive experience in domains with similar challenges has shown that the kinds of complex automation often seen in NOCs today do not adequately leverage human creativity, ingenuity, and flexibility—besides actually hindering analyst effectiveness in some ways. Though ongoing efforts to increase computing resources and improve technology is essential, the point of providing these enhanced proficiencies is not merely to make computational tools more capable in and of themselves, but also to make analysts more capable through the use of such technologies [75]. To better empower these professionals, we need to seriously rethink the way cyber operations tools and approaches have been conceived, developed, and deployed.

In this article, we focus on selected problems for distributed sensemaking and response in Cyber Defense Analysis and other roles in cyber operations. In particular, we describe our experiences in applying knowledge about the cognitive sciences to help analysts working in large-scale NOCs. Though it will be impossible in this article to discuss more than a sampling of relevant research, we will survey some concepts and findings running the gamut from basic cognitive science (e.g., perception, attention, inference, individual differences) to socio-cognitive issues (e.g., theories of social interaction, human-automation teamwork).

As rationale for the principles used in our work design, we present the results of a macrocognitive work analysis (Section 2). In that analysis, we identified seven target topics that would be the focus of our research: engaging automation as a partner, reducing the volume of uncorrelated events, continous knowledge discovery, more effective visualizations, collaboration and sharing, minimizing tedious work, architecting for scalability and resilience. In addressing the first target topic, we describe the Klein, *et al.* Data/Frame theory of sensemaking and introduce the concept of *coactive emergence*. In subsequent sections, we describe our efforts to address the remaining six target topics as part of the design and development of a cyber operations framework called Sol (Sections 4-9). Specifically, we describe the use of software agents, policies, and visualization to enact a sensemaking strategy for taskwork and teamwork inspired by the phenomenon of coactive emergence. We also show how policy-governed agents, working in tandem with people, can help in additional ways. We introduce the primary implementation frameworks that provide the core capabilities of our Sol cyber framework: the *Luna Software Agent Framework,* the *VIA Cross-Layer Communications Substrate*, and the *KAoS Policy Services Framework*. Finally, we describe results of empirical studies addressing some of the issues raised in this article (Section 10), as well as anticipated trajectories for future development of the Sol framework (Section 11).

2. Macrocognitive Work Analysis

Macrocognitive work is how cognition adapts to complexity [5]. Distinguished from the phenomena of cognition that are studied in the traditional psychology laboratory, macrocognition includes such functions as sensemaking, adapting, and collaborating. The study of macrocognitive work involves methods of cognitive task analysis, although we recognize that the term "task," as it is traditionally used, is less apt than the term "work."

2.1. Approach

For the project that we report here, we engaged in a literature survey, obtrusive workplace observations, participation and discussions as part of training exercises, and semi-structured interviews, case study reviews, and discussions with cyber defense analysts in government and private industry. Concept maps, text notes, and drawings were used to record our sessions, however no formal methods of knowledge modeling were used and no formal analysis of the results was undertaken.

Our approach was oriented around four major kinds of inquiries:

1. *Finding out what aspects of the work-shaping technologies were most important yet caused the most difficulty.* Of prime value to gaining an understanding of the analyst's work and its requirements was to understand what activities are the most *important* for conducting work effectively and why. We tried to learn which of these important activities were the most difficult to manage or overcome. Subsequently, we explored some of the perceived reasons for this difficulty. This kind of exercise starts to give us focus in our inquiries and research directions, in order to assure that we are working on problems of high value [1]. We call these areas of interest "target topics."
2. *Inquisitive observation of practice and discussion of case studies to understand the "actual work."* We supplemented our observation of experts through readings and discussions of case studies and work practices. In addition to studying guidelines for standard operations, we have been interested in deviations from these expected practices, the presence of "invisible" (vs. overt) work, and contextual adaptations in the face of field expediency [2]. We reviewed case studies with experts under a

modified "think aloud" procedure. That is, we asked analysts to tell us generally what they were doing at different stages of the activities being reviewed, and we were able to ask questions as their activities interacted with particular points that we were trying to understand. We paid particular attention to any encounters with the "target topics."

Of particular interest are cases that may be seen as challenging analysts for reasons such as the following: 1. they taxed the limits of their expertise (e.g., the solving of an analytic "puzzle"); 2. they required various workarounds (e.g., technology gaps; organizational or procedural inconveniences that necessitated "extra" steps in the work); or 3. they raised personal, organizational, or policy dilemmas (e.g., situations where simply following the accepted procedure would have produced an unacceptable result, or where invisible or explicit organizational and policy structures created barriers to effective performance). Such inquiries identify leverage points for technological interventions, and reveal ineffective problem-solving strategies that affect individual work performance and collaboration (see, e.g., [3]).

3. *Finding out the analysts' "desirements"* [76], that is, functionalities and features they would like to have that would make it easier for them to achieve their work goals. We conducted additional structured discussions on specific questions with analysts to get feedback on design ideas that the team had generated. These discussions continued throughout the project, feeding a spiral development process on the major technological capabilities developed.
4. *Creation and refinement of a scenario as part of the quest for generalizability*. Based on information gleaned from the activities described above, we created a detailed scenario of a 24/7 network operations context. The scenario provided a narrative that would illustrate, and qualitatively represent, the policy-driven, agent-based monitoring and control capabilities being developed. The scenario was reviewed, discussed, and refined with project sponsors, with professional colleagues, and with practicing analysts. Discussions of the scenario helped reveal hidden requirements and concerns that were not always revealed directly by the work analysis itself.

Cognitive engineering approaches of this sort entail a level of complexity and nuance that is not encountered in more traditional classroom or laboratory studies. However, because of the broader range of issues considered in our "field research" approach, we believe that it is more likely than laboratory experimentation to reveal underlying factors that will enable recommended improvements in organizational, policy, and work systems design, and would enable technology support to have a more powerful, predictable, and lasting impact.

2.2. Target Topics and "Desirements"

Among the target topics (challenges to the macrocognitive work) that emerged from our observations and discussions were the following. Most of these are specific instances of problems that were actually created when tool developers took a designer-centered rather than a human-centered approach to design:

1. *Engaging automation as a partner in the rapidly-evolving process of sensemaking and response.* Analysts are accustomed to using a piecemeal set of software tools in the accomplishment of their work, pulling out a software "wrench" when a wrench was called for, and a software "hammer" when a hammer was called for. Each tool had been designed to perform one or more specific, generic tasks, but no tool really "understands" the overall work in which the analyst might be engaged. It was people who provided the know-how needed to use the tools, the sometimes-arcane routines needed to transfer data among them, and, most importantly, the understanding of the overall context and objectives that motivated and shaped the effort. When the tools were not merely passive, they were seen as adversarial—targets of pointed cursing because of their limitations (a phenomenon called "automation abuse" [77]).

   The dream of analysts was not a toolset, but a software teammate that would understand something about what they were trying to do and could actively assist them in overall sensemaking and response processes—both teaching them and being taught in an iterative process of mutual interdependence. Could today's stove-piped tools be integrated into a context-sensitive, task-aware, and assistive capability? A related problem is that both the nature of attacks and the details of work practice inevitably change much more rapidly than the traditional software development and release cycles currently support. Would it be possible to build technologies that could evolve as quickly as threats and responses do? Could a system be made to straightforwardly assimilate future analytic and response innovations that cannot presently be anticipated? Could the tools for creating that new work system be made simple and yet adaptive enough such that analysts could use them in do-it-yourself fashion?
2. *Reducing the great volume of uncorrelated low-level events.* Analysts tasked with monitoring and performing triage on network events can be overwhelmed by the massive volume of uncorrelated. low-level, and simplistic alerts and alarms with which they were continuously confronted. Analysts asked for better tools for the detection of complex anomalies, especially those that are context-specific or involve correlations across multiple data sources. They wanted help in understanding history and trends, so they could better understand what was

normal and recognize when significant long-term or short-term deviations in expected findings are taking place.

3. *Enabling continuous knowledge discovery and enrichment.* Analysts continually divide their time among a multitude of tasks. Their work in pursuing a given objective may be interrupted for hours or days while they deal with a sudden emergency. Tools that could continue to monitor relevant data sources in their absence, enrich results with pertinent information (e.g., geographic localization, entity identification and elaboration, database correlations), and organize those results on their own for later review by the human analyst were seen as having great potential.
4. *Overcoming the inadequacies of visualization tools.* Visualization tools were seen as inadequate in several respects. One problem is scalability. For example, parallel coordinate displays are not intelligible for any more than a few dozen network traffic records. Another problem was the form and content of what was presented. For instance, dashboard-style displays do not present information of different types in an integrated and meaningful fashion that directly answers analyst questions of central interest. Displays are typically technology-centered—focusing on what can easily be shown—rather than human-centered—focusing on what needs to be known. Display designs are fatiguing rather than appropriately stimulating to the eye and the imagination because they do not reflect sensitivity to issues of human perception and cognition.

   Another issue is a lack of interactivity—effective sensemaking requires not just "seeing" the data but also being able to probe and interact with it—and, in addition, requires the capability for the analyst to take action when necessary without having to move to a different display or software application. Displays are typically retrospective, showing something that had happened, rather than helping analysts anticipate what might happen next through the extrapolation of current trends, and assisting them in taking proactive measures when appropriate.
5. *Encouraging collaboration and sharing across individuals and distributed groups.* They face a plethora of information sharing challenges that sometimes lead to critical failures in achieving the common ground needed for understanding and effective action. First, analysts were sometimes unaware that information they possessed could be useful to someone else, or vice versa. Second, analysts are limited to specific means of communication (e.g., phone calls, chats) that can make it difficult and time-consuming to convey the richness of their observations. Third, the simplistic nature of today's digital policy management systems results in ambiguities about what could be shared with whom, and sometimes leads to out-of-band workarounds to circumvent inflexible systems when all else failed. Fourth, and most fundamentally, shared visualizations, such as those that might appear on large displays at the front of a room housing a NOC, have generally suffered from a lack of careful study of what kinds of information might actually be useful in such contexts.
6. *Minimizing the burdens of tedious everyday work.* Analysts complained about the amount of tedious and time-consuming work, including writing of a variety of report types. Awkward adaptations have proliferated as means to manage their burdens and to deal with the rigidity of tools and procedures. The ability to assess the status and progress of ongoing individual and group activities was sorely lacking. The need for a means of capturing and sharing knowledge with less-experienced analysts was expressed. Related to this problem was the loss of important "organizational memory" when analysts left or retired or when a case was "finished."
7. *Architecting for scalability and resilience.* Our interviewees said that they imagined that future analysts would need to be able to work securely and effectively in increasingly heterogeneous computing environments. Unfortunately, software systems are not usually designed with this forward look in mind. On the one hand, there is a need for a computing architecture that can automatically scale to varying computing and network resources. On the other hand, new kinds of computing devices, large and small, will continue to proliferate, and analyst will want to be able to use and synchronize their information across all of them. In addition, organizations will increasingly expect their technological support systems to be engineered for resilience, ensuring mission continuity, even when under attack or experiencing failures.

We are using the above target topics and "desirements" to guide the design and development of a cyber operations framework called Sol [4]. In the next sections we will describe our efforts to address the first target topic: engaging automation as a partner. In Sections 4-9, we will do likewise for the other topics.

3. Engaging Automation as a Partner

With respect to our first target topic, the analysts we interviewed were interested in engaging automation as a partner in the process of sensemaking and response. In order to lay the groundwork for a subsequent discussion of the details of the design of Sol, we first give an overview of what we mean by the term "sensemaking" (Section 3.1). We outline the role of software agents as partners in sensemaking (Section 3.2). We then introduce the concept of *coactive emergence* (Section 3.3). In doing this, we draw on the work of Johnson, who coined the term "coactive design" as a way of highlighting *interdependence* as the central organizing principle

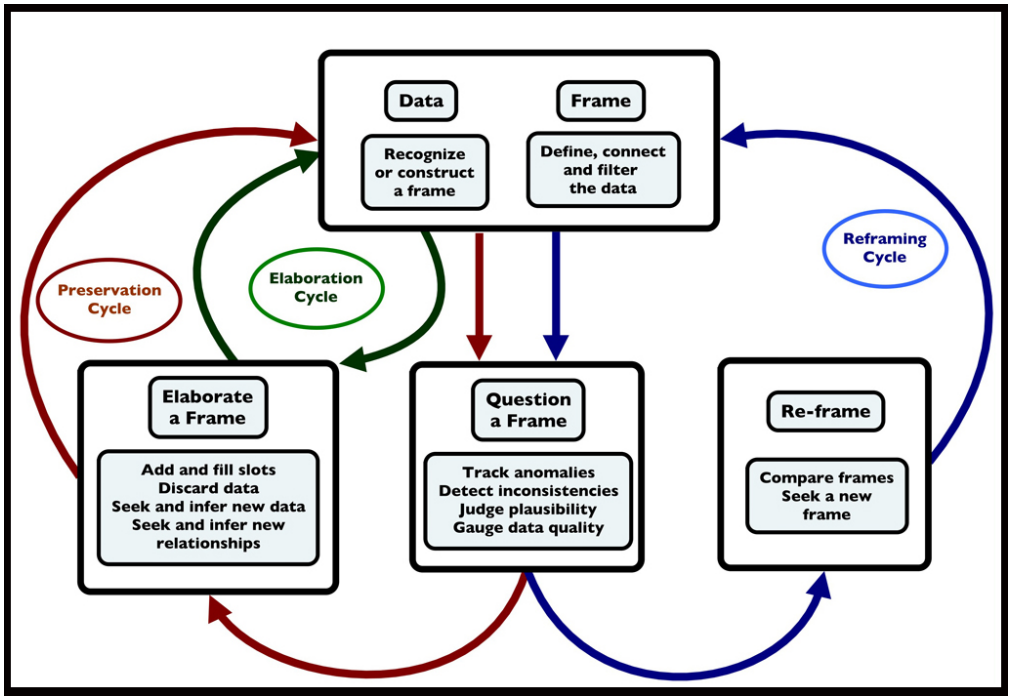

Figure 1. The Data/Frame Theory of Sensemaking

among people and agents working together [7][8][9][10]. We see coactive emergence both as a phenomenon of complexity and also as a strategy for the design of sensemaking work that combines the efforts of humans and software agents[†] in understanding, anticipating, and responding to unfolding events—both the foreseen and the unforeseen.

3.1. Sensemaking

As defined by Klein, *et al.* ([5], p. 71), sensemaking "is a motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively."

Figure 1 illustrates what Klein and his colleagues call the "data/frame theory of sensemaking" ([6], p. 89). At the most basic level, the theory acknowledges that understanding situations always occurs with respect to a framing perspective. The frame constitutes a set of more or less coherent hypotheses about the data to be understood, and serves both to determine what counts as data of interest and to shape the interpretation of the data. Note the absence of input and output arrows in the diagram. The sensemaking process can start, or recommence at any point, even though it is often triggered by surprise.

As data accumulate, the sensemaker may be confronted with the question of whether to elaborate a current frame by incorporating new details, or to seek a new frame that better accounts for current findings. The process involved in the ongoing evaluation of a given frame includes the possibility of a closed-loop alternation between backward-looking mental model formation—which seeks to explain past events—and forward-looking mental simulation—which anticipates future events.

The application of sensemaking concepts to the field of intelligence analysis (e.g., [11]) has looked at the ways to shape the sensemakers' investigative procedures in order to help them counteract lines of reasoning that might lead to misconceptions. A basic foundation for analyst sensemaking having been laid already in the research literature, a next step is toward implementation of a sensemaking support system that can harness the joint power of humans and machines. In particular, an understanding is needed of the potential impact of new forms of visualization and automation on the sensemaking process, and how such tools ought to be designed in light of what we already know. The emphasis of our own work on sensemaking is to put questions about the role and benefits of computer interaction with people in center stage.

In their discussion of the data/frame theory, Klein, *et al.* conjecture that the role of machines in assisting people with sensemaking may not be merely to confirm or disconfirm the accuracy of a particular interpretation with respect to a given frame, but also as an aid in the reasoning process that leads to the possibility of reframing: "The implication is that people might benefit more from intelligent systems that guide the improvement of frames than from systems that generate alternative understandings and hypotheses and foist them on the human" ([6], p. 89). This conjecture is consistent with the view of Woods, *et al.*, who have adopted a stance to resilience engineering that takes as its basic assumption that "human systems [are] able to examine, reflect,

[†] In this article, the term "agent," standing alone, will always refer to a software agent. Likewise "analyst" will always refer to a human analyst.

anticipate, and learn [i.e., engage in sensemaking] about [their] own adaptive capacity" ([14], p. 128).

It is in the spirit of these observations that we are exploring the concept of *coactive emergence* as a phenomenon that occurs in macrocognitive work systems [15]. Moreover, precisely because sensemaking provides a good model of macrocognitive work, coactive emergence can be used simultaneously as a strategy to guide work design. Below, we motivate the use of software agents as active peers in sensemaking and response processes (Section 3.2). Then we describe the concept of coactive emergence (Section 3.3). In later sections of the paper, we will describe our implementation of this approach within the Sol framework.

3.2. Agents as Sensemaking Partners

Many advantages of direct manipulation interfaces—those based on windows, mouse, and keyboard interaction—begin to fade as tasks grow in scale or complexity. Among other challenges, people are likely to encounter problems in dealing with large search spaces, passive reactions that respond only to immediate user actions, lack of composability of basic actions and objects, lack of sensitivity to context, orientation to generic software functions rather than an orientation to context-sensitive worker tasks and needs, lack of long-term temporal continuity, and no improvement of behavior. Researchers at IHMC have been pioneers and innovators in software agent technology [16][17][18] which addresses these problems by combining the expression of user intention through direct manipulation with the notion of an *indirect management* style of interaction [19][20]. By their ability to operate independently in complex situations without constant human supervision, collaborating teams of agents can perform tasks on a scale that would be impossible for other approaches to duplicate.

Software agents are typified by their active, adaptive nature. This quality is often characterized in the Artificial Intelligence literature by the word "autonomy." However, as we have argued elsewhere [21][22], autonomy sounds like just the *wrong* word for characterizing agents like ours that are designed to assist, rather than replace, people. Though we are certainly interested in making these agents more active, adaptive, and functional, the point of increasing these proficiencies is not merely to make the machines more independent when independence is required, but also to make them more capable of sophisticated *interdependent* joint activity with people [7][8][9][10]. In addition to being able to hand off their tasks to such agents, people need to be able to work in simultaneous collaboration—i.e., coactively—with them, participating in joint activity in a fluid and coordinated manner [23]. In this way, well-designed software agents, in their ultimate manifestations, become teammates rather than tools [6][21][24]. This is consistent with our ultimate goal that these agents be more than mere processors of data, but, in addition, that they be capable of the more demanding requirement of full engagement as assistants to analysts in the sensemaking process itself.

Our cognitive task analysis identified three aspects of sensemaking that could be supported by software agents:

1. *Identifying and understanding cyber threats (making sense of taskwork).* Not surprisingly, taskwork receives more attention than other facets of sensemaking in analyst training. Identifying the central challenge of human analysts in this regard, Branlat, *et al.* ([12], p. 7) have insightfully observed that, in cyber defense analysis:
   > … the detection of elementary and potentially suspicious traces of activity does not seem to be the main problem, apart from the latency [when analysts and their tools cannot keep up with events]. The bigger issues are determining… what it means in terms of purposeful actions perpetrated by the attacking team;
2. *Helping maintain common ground and facilitating coordination among human and agent team members (making sense of teamwork).* Among other things, such information helps analysts become aware of pertinent information coming from others, synchronize handoffs, and realize when progress is running ahead or behind expectations;
3. *Helping people and agents maintain awareness of background information relevant to their activities (making sense of work context).* In NOC, this facet of sensemaking is addressed in part by large displays in the front of the room where breaking news and statistical summaries are posted for all to see. Much can be done to increase the effectiveness of current wall displays of this sort.

We now discuss the first aspect of sensemaking describe above. The other two aspects will be discussed briefly later in the paper.

3.3. Coactive Emergence

Coactive emergence describes an iterative process whereby secure system configurations, effective responses to threats, and useful interpretations of data are continuously developed through the interplay of joint sensemaking and decision-making activities undertaken by analysts and software agents. The word "coactive" emphasizes the joint, simultaneous, and interdependent nature of such collaboration among analysts and agents.

Ideally, the process of coactive emergence is synergistic, leading to progressive convergence on threat hypotheses. Of course, competition among hypotheses is also desirable in sensemaking in order to encourage the exploration of the same space (or a wider space) from different perspectives and to avoid premature closure.

We will explain our application of the phenomenon of coactive emergence in more detail below, following a discussion of the task-artifact cycle.

The task-artifact cycle. In contrast to more typical software development practice, human-centered design requires a co-evolution of the worker task and the technology artifact, as articulated more than two decades ago in John B. Carroll's task-artifact cycle [25]. The task-artifact cycle (Figure 2) includes two parts: the first involves the design and development of artifacts to help workers perform their assigned tasks; the second concerns the way that the use of the artifacts defines new perceptions, possibilities, or constraints of use that change the way the task is performed. Though the basic concept is a good one, the development cycle as typically implemented is too slow to keep up with the fast pace of change in threats and analyst practice [78]. To speed up the process of parallel evolution of tasks and work practices, we are proposing tools and methodology based on the concept of coactive emergence.

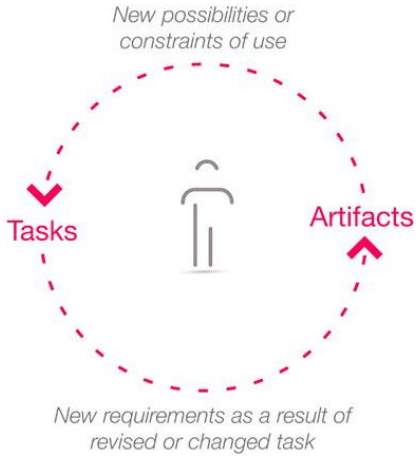

Figure 2. The Task-Artifact Cycle

The coactive emergence cycle. Because the definition of new agents, agent tasks, and redirection of agent activity can occur interactively at run-time, the rate of co-evolution of work practices and system activity can better keep up with continuous changes in threats than it currently does in typical software development practice.

The goal is to combine the know-how of people and agents in an ongoing scaffolding process. As new kinds of threats are discovered, agents can be directed to detect patterns that will identify and characterize them. In addition, some patterns can be learned by agents automatically and presented to analysts for validation, as illustrated in Section 5.3. Moreover, as new analytic innovations are developed, new kinds of sensing agents also can be straightforwardly added. The current version of Sol contains graphical interfaces to allow analysts to perform all these tasks in specific instances. To make these into general-purpose capabilities will require considerable elaboration of these interfaces, but not of the basic architectural framework that already supports them.

Figure 3 illustrates how a cycle of coactive emergence applies to cyber sensemaking. There is an analogue here to the task-artifact cycle, except that we consider the artifacts in this case—i.e., the agents—to be coactive with

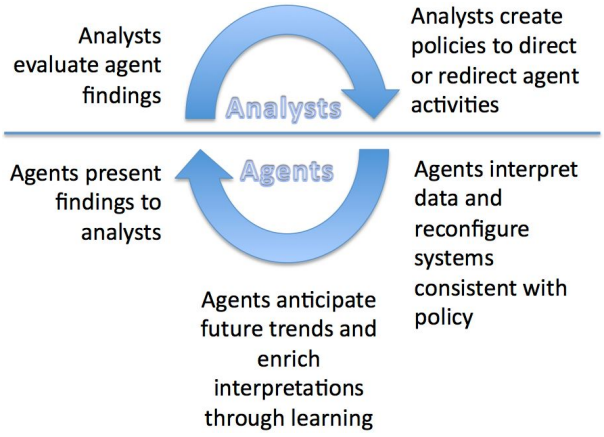

Figure 3. The Coactive Emergence Cycle

the analysts in their efforts to improve the performance of the macrocognitive system. Note that the diagram showing a single loop is somewhat misleading, since multiple threads of agent and human activity would be operating on individual schedules rather than in lockstep as the figure implies:

1. Agents are pre-coded in Java to perform particular classes of common tasks (e.g., tagging, correlation—see Section 4.2). Analysts use their knowledge to characterize previously-known and hypothesized patterns of attacks and to encode these patterns into high-level declarative policies that enable the agents to detect and monitor them in a secure, predictable, and controllable manner.
2. Subsequently, agents tag real-time data containing these patterns for presentation to analysts. Within the constraints of policy, agents may not only sense but also act—for example, manipulating system configurations to improve security.
3. Agents may optionally enrich their findings with additional information gleaned through learning (e.g., hypothesized correlations between sets of suspicious flows, anticipated future trends). Because of their built-in abilities to work together dynamically to analyze and synthesize meaningful events from the raw data, agent interpretations can be more easily made to match the kinds of abstractions found in human interpretations more closely than those that rely exclusively on low-level sensors.
4. Agents may aggregate and present their findings by visually annotating graphical displays in real-time in order to highlight and draw the attention of the analyst to anomalous or otherwise interesting elements, such as possible attacks. We call such displays *mediating representations* — highly-communicative visual models of the situation that can be simultaneously used by mixed teams of people and software agents in order to come to a common understanding of a situation [26][27]. For example, in our Flow Capacitor visual display (to be discussed in Section 6, and shown in Figures 11 and 12), highlighting and coloration of threat information

reflects real-time agent-tagging. In addition, we capitalized on additional suggestions by practitioners by allowing flows of interest to be further annotated by attaching "flags" to the end of colored segments of particular flows. The flag colors and what they indicate (e.g., type of attack, presence of flow source in blacklist) can be customized by the analyst. Analysts interact with these displays in order to explore and evaluate how agent findings bear on their hypotheses.

5. As agent-derived information is presented to analysts, they may agree or disagree with agent findings, leading to further corrections and refinements of interpretations, and consideration of response options.
6. Analysts continue to direct and redirect ongoing agent activity through the construction of new agents, modification of agent policies, and extensions to lines of inquiry.

Note that the process of emergence operates at two levels:

- First-order patterns emerge from agent and analyst interpretations of data that are shaped by problem-space constraints currently expressed within policies and tool configurations (cf., e.g., [28], pp. 117-118). This is related to the process of frame elaboration in the D/F model of sensemaking, where the agent policies currently in place drive the interpretation of incoming data. For example, through the application of analyst-defined policy-based pattern recognition, agents may tag and display selected network data as instances of emergent threats. Likewise, a display of current agent results may lead analysts to recognize the possibility of additional emergent threat patterns that the agents may have missed.
- Second-order emergence arises from dynamic changes made by agents and analysts to the problem-space constraints (cf., e.g., [29], p. 90). This is related to reframing processes in the D/F model of sensemaking, where agent policy modifications made by people or the agents themselves change the way data is being interpreted. For example, analysts may add, delete, or change agent policies in order to refine their data interpretations or to modify their responses to threats. Agents may also change their own policies through policy learning. When permitted, agents may also propagate learned policies to other agents.

Benefits of the approach. As with all forms of macrocognitive work, our emphasis on the coactive participation of humans and machines in threat understanding proceeds from the premise that the use of teams involving such a mix can increase the range, richness, and utility of models that could be explored by people or computers alone. In human-agent teams, people occupy a privileged position as compared to machines because, among other things, they generally know more about the way that joint tasks interact with broader ongoing activities and with the situation at large. For these reasons, humans have an important role in keeping software agent taskwork aligned with its wider contexts [30]. In their complementary role, software agents can help people cope, for example, with the volume, tempo, computational complexity, and highly-distributed nature of joint tasks. In addition to supporting appropriate aspects of taskwork, agents can be used to help various aspects of team process, as will be discussed in more detail in the next section.

Having discussed our general strategy for engaging automation as a partner in sensemaking and response through software agents and analysts working together in a cycle of coactive emergence, we will now discuss a specific role of agents in addressing target topic two: reducing the high volume of uncorrelated low-level events.

4. Reducing the Volume of Uncorrelated Events

In monitoring complex, high-tempo data streams, it is impossible for a human to keep up with the typical flow of uncorrelated low-level events. Rather than requiring workers to rely on direct sensing of the network alone, context-sensitive software agents enable analysts to have mediated access to correlated data and information. For example, one of the agents' principal sources of data is NetFlow records [79]. These records contain information about source and destination addresses of network packets, protocols and ports used, size and rate of the flow, and other information. Agents are organized hierarchically to facilitate the enrichment of NetFlow records at multiple levels of abstraction. In this way, agent annotations do not merely highlight low-level indicators of intrusion patterns, but instead directly identify the type of intrusion itself. For instance, instead of requiring the analyst to notice that a configuration of connecting lines (some of which may be obscured in a typical display) indicates a distributed port scan, agents working on abstracted data semantics can directly indicate the source and nature of the attack. As another example, if a message is anomalous because it is sending oversized packets to a port associated with an SQL database, higher-level agents can abstract that message and represent it as an instance of an SQL injection attack. This ability to reduce perception and reasoning requirements on the analyst is a major benefit of agent-based analytics.

Agent characterization of the data in terms of identifiable intrusions enables analysts to carry out standard procedures in response. These procedures could include the automatic configuration of visual displays that allow the analyst to isolate intruder actions, or the spawning of new agents to collect data related to the identity of the network threats. In related projects, we are using agents to perform interdictory actions to prevent the

intrusion from propagating further or wasting more network resources.

Before giving specific examples of how this is done, we will present an overview of our Luna agent framework, named for the founder of Pensacola, Tristán de Luna y Arellano (1519 – 1571).

4.1. Luna Agent Framework Overview

IHMC's Luna is an agent framework designed for the demands of cyber operations. Within the Sol cyber framework, Luna agents function both as interactive assistants to analysts and as continuously-running background aids to data processing and knowledge discovery. To facilitat e their use as sensemaking partners to analysts, we have designed our software agents to be comprehensively governed by semantically-rich policies that are defined, analyzed, and enforced by IHMC's KAoS policy services framework [31][32], enabling a high level of assurance in their deployment.

KAoS policies are of two primary types: 1) authorization policies that permit or forbid a given action by a given agent or class of agents in a given context, and 2) obligation policies that require or waive requirements for a given action to be performed when triggered in a specific situation. More complex policies governing things like delegation, goal refinement, and collective obligations are built out of these two basic kinds of policy.

In addition to their role in directing the taskwork of agents and in assuring safe and secure operations, machine-interpretable policies, enforced by KAoS independently of agent code, are also the primary means by which good teamwork practices by software agents are assured [33][34]. In addition to regulating task-specific behavior whose details may also be directed by runtime-modifiable policy constraints, each agent is governed by policies designed to assure its observability (e.g., mandatory status updates at an appropriate frequency, or in response to specified events), directability (e.g., immediate responsiveness to redirection due to policy changes), interpredictability (e.g., obligation policies assuring that required behavior will be executed within a specified time period), adaptation (e.g., policies governing the range of adaptations permitted and the process of propagation to other agents), support for multiplicity (e.g., policies governing synchronization of multiple perspectives), and trustworthiness (e.g., policies assuring the observability of parameters indicating the reliability of agent operations). Luna also relies on KAoS for capabilities such as registration, discovery, self-description of actions and capabilities, communications transport, and messaging. A detailed technical overview of Luna with many examples of how it exploits policy governance can be found in [35].

Figure 4 shows how KAoS integrates with the Luna environment and individual agents (A, B, C, and D) to provide services and to enforce policies. An OWL

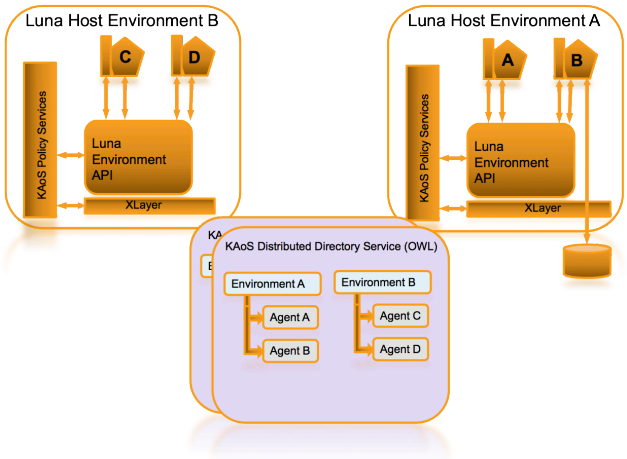

Figure 4. Luna conceptual architecture

representation of Luna is maintained within the KAoS Distributed Directory Service. Through its interactions with the Luna host environment, KAoS regulates the lifecycle of both the environment (e.g., start and stop Luna) and the agents (e.g., create, pause, resume, stop, and move agents). Policy can also regulate environment context for shared agent memory (e.g., getting and setting its properties), allowing efficient parallel processing of large data sets. In the future, KAoS will also integrate with our Xlayer capability, VIA, to provide a means of policy enforcement outside the Luna host enironment (e.g., to govern agent B's access to a networked database, as shown in the figure). An agent-based implementation of context mirroring across different Luna environments is provided. Through policy, the Luna host environment also governs agent progress appraisal—a subject to which we will return later.

In order to support dynamic scalability, load balancing, adaptive resource management, and specific application needs, the Luna platform supports the policy-governed option of allowing the *state* of agents (vs. *code* of agents) to migrate between operating environments and hosts. The Luna environment maintains agent mailboxes with message forwarding when agents migrate. Luna state mobility will provide the foundation for future implementation of agent persistence (i.e., saving and loading agent state to a persistent store).

Within the base class for Luna cyber agents are defined some common agent tasks that can be called through OWL descriptions. However, one of the most important innovations in Luna is the ability to add custom agent actions to the policy ontology, based on their Java equivalent. This allows any newly defined Java-based agent capability to be brought under full policy governance. IHMC provides a Java2OWL tool to assist with this task. The Java2OWL tool can be used to browse custom agent code, select methods to bring under policy control, and automatically generate an OWL description for the selected method signatures. These methods are then immediately available for policies as Actions performed by Agents of that type.

4.2. Applying Luna to Event Processing

A demanding role played by Luna agents within our Sol cyber framework is its responsibility for multi-layer agent processing and tagging of live or retrospectively played-back NetFlow data representing worldwide Internet traffic. A high-level view of roles and relationships among agents relating to these functions is shown in Figure 5.

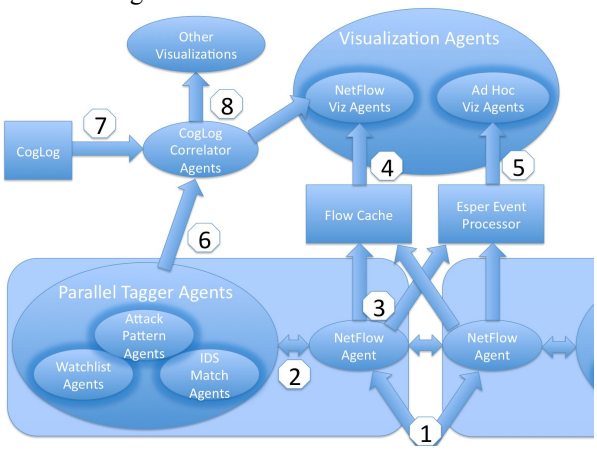

Figure 5. Agent processing and tagging of NetFlow

Incoming UDP traffic goes to a NetFlow agent for parsing and transformation into Java objects (1). In principle, the same or different data could be routed to multiple NetFlow agents on the same or different hosts to share the processing load. The NetFlow agent sends the data to any number of Tagger agents that work in parallel in real-time to tag the data (2). For example, Watchlist agents tag data that appears on whitelists or blacklists while IDS Match agents tag data corresponding to intrusion detection alerts. Drawing on selected results from low-level tagging agents, Attack pattern agents may be defined to look for higher-level attack patterns. By this means, agent annotations do not merely highlight low-level indicators of threat patterns, but can directly identify the type of threat itself, as described earlier. A system of semaphores ensures that all the Tagger agents have completed their work before the NetFlow agent sends results to the Flow Cache (3).

NetFlow Visualization agents enforce policies that mediate data being sent to analyst displays, ensuring, among other things, that data not authorized for viewing by particular users are automatically filtered out (4).

The Esper complex event processor [36] provides support for efficient ad hoc queries of many types that can be initiated and consumed by other visualization agents (e.g., our Stripchart View agent) or by agents of other types for further processing (5). We are also considering the use of Esper for data stream handling further upstream in the agent analytic process.

CogLog Correlator agents ingest combined data from selected Tagger agents operating on real-time data (6) and historical data. within an interactive archiving tool called the CogLog, short for Cognitive Case Log (7). The CogLog is described in more detail in Section 5 below. Unlike the real-time Tagger agents, the Correlator agent can perform deeper kinds of analytics in "out of band" mode. Among other things, this correlated information can help different analysts "connect the dots" between related investigative efforts—e.g., when one or more ongoing cases might overlap in interesting ways with cases recorded within the CogLog. The Correlator agents may send additional data annotations to NetFlow Visualization agents and/or to agents supporting other visualizations (e.g., Connection Graph view, as shown in Figure 8) (8). Our Attack Pattern Agents provide another example of an out-of-band agent type. These agents consume and process all NetFlows (rather than just subsets of tagged data produced by Tagger agents) in order to learn and propagate useful threat patterns.

In the future, exploration of larger questions of adversarial intent, attack strategies, and social connections among attackers could also proceed along similar lines of increasing abstraction in agent processing. The ability to reduce perception and reasoning requirements on the analyst through fixed or ad hoc organizations of agents processing low-level or abstracted visual and logical data dimensions in a correlated way is a major benefit of agent-based analytics.

In the next section, we will describe some of the ways we use agents to facilitate continuous knowledge discovery and enrichment.

5. Continuous Knowledge Discovery

In this section, we will introduce three capabilities meant to address aspects of the problem of continuous knowledge discovery and enrichment. Unlike the mature agent, policy, sharing, and visualization capabilities of Sol, these tools are working demonstration prototypes.

5.1. The Analyst Chat Assistant

Agents promote continuity in investigation by continuing to work when analysts are unavailable. They can free up analyst time by performing tedious, distracting, complex, and high-tempo chores. For example, agents can not only keep up with real-time tagging of individual flows, but can also work continuously in the background to discover higher-level patterns, such as significant deviations from expected network traffic levels.

We created the Analyst Chat Assistant as a working prototype to demonstrate the potential of agents for such tasks (Figure 6). Within this tool, agents monitor background chat sessions and annotate specified data of interest that match certain criteria (e.g., IP addresses contained within a watchlist). Such addresses are automatically enriched by other agents that are tasked to look up additional metadata. Significant findings may be categorized and posted automatically by additional agents to the CogLog, described below.

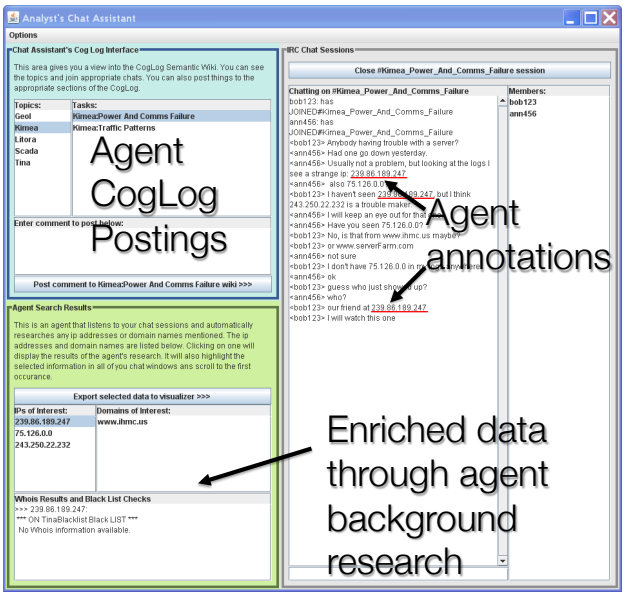

Figure 6. The Analyst Chat Assistant

5.2. The CogLog

As an aid for analyst investigative and reporting tasks, agents can also collect specified types of information concerning workflow and investigation results into a working demonstration prototype we call the CogLog (Figure 7). The CogLog is a semantic Wiki-based tool within Sol that contains a log of findings pertinent to a given investigation, while also linking to other relevant information from prior cases. Information associated with each case can be logged and maintained while analysts jump from chore to chore, and from case to case. As the figure shows, both analysts and agents can post data. These posts can range from the mundane (e.g., IP addresses, names, pictures) to more abstract entities like lines of inquiry or "blind alleys." It is easy to envision how libraries of data of this sort might represent an important kind of knowledge management capability for analytic work. Such data could be cross-referenced in future investigations, supporting a form of case-based inquiry.

We have prototyped correlation agents that implement capabilities for making connections of different types by continuously doing knowledge discovery: looking for relationships among items of data, people, cases, analysts' activities, and lines of inquiry across individuals and groups of analysts. For example, Sol supports the ability for a KAoS obligation policy to be defined to enable the automatic creation and commissioning of a new agent to look for additional data or metadata relevant to a set of flows whenever the analyst makes a selection using a pointer gesture. As a result, the agent might signal to the analyst that others are also working on related threats when it discovers a given IP address in a live chat interface or within a previous case record in the CogLog.

In Figure 7, we've colored different sections to show how different kinds of CogLog postings can combined within the same case record. The blue pane at the top shows excerpts from chats within the Analyst Chat Assistant that were posted by an analyst. The green section, featuring a graph showing spikes in network activity, was automatically posted by an agent. The yellow section shows an image created in the Flow Capacitor visualization based on information exported from the Analyst Chat Assistant.

Figure 7. The CogLog

5.3. Agent Learning

Agents can augment human pattern recognition by learning new threat patterns and presenting them to the analyst for validation. For instance, in order to identify additional attacks and targets that analysts may have missed, a group of attacking flows and their targets could be selected, and an agent using our working prototype of biologically-inspired learning mechanisms [37][38] can be launched to find additional, similar flow patterns. Figure 8 shows an example where an agent has posted the results of its learning to a connectivity graph display. The green node at the upper right-of-center represents one of the power plants belonging to an analyst's own organization, along with the tan-colored attackers and their presumed command-and-control node. At the lower right is a green node that is a likely next target, due to the fact that it is now experiencing scan attacks from two tan nodes and has the same configuration and vulnerabilities as the first power plant. The large node just to the left of center is another likely target that sits outside the analyst's own network. In this way, agent learning can help the analyst anticipate additional attacks and potential new targets that otherwise might have been overlooked.

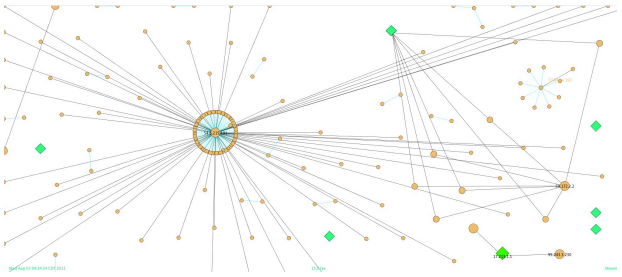

Figure 8. Connectivity Graph showing agent learning results

6. Better Interactive Visualizations

Some of the most important unresolved issues about visualization and sensemaking concern how displays should be designed and evaluated. Most of the past work in this vein has been guided by intuition rather than principle, and has been evaluated by anecdote rather than empirical analysis. While some amount of this is unavoidable (and, in fact, desirable), we aim to do more to develop a theory-based visualization design methodology.

Building on lessons learned at IHMC about principles of effective interactive visualization design, we will give examples of how knowledge about human perception, cognition, and collaboration relevant to real-time sensemaking tasks can inform the design of particular instances of visualization. We will emphasize the role of exploration and real-time interaction with such displays as a means of enhancing human understanding.

6.1. Principles for Effective Visualizations

Our approach to real-time cyber sensemaking displays is informed by lessons learned in the design of IHMC's highly-successful OZ‡ flight display [39][40]. Though the display's reliance on colored lines and dots on a black background may seem a primitive throwback to first-generation video games, this simplicity is by design, based on a sophisticated understanding of the latest research results in human perception and cognition [82]. Due to its specially-designed features, experimentation has repeatedly demonstrated the superiority of OZ over traditional displays in minimizing pilot error, reducing pilot disorientation, and maintaining situation awareness.

We now discuss in more depth some of the lessons learned that have been applied in the development and refinement of OZ, and that we have relied on, when appropriate, in the design of cyber sensemaking displays.

Ambient Vision Channel. The visual field can be divided into three channels, the focal, the peripheral, and the ambient. The focal channel is used for tasks such as reading, which require directed attention. The peripheral channel is useful in noticing movement, and may be performed with or without directed attention. The ambient channel is used primarily for tasks involving both focus and movement, such as locomotion that can be accomplished without conscious effort or even awareness. For example, ambient vision is used by people to quickly and successfully navigate crowded hallways without conscious thought or to catch a football on the run [41][42][43][44]. In the normal environment all of these channels are simultaneously active, as when a running quarterback passes the ball to a receiver or when a driver reads a sign while controlling an automobile during a turn.

In designing interfaces, perhaps the most restrictive visual channel is the peripheral, which has been primarily used for alerting the operator to changes in the work environment [45]. There is considerable flexibility in designing interfaces for focal vision, as the full range of reading and symbol comprehension can be utilized. However, this flexibility can often occur at the expense of speed: requiring foveating on a number of spots to obtain needed data can put the operator behind the pace of operations and can result in an overall decrease in performance. This is an odd situation, where more information results in lower performance. Displays relying on ambient vision occupy a middle ground between displays designed for use by the peripheral and foveal vision channels. Because of this, ambient displays can excel when there exists a large amount of information requiring continual monitoring and response.

Proportionately-Scaled Symbology. This widely-applicable design principle can be used in virtually any type of interface design. In this approach, symbology is proportionately-scaled when the communicative aspects of the symbology are not overshadowed by the size, shape, or change in other neighboring symbologies. For example, having small, slow-color-changing glyphs arranged next to large, fast-color-changing glyphs may reduce the operator's ability to detect and respond to the slow color changes. Sizing these appropriately ensures that the information to be communicated by the symbology is neither muted nor excessively exaggerated. The symbology is constructed of visual primitives that are resilient to optical and neurological demodulation, which exploits both ambient and focal vision. For example, OZ uses color, shape and scale (i.e., spatial frequency) to construct primitives that, when viewed over a high-contrast dark background, have increased legibility, allowing pilots to distinguish the elements clearly.

Holistic Foreground Against Contextual Background. Displays designed to be processed by the ambient visual channel can take advantage of movement sensitivity and large field of view when constructing the visual elements of the display. One effective approach relies the principle of constructing symbology relating to the subject of the interface as a holistic foreground element of the display and filling the background behind this element with symbology that conveys contextual information [45][46]. The starfield and tri-plane wings of OZ are a good example of this. The starfield continually conveys

‡ OZ relates to the classic film "The Wizard of OZ" and is not an acronym.

contextual information concerning aircraft altitude, attitude, heading and relation to other objects, while the foreground elements indicate aircraft performance.

Structure from Motion. This is the phenomenon in which people naturally construct meaningful objects based on the movement of a small number of elements. Although this example has many more points than necessary for a real-time demonstration it is helpful to show as a printed image. As the points move, it is easy to understand the picture as a rotating sphere. OZ exploits these capabilities by using movement to convey difficult, correlated information [48][49].

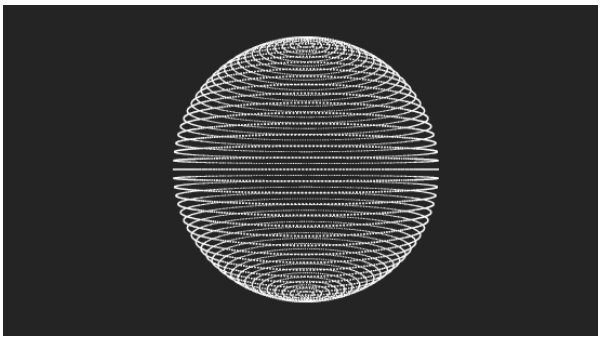

Figure 9. Structure from motion example

Pop-Out. Pop-out is a phenomenon that occurs with visual search when features of the search target are significantly different from its surroundings that the target becomes the most salient element of the visual search field [45][50]. In the example below, finding the red circle is much easier in the first field than in the second. In the second field, the design of the dots in the field have overlapped too closely, making the task of distinguishing the target much more difficult. In other examples, the misalignment of a foreground element to a background element can be seen as "popping out", too. As part of a control interface, noticing and responding to these differences early can significantly improve overall operator performance.

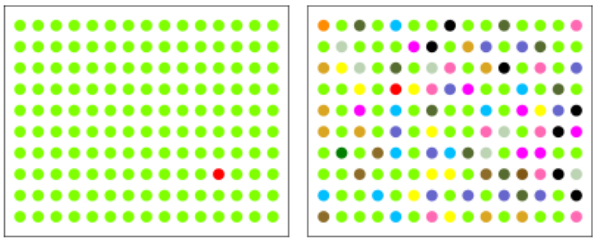

Figure 10. Pop-out example

Chunking. Visual displays of complex data can benefit from "chunking" conceptually-interrelated stimulus units [51]. This is because the human mind can commit more of these visual elements to short-term memory when they are organized within such chunks than if these elements were presented in a disassociated manner. This principle is used to great advantage in the OZ display, where complex interrelationships within and between the aircraft and starfield metaphors help the pilot to retain the current state of the world.

6.2. OZ Principles in the Flow Capacitor

We have used lessons learned from the OZ flight display in our design of visualizations for cyber situation awareness. Consider, for example, a visualization we call the Flow Capacitor (Figure 11).

The Flow Capacitor. The Flow Capacitor is a highly-configurable, interactive 3D visualization of Internet traffic. The input to this visualization is NetFlow records.

The two planes at the top and bottom of the display are mirror images of each other. The top plane shows a "Source IP Map" of the NetFlow records and the bottom plane shows a "Destination IP Map." Each of the two planes shown in Figure 6 represents the full IPv4 address space, where each point on a plane is a unique IP address—defining, in this case, a model of 65,536 pixels cubed. The 256 grid boxes on each plane divide the IPv4 space by the first octet in the address, the class A network. Due to the modularity of the agent architecture, upgrading to IPv6 will be straightforward.

The record of a given flow at a specific moment of time is represented as points on the source and destination planes, creating a result similar to heat maps. The color of the source and destination points encodes the first three octets of the IP address (i.e., the class C network address). Users can drill down at any time to see a more detailed projection of the traffic on a plane, displaying, for example, current flow records from or to all addresses *within* a given Class A network.

As alternatives to the IPv4 maps shown, any number of alternate plane types can be defined. For instance, the framework can geo-locate the IP addresses and project the source and destination locations as latitude and longitude on a map of the world. Conceptually-defined planes, categorizing flows from certain types of groups (e.g., criminals, nation-state attacks) or economic sectors (e.g., financial, energy) can also be defined.

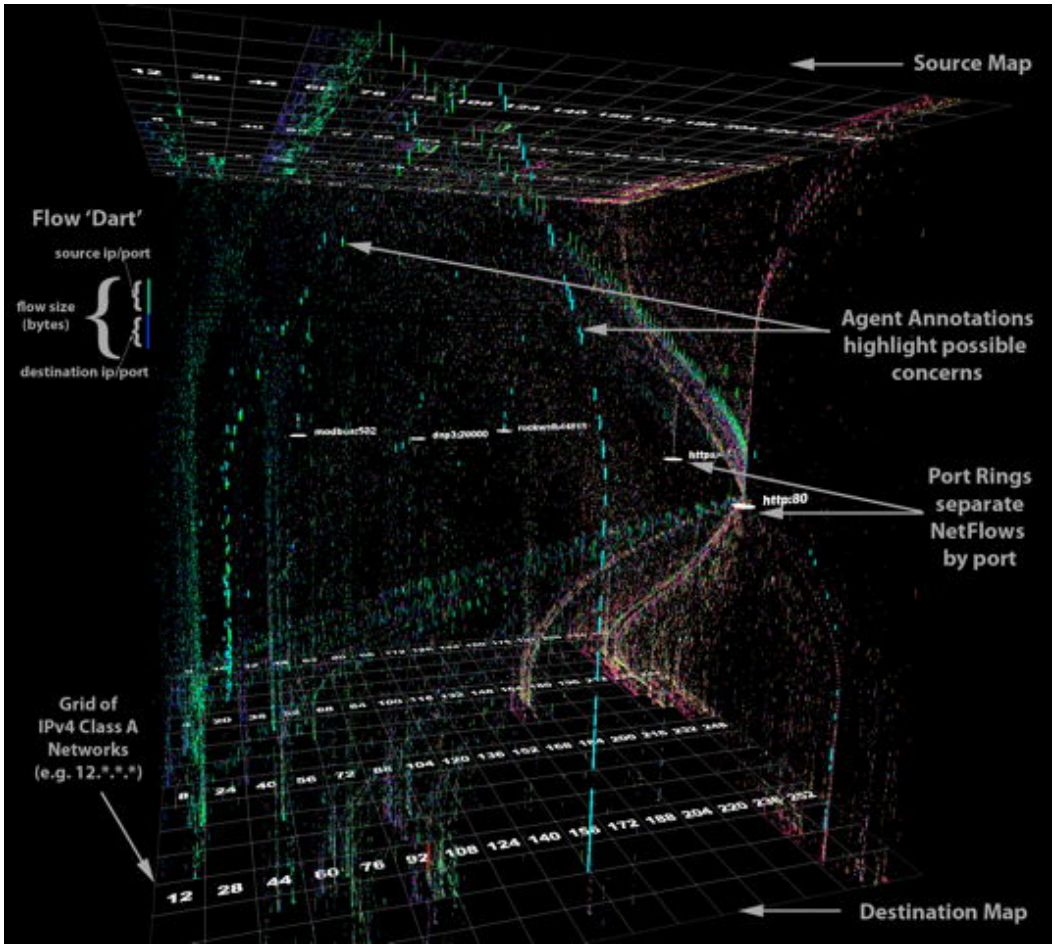

Figure 11. Annotated flow capacitor example

NetFlow "darts." In addition to being shown on the source and destination map planes, each NetFlow is also represented as a short line segment or "dart" that moves in real time from a source in the top plane to a destination in the bottom one. The length of the dart is proportional to the number of bytes that are being transferred between the source and destination by that flow. The appearance of the top half of the dart reflects attributes of the source plane, while the bottom half reflects attributes of the destination plane. For example, the two halves of the dart may be shown in one of 65,536 unique colors corresponding to the source IP and the destination IP. Alternatively, for example, the colors could be defined to correspond to the port number. The properties on which the colors are based and the particular colors chosen for a given property value can be easily redefined to represent other flow attributes such as protocol, duration, and TCP flags.

Port rings. The white rings labeled with protocols and port numbers (e.g., http:80, https:443) "attract" NetFlows that have a matching source or destination port value. This allows them to be visually grouped by the ring as they travel downward. The rings are initially placed in sorted order, but can be manipulated with a pointing device. For example, an analyst can interactively move the ring to a less congested area of the display in order to more easily separate and monitor certain kinds of traffic. Besides ports, other kinds of properties can also be used to define rings.

User controls. Configuration of user controls is performed graphically on-the-fly in auxiliary window panes. A pointing device can be used to rotate, zoom, and pan the view interactively. Modifier keys are used in conjunction with mouse actions (e.g., click, drag) in order to differentiate user intent. A vertical timeline with configurable color-coded key event annotations provides a temporal overview of the unfolding situation and incorporates a slider control for quick navigation through time (see Figure 12). The user can pause, rewind, and fast-forward the display for instant replay in slow- or fast-motion—enabling users to engage in different kinds of attentive and preattentive visual processing of the information [83].

Pausing the display enables the user to mouse-over individual flow darts to display flow metadata. To allow easy selection, darts can be made "bolder" automatically when the display is paused. In addition to specific dart selection, individual flows or groups of flows can be selected for more detailed analysis by software agents or for viewing in other kinds of displays. Selections of interest can also be shared among different individuals and groups.

The period of time represented between the top and bottom planes can be configured to any length, from

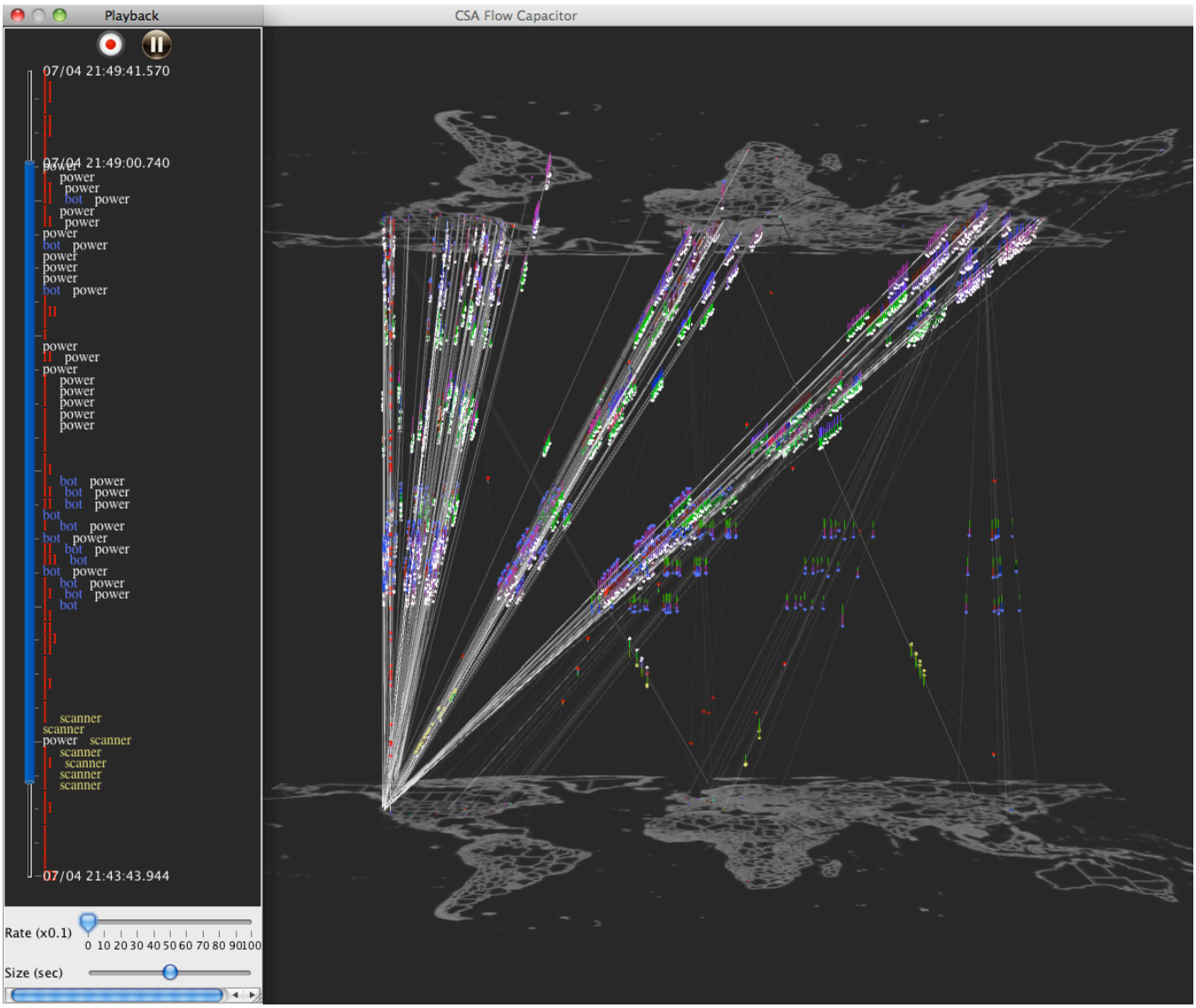

Figure 12. Distributed denial-of-service attack example

weeks or days to milliseconds. Slider controls below the timeline allow the user to specify the time frame of interest and the rate at which time passes. The slider control on the vertical timeline is automatically sized to indicate the proportion of time in the currently-displayed slice relative to the length of the overall timeline. The user also determines whether to render all of the NetFlow records or to filter them based on a combination of protocol, port, IP, and country.

6.3. Distributed Denial-of-Service Example

From a single snapshot of the Flow Capacitor in Figure 12, we can see the sequence of events leading up to a distributed denial-of-service attack portrayed in graphic clarity. Reading from bottom (oldest events) to top (most recent events):

1. Blacklisted scanners [yellow] get control signals from some unknown command-and-control node not yet on our blacklist (yellow flows over Italy)
2. Blacklisted scanners [yellow] hit whitelisted power infrastructure nodes [white] on US west coast (four streaks of yellow)
3. Some power infrastructure nodes respond to the scanners (yellow and white flows cross the tail of the scan attacks with the yellow tags at the opposite end of the darts). There are two sets of four darts moving diagonally from left to right. The set on the left (over the Atlantic) consists of responses from California and Washington to the scanners in Italy. The set on the right consists of the scanners in Italy subsequently passing these responses on to a C2 node in China.
4. Blacklisted bots [blue] receive control signals from their C2 (burst of blue from one to many on the right)
5. Blacklisted bots attack whitelisted power infrastructure (blue and white "tornados")
6. Unknown nodes, not yet on our blacklist, attack whitelisted power infrastructure nodes (white "tornados").

6.4. Parallel Coordinates 3D Observatory

The Flow Capacitor is a robust and mature capability that can be used to answer a core set of important questions about the number and nature of flows between sources and destinations in networks of any size. In order to broaden the range of questions that can be asked, we devised an additional demonstration prototype showing a generalization of the concept of a Flow Capacitor called the Parallel Coordinates 3D Observatory—"PC3O" or "Observatory" for short.

Toward a performance model for network analysis. Although nearly all of the principles behind the OZ flight display design have been applied in new ways in our work on cyber sensemaking, there are some aspects that have proven more challenging. These additional challenges help inform the design of the PC3O.

One of the most important differences between these two applications is the difficulty in finding the equivalent of the flight performance model for network analysis. Whereas the primary task of the pilot is to fly effectively within the known parameters of a fixed aerodynamic model, the job of the NOC analyst is to understand emerging threats accurately against the moving target of a network that is constantly changing. With this fact in mind, it is easy to see that what the analysts need is not a control device, nor merely an informative picture of the world, but rather a tool for *formulation, exploration, and testing of hypotheses* about a situation—essentially the framing and reframing aspects of sensemaking ([6][52], p. 286). In our view, the utility of a given approach to cyber sensemaking should be evaluated pragmatically in terms of its effectiveness in asking and answering a serviceable range of relevant questions.

Comparison of PC3O to Parallel Coordinate Graph approaches. Parallel coordinate graphs are a common way of visualizing data with a large number of constituent features.[§] These graphs show connections between feature values based on a given set of data, usually with each feature dimension represented by a vertical line, which normalizes that features values in to a continuous range over the length of the line, or in equally spaced points for discrete feature values. For example, Figure 5 shows a parallel coordinates type display called VisFlowConnect [53]. External senders are shown on the left, internal hosts in the center, and external receivers on the right. This visualization facilitates recognition of intrusions such as port scans or distributed denial-of-service attacks.

While such interfaces are easy to read in low-volume, small network situations, they place a large burden on the operator to notice the patterns indicative of intrusions. Even with large or multiple screens, clutter from overlapping connection lines in larger networks can increase to the point where important information needed by the analyst to recognize the patterns indicative of intrusions may be obscured. Our PC3O approach, coupled with the agent annotations described in the next section, helps address these and other of the drawbacks of conventional parallel coordinate graphs.

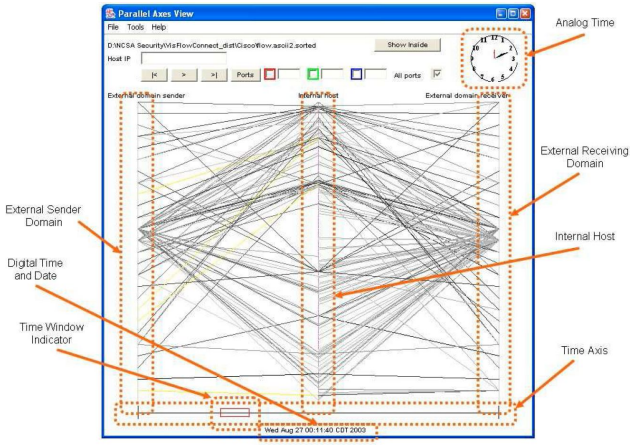

Figure 13. VisFlowConnect parallel coordinates view

Enhanced visual separation of anomalies using custom configurations of multiple planes. The Flow Capacitor can be seen as a base configuration of the Observatory, with two identical planes being shown. PC3O extends this idea by allowing any number of additional planes to be vertically layered between the end planes so they sort the downward path of the flow darts. Because the data are shown in planar form, combinations of features can be displayed in two dimensions (e.g., packet size *vs.* packets per second). In this way, each plane itself contributes to the understanding of the network situation, as well as contributing as a component of the overall PC3O configuration.

At each vertical layer, all the flows may pass through a single plane that visually highlights their individual features. Alternatively, the flows can be routed by Boolean operators into one of multiple planes (e.g., a plane that captures flows within our network vs. a second plane that captures flows outside our network), allowing analysts to distinguish via visual separation the interesting characteristics of the data versus the mundane. By building visual separation into the graphical model, the analyst gains *comparative information* (e.g., proportion of threats going to hosts in the energy sector vs. the financial sector) and *correlative information*, by seeing untagged flows that are behaving similarly to tagged flows. By allowing analysts to construct a custom environment of heterogeneous planes that separate and characterize the flows, the Observatory allows the incremental formulation of a whole series of hypotheses constituting a *line of inquiry*, at the price of some added complexity for the novice user. Useful configurations of PC3O planes (i.e., lines of inquiry) can be archived in the CogLog for future reuse in analogous situations. One could envision whole libraries of such inquiry tools.

Example: Exploring a line of inquiry. As an example of how the Observatory supports a line of inquiry, consider a network analyst who is investigating a series of attacks on

[§] See [54] for a survey of visualization approaches for network situation awareness.

port 20000 to the critical infrastructure of a set of electrical power plants. Wondering whether any attackers were missed in the original report, the analyst widens the search for attackers to include flows using SCADA-related protocols originating from a larger geographical area and using not only port 20000 but also neighboring ports of significance to SCADA systems. The analyst uses the Observatory to define a first plane that plots the use of SCADA protocols on all related ports for the larger geographical region.

Having discovered some previously-unrecognized attackers in this way, the analyst creates a second vertical layer in order to answer the question of whether a particular regional utility company is the sole target of the of the attack, or whether a second utility in the same region is also being threatened. The new layer consists of two planes, one of which captures flows going to portions of the IP space corresponding to one regional utility company and the second of which captures flows going to portions of the IP space used by a second company.

Having discovered that attacks are targeting all power utilities in the region, and not just one particular supplier, the analyst now wants to know who needs to be advised of the situation. The analyst constructs a third layer, consisting of two geographical planes that respectively capture the physical locations of the plants under attack. PC3O enables the analyst to discover that, in the case of the first utility, only the supervisor for a small region needs notification, while in the case of the second utility, multiple regional supervisors need to be advised.

7. Collaboration and Sharing

In our observations of groups of analysts at work, we have noticed a tendency for them to work solo, even when coordination would be easy and beneficial. For example, at Tracer FIRE exercises (see Section 10 below), analysts were asked to work as team competitively to solve a series of problems. It was not unusual for two or more individuals to be working independently of each other, not even making it known to others around the table which aspect of which problem they were currently focused on. Based on anecdotal evidence, we surmise that some of this is due to the selection traits of those entering the field, as well as to the tendency, especially within small organizations, to place such individuals in an isolated role where they may be the sole performer and/or where they are rewarded for individual rather than group performance. We could do very little to address such problems, but our cognitive task analysis gave us some indication that there were other areas where we could provide help.

First, correlation agents such as those described above could help make analysts aware of situations where information sharing could be useful, including the sharing of data about relevant past cases. Second, we could use agents to enable the sharing of richer kinds of information among distributed groups. Third, we could use the KAoS policy framework to define and govern information sharing opportunities in line with organizational imperatives. Fourth, we can consider how shared visualizations, such as those that might appear on large displays at the front of a room housing a NOC, could be used to better purposes.

Making Sense of Teamwork. One way which Sol agents support teamwork is through agent-enabled shared windowing and selection in analyst displays. Our current implementation enables efficient joint control and remote viewing of all or part of a visual perspective while minimizing network loads. Selections of objects within views can also be shared across platforms and exploited across different types of views or in directing agent processing of information. The content being shared can be governed by digital policy, as described in more details below in the discussion of the Live Advisory (Section 8 below). In the future, new kinds of visualizations can straightforwardly reuse these foundational capabilities.

Making Sense of Work Context. Shared visualizations, such as might appear on a large display at the front of a room housing a NOC, have generally suffered from a lack of careful study of what kinds of information might be most useful to display in such a fashion in a given context. During our discussions with practitioners, we have considered various possibilities for the kinds of information that could be included in such displays. Among examples seen as most promisingare: status of progress on individual and group goals and tasks, tasks where help is needed or for which help can be given, availability of others to help (or not), questions that still remain and answers that have been obtained, requests for alternative interpretations, notices of damaging or potentially damaging events, timelines of critical events, and graphical summaries useful for commanders trying to get a quick picture of the current situation or needing to develop a quick status report or alarm.

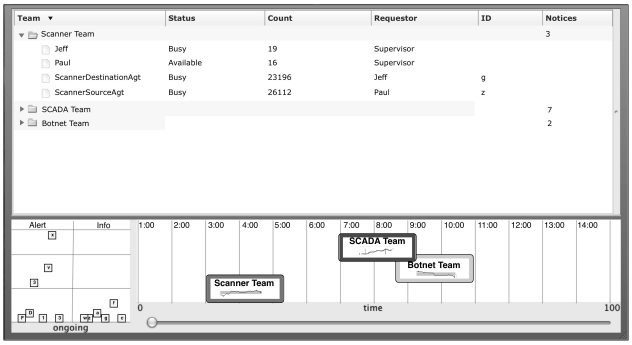

Figure 14. Mock-up of a FishTank display. Note that this hand-drawn image contains some additional graphics (e.g., the sparklines) that were not implemented in the prototype

In support of this objective, we implemented an initial demonstration prototype of what we have called the "Fishtank" (Figure 14). The idea of the FishTank display is to enable continuous progress appraisal [55] by groups of analysts through a visualization that might help them to easily see what tasks and which human and/or agent team members are significantly ahead or behind schedule, and

thus replan their own efforts on interdependent tasks accordingly. The name "FishTank" for the concept comes from the idea of tasks needing attention and team members needing help rising gradually upward on the display according to their urgency, like dead fish floating to the top of a fishbowl.

8. Minimizing Tedious Work

There is much that could be done to help analysts with the burdens of tedious everyday work. We considered that, in general, agents provide an interesting solution both to the problem of sharing of actionable expert knowledge with less-experienced analysts and preserving such knowledge when analysts left or retired. For example, in addition to sharing know-how verbally with close colleagues, analysts could share such knowledge quickly across an entire organization by creating an agent that embodied the task in question and putting it in a library that could be accessed by others.

We also developed a prototype to deal with the burden of generating and sharing warnings and advisories—typically one by chat or phone at present We reasoned that agents could provide rich, active, and actionable information by generating advisories, indications, and warnings in the form of intelligent, dynamic, multimedia components that can be shared remotely. For instance, in order to notify the power plants that are likely next targets of attack, as discussed in the learning example above, the analyst can graphically select the nodes in question and send what we call a "live advisory" in order to notify, and even provide active assistance to, remote colleagues.

Our "live advisory" is an agent that contains not "just the facts" of a situation, but also contains active analytic tools, views, and capabilities useful in ongoing monitoring and response to a threat. In addition, analyst expertise can be embodied actively in the live agents that are sent to colleagues, rather than included passively in "dead" notes and reports. Because the Live Advisory is encapsulated within an agent, every aspect of its actions can be governed by policy—from the decision about whether or not the receiver can accept delivery, to the dynamic determination of which parts of the content of the displays can be viewed by a given recipient, to the determination of whether or not the protective action recommended by the sender can be trusted.

Figure 15 shows an example. Once remote colleagues receive a Live Advisory, they can open it up (if security policy permits) to view the rationale of the sender for sharing this information with them. In addition to the summary text at the top of the display, the analyst may replay past data or connect to live data relevant to their problem through one or more views encapsulated as agents. The tabs labeled "defend" and "respond" (not yet operational) were intended to illustrate the future possibility of using the "Live Advisory" directly to engage in protective actions, thus saving valuable time.

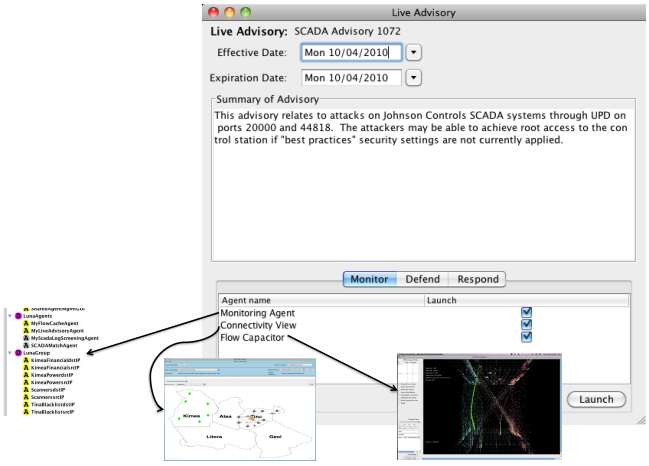

Figure 15. Live Advisory Example

9. Scalability and Resilience

The Sol framework has been architected for dynamic scalability across varying computing and network resources. In addition, the principles we have used to create the underlying agent framework lend themselves well to resilient performance, ensuring mission continuity, even when under attack or experiencing failures. We explain these concepts in more detail below.

9.1. Dynamic Scalability

Agents enhance system scalability in four ways. First, by their capacity to automatically adapt to changes in arrangements among highly distributed, rapidly reconfigurable, service-oriented computing platforms. Policy-based governance of agents allows any change in the state of the world or in the availability or configuration of computing resources to be reflected in changes to agent behavior. Second, multi-agent systems facilitate the augmentation of system capabilities at runtime—thus, extensibility to new kinds of threats is as easy as plugging in a new agent—or adding new behaviors to existing ones. Third, because the overall Sol architecture leverages the inherent distributed computing capabilities of the agent platform, virtually every aspect of system performance can be multiplied in proportion to the amount of distributed computing resources available—from a single standalone host or device to a cloud.

Finally, dynamic reconfiguration of processing among different servers or between clients and servers is made possible by Luna. As briefly mentioned earlier in the article, the Luna platform supports the option of allowing agents to migrate between operating environments and hosts. In principle, this would allow Sol to maintain session continuity when an analyst moves to a different workstation host or even to a portable device. Most mobile agent platforms support only strong mobility, where executable code is moved, or weak mobility, where agents can move while preserving essential aspects of their execution state [56]. In addition to *voluntary weak*

mobility, Luna supports *forced mobility* where, with complete transparency to the agents themselves, agents may be moved from one system to another by an external asynchronous request. Since only the agent execution state is moved, not the agent software itself, the Luna platform is protected from the security vulnerabilities of typical code migration approaches to agent mobility.

We have anticipated the benefits of parallel processing in certain portions of our framework. For example, while the impressive performance of our current version of the Observatory (manipulation of 6-8 million particles in real time) relies on highly-efficient single-chip OpenCL processing [57], the parallel processing enabled by this architectural approach could be fully exploited in the future.

Resilient performance of the macrocognitive system is a key objective of our design. We discuss our future plans to address this issue through polycentric governance below.

9.2. Resilience Through Polycentric Governance

Within the framework of resilient systems engineering, Branlat and Woods have discussed important patterns that lead to failure in complex systems [14]. The focus of this section is how agents might be used to provide support for adaptive performance in the face of stressors and surprise through the principles of polycentric governance [81].

A related notion of organic resilience [58] was inspired by the concept of "organic computing" proposed in Müller-Schloer [59]. Organic resilience relies heavily on biologically-inspired analogues and self-organizing strategies for the management and defense of distributed complex systems. Carvalho, *et al.* have previously applied the concept for the defense of tactical communication systems [58] and mission-critical cloud applications [60]. The concept focuses on the design of emergent coordination mechanisms through local gradients and implicit signaling. Multi-layer defense frameworks following the same principles were later developed for critical infrastructure protection and distributed control systems [61][62][63]. These infrastructures included humans as an integral part of the system, working in collaboration with software agents to improve system resilience. This approach seems well-suited to applications such as the one described in this article.

The use of semantically-rich policies to help achieve polycentric governance builds on our contributions to the DARPA Ultra*Log program. In that effort, IHMC's KAoS Policy Services Framework [31][32] was used in conjunction with software agents to assure the scalability, robustness, and survivability of logistics functionality in the face of information warfare attacks or severely constrained or compromised computing and network resources [64][65]. In a review of alternative policy language approaches presented by the NSA-sponsored Digital Policy Management (DPM) Architecture Group, KAoS was highlighted as the "recommended policy ontology starting point" [66]. Following subsequent collaborative efforts by DPM and IHMC, the KAoS core ontology was adopted as the basis for future standards efforts in DPM [67]. Impressive system performance results have being demonstrated in a simulated environment within the AFRL Tactical Service-Oriented Architecture (SOA) program and also as part of the Army CERDEC JTEN (Joint Tactical Edge Network) program [32]. In addition to the work mentioned above, we have drawn on concepts and an initial implementation of the notion of collective obligation policies developed by van Diggelen, *et al.* [68][69].

Because the latest evolution of these particular aspects of our approach to increasing resilience in Sol is currently the subject of active research and has not yet been fully implemented, we sketch its major elements only briefly.

As with many biological systems, the goal of an approach that relies on polycentric governance is to avoid static and centralized single-point-of-failure solutions for organizing work to the greatest degree practical. Thus, although groups of agents within the system are collectively responsible for jointly executing various tasks, the specific responsibilities assigned to agents are not fully determined in advance. The goal is to allow the agents to self-organize within the constraints of their individual capabilities and current availability. As described in Carvalho, *et al.* [58][60], the premise of such resilience depends on understanding the advantages and disadvantages of particular techniques for self-organization for different problems within a given situation and computing environment.

The use of collective obligations is critical for practical applications of polycentric governance. Whereas an individual obligation is a policy constraint that describes what must be done by a particular individual, collective obligations are used to explicitly represent a given agent's responsibilities within a group to which it belongs, without specifying in advance who must do what. In other words, in a collective obligation, it is the group as a whole that becomes responsible, with individual members of the group sharing the obligation at an abstract level.

The execution and enforcement of collective obligations requires different mechanisms for different contexts. For some applications, a specialized planning system, spanning a group of agents, may be the best approach. However, in this case our commitment to a biologically-inspired approach requires that the agents themselves, rather than some centralized capability, organize the work. In our case, we expect that the agents themselves usually will be in the best position to detect local triggers for collective obligations (e.g., potential threats or opportunities), to determine what support they can offer through their own resources and individual capabilities, and what information should be shared among peers and with agents elsewhere in the system. The self-organizing nature of the system enables the agents to revisit responsibilities and resource allocations themselves, as needed, on an ongoing basis.

Applied in a manner consistent with polycentric governance, we believe that policy-based collective obligations provide the regulatory mechanisms to enable effective and coactive coordination algorithms for agents. Moreover, we envision the implementation of policy-learning mechanisms that could rapidly propagate lessons learned about productive and unproductive actions to whole classes of actors.

10. Performance Studies

We have not yet been provided with the opportunity for robust experimental validation of the framework, but we have relied on literature reviews, observations, and, whenever possible, direct feedback from analysts representing several organizations to guide the work. In large measure, we attribute the enthusiastic feedback about the potential of the framework in our interactions with practitioners to the design principles whose foundations lie in research in the cognitive and social sciences. In this section, we describe recent results from empirical studies that form a foundation for our efforts to formally evaluate the effectiveness of the Sol framework in the future, as well as to investigate variability in individuals and teams.

10.1. Accommodating Variability in Individual Processes

A common simplifying assumption in the development of software tools and organizational and team processes is to assume that individuals and teams are similar, and that there is a single best solution that may be employed across organizations, teams, and individuals. More realistically, however, it has been repeatedly demonstrated that people will explore a range of strategies and depending on operational constraints and individual attributes, will gravitate to the strategy that has the greatest perceived utility. Consequently, designers must accommodate this variability by avoiding design features that unnecessarily constrain users.

Individual differences in cognitive and psychological characteristics and aptitudes have long been a topic of research interests and numerous attributes have been identified and associated measures developed to assess specific individuals. Likewise, differences in team and organizational processes have been studied extensively, and it is recognized that there are basic differences that impact operations and the effectiveness with which teams function within different contexts [80].

At an individual level, it would be valuable to be able to anticipate the strategy someone would select when presented with a task offering clear strategy alternatives, based on an understanding of relevant cognitive attributes. This hypothesis was examined in a series of studies conducted jointly by Sandia National Laboratories, the University of Notre Dame and the University of Memphis [70]. In these studies, individuals completed an extensive battery of tests to characterize their cognitive aptitudes (e.g., working memory capacity, spatial reasoning, analogical reasoning) When presented various tasks that could be effectively performed using multiple strategies (e.g., line tracing, binary decision, NASA Multi-Attribute Test Battery), participants explored alternative strategies and generally settled on a preferred strategy. Unfortunately, there was very little success in linking cognitive attributes to selected strategies. However, cognitive attributes as measured by the Random Associates Test (RAT) did prove to be effective predictors of the extent to which individuals would explore alternative strategies and the OSPAN (a measure of working memory capacity) was correlated with the propensity of individuals to switch strategies over a series of trials.

These findings regarding strategy switching imply that placed in an operational work environment, individuals differ in their tendency to explore alternative strategies for accomplishing task objectives. While this may often be attributed to increasing knowledge and skill, with some individuals, there appears to be a restlessness that occurs after doing the same thing the same way for some period of time. Consequently, designers must recognize that within some portion of the user population, there is going to be a basic tendency to experiment, using tools in different ways—some of which will surely not have been anticipated by designers.

It seems safe to extend these conclusions beyond the behavior of individuals to also describe team cognition. It is conjectured that different teams will demonstrate a differential tendency to explore different strategies over time. This premise has recently been observed within teams of cyber analysts participating in red versus blue team exercises.

10.2. Tracer FIRE Studies

The Tracer FIRE (Forensic Incident Response Exercise) is a government-coordinated event in which cyber analysts from various government agencies participate as teams in a red versus blue exercise. Observations were made by Forsythe and Bradshaw at an event that occurred in February 2012. This event involved ten teams, each composed of five to eight individuals. Teams were provided a simulated enterprise network and presented various challenges that required they defend their network against various attacks, including detecting and reverse engineering malware infecting their network. The exercise extended over three days and teams were awarded points based on their success in responding to various challenges.

Within the context of the Tracer FIRE exercise, the tools and competition placed minimal constraints on the team processes adopted by a given team or the strategies pursued with respect to the competition. This proved advantageous because it allowed the teams to not only

exercise what they knew about cyber defense analysis, but to also gain experience working on a team with different individuals, who possessed different skills and levels of experience.

During days one and two of the event, each team was interviewed by Forsythe regarding their strategy toward the game, as well as their team processes. Of the ten teams, there were nine unique strategies observed in the competition. These strategies involved various approaches to prioritizing the challenges, as well as approaches for gaming the competition (e.g., scare off other teams by making a concerted effort to be the first team to get points on a given challenge following its introduction).

Additionally, each of the teams described a somewhat unique team organization and team processes. For instance, some teams would discuss activities and divide up the tasks based on who had the best experience. Other teams had two or more experienced individuals who took the lead with the remaining team members focused on supporting them. Still other teams divided into subteams with groups of two or three individuals working together. Of particular interest, most of the teams noted that they had switched strategies one or more times over the course of the event. Often, strategy switches reflected their having gained a better understanding of the competition, but in other cases the switching corresponded to the teams actively seeking a more effective division of labor and approach to the competition.

Within operational settings, one may assume that both individuals and teams are going to similarly explore alternative strategies and developers of tools and processes implemented within these settings must be conscious of this diversity and enable teams to be maximally effective given the strategies and team processes that they select.

11. Future Work

Below we outline a few areas for future work: coactivity in adversarial situations, system-level policy enforcement, and formal evaluations of the effectiveness of Sol.

Coactivity in adversarial situations. Of course, not only participants in cyber defense, but also their adversaries are engaged in a coactive process involving mutual intentional adaptation to peers and to foes [12]. In contrast to peer-oriented adaptations, the intent of adaptations to foes is to disrupt any activity thought to be useful to one's adversary. As an example, "Moving Target Defense" (MTD) strategies allow networked computers to change their structure and configuration dynamically while maintaining their functionality and availability to legitimate users [71]. The goal of these constant changes is to present attackers with an uncertain and unpredictable target. If the target changes quickly enough, it will be too difficult for attackers to succeed in their malicious intent.

While encouraging results have been realized for some of the proof-of-concept implementations of the proposed MTD concepts, there are still questions regarding their applicability and practical use. There are important interdependencies between individual defense tools and the functionality of critical applications and services. Furthermore, different operational contexts are likely to require different configuration requirements for individual defense tools or groups of tools. This is especially important when taking into account the adaptation (or co-evolution) of the adversary. Thus, it is important to start addressing the coordination, or the command and control aspects of moving target defense tools.

We believe that a MT defense infrastructure must be able to combine, manage and optimize the use of multiple moving target defenses, under different operational conditional and mission requirements. We also recognize that effective coordination mechanism for these complex environments must account for both the high-level understanding and framing on operational settings, as well as the low level distributed monitoring and control enabled by intelligent software components. To this end, we are currently working on a human-agent Teamwork approach for MTD Command and Control [72].

System-level policy enforcement. In the future, KAoS will also integrate with VIA to provide a means of policy enforcement outside the Luna host enironment.

VIA [73][74] is a next generation cross-layer communications substrate for tactical networks and information systems. Operating below the network layer, VIA enables applications to adapt and leverage the characteristics of the dynamic communication environment and enables the underlying communications infrastructure to better support application Quality of Service requirements and constraints. VIA enables the full control of all network communication between processes and application within, and across machines. We intend to integrate KAoS with VIA in order to allow fine-grained enforcement of policy down to system-level operations such as the opening of a socket. Operating at lower levels in the communications stack, VIA would provide full visibility and control of all network interactions to the policy services, without requiring any collaboration with or changes to applications.

Formal evaluations of the effectiveness of Sol. We are currently planning controlled experimentation studies that will allow us to formally evaluate the effectiveness of the Sol framework in the context of exercises such as Tracer FIRE. In addition to evaluating the tools themselves, we aim to understand how to support individual and team diversity, to support collaborative processes, and to accelerate individual and team learning. Moreover, in contrast to the many standalone models of cyber defense or attack processes, a joint model based on simultaneous investigation of both processes is sorely needed. The questions and results discussed by Branlat and his colleagues [12] point the way forward from previous studies that typically focused on technological dimensions of the domain and associated knowledge and skills to future studies incorporating human-centered research to uncover and address the difficulties experienced by network defenders.

References

[1] Feltovich, P.J., Spiro, R.J., and Coulson, R.L. (1993). "Learning, teaching, and testing for complex conceptual understanding." In N. Frederiksen, R.J. Mislevy, and I.I. Bejar (Eds.), *Test Theory for a New Generation of Tests* (pp. 181-217). Hillsdale, NJ: Lawrence Erlbaum.

[2] Clancey, W.J. "Observation of work practices in natural settings." In *The Cambridge Handbook of Expertise and Expert Performance,* edited by K.A. Ericsson, N. Charness, P.J. Feltovich and R.R. Hoffman, 127-45. Cambridge, England: Cambridge University Press, 2006.

[3] Meyer, M.A., J.M. Booker, and J.M. Bradshaw. "A flexible six-step program for defining and handling bias in knowledge acquisition." In *Current Trends in Knowledge Acquisition,* edited by B. Wielinga, J.H. Boose, B.R. Gaines, G. Schreiber and M. Van Someren. Amsterdam: IOS Press, 1990.

[4] Bradshaw, J.M., M. Carvalho, L. Bunch, T. Eskridge, P.J. Feltovich, M. Johnson, and D. Kidwell. Sol: An Agent-Based Framework for Cyber Situation Awareness. *Künstliche Intelligenz*: Volume 26, Issue 2 (2012), pp. 127-140.

[5] Klein, G., Moon, B., and Hoffman, R.R. (2006, July/August). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, pp. 70-73.

[6] Klein, G., Moon, B., and Hoffman, R.R. (2006, November/December). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, pp. 88-92.

[7] Johnson, Matthew, J.M. Bradshaw, Paul J. Feltovich, Catholijn Jonker, Birna van Riemsdijk, and Maarten Sierhuis. Autonomy and Interdependence in Human-Agent-Robot Teams. *IEEE Intelligent Systems*, March/April 2012 (vol. 27 iss. 2), pp. 43-51.

[8] Johnson, Matthew, J.M. Bradshaw, Paul J. Feltovich, Robert R. Hoffman, Catholijn Jonker, Birna van Riemsdijk, and Maarten Sierhuis. Beyond Cooperative Robotics: The Central Role of Interdependence in Coactive Design. *IEEE Intelligent Systems*, May/June 2011 (vol. 26 iss. 3), pp. 81-88.

[9] Feltovich, P.J., J.M. Bradshaw, W.J. Clancey, and M. Johnson. "We regulate to coordinate: Limits to human and machine joint activity." *Proceedings of ESAW 2006*, Dublin, Ireland, 6-8 September, 2006.

[10] Feltovich, P.J., J.M. Bradshaw, R. Jeffers, N. Suri, and A. Uszok. "Social order and adaptability in animal and human cultures as an analogue for agent communities: Toward a policy-based approach." In *Engineering Societies in the Agents World IV*. LNAI 3071. Lecture Notes in Computer Science, 21-48. Berlin, Germany: Springer-Verlag, 2004.

[11] Moore, D.T. (2011). *Sensemaking: A Structure for an Intelligence Revolution*. Clift Series on the Intelligence Profession. Washington, DC: National Defense Intelligence College.

[12] Branlat, M., Morison, A. and Woods, D.D. (2011). Challenges in managing uncertainty during cyber events: Lessons from the staged-world study of a large-scale adversarial cyber security exercise. *Human Systems Integration Symposium,* Vienna VA, 10-25 to 10-27-2011

[13] Feltovich, P., J. M. Bradshaw, R. Jeffers, N. Suri, and A. Uszok. "Social order and adaptability in animal and human cultures as an analogue for agent communities: Toward a policy-based approach." In *Engineering Societies in the Agents World IV*. LNAI 3071. Lecture Notes in Computer Science, 21-48. Berlin, Germany: Springer-Verlag, 2004.

[14] Woods, D.D., & Branlat, M. (2008). Basic patterns in how complex systems fail. In E. Hollnagel, Jean Paries, D.D. Woods, & J. Wreathall (Eds.), *Resilience Engineering in Practice* (pp. 127-143). Burlington, VT.: Ashgate Publishing.

[15] Klein, G., K.G. Ross, B.M. Moon, D.E. Klein, R.R. Hoffman, and E. Hollnagel. "Macrocognition." *IEEE Intelligent Systems 18*, no. 3 (May-June 2003): 81-85.

[16] Carvalho, M., T.B. Cowin, and N. Suri. "MAST: A mobile agent based security tool." *Proceedings of the Seventh World Multiconference on Systemics, Cybernetics, and Informatics* (SCI 2003), 2003.

[17] Bradshaw, J.M., S. Dutfield, P. Benoit, and J. D. Woolley. "KAoS: Toward an industrial-strength generic agent architecture." In *Software Agents*, edited by J. M. Bradshaw, 375-418. Cambridge, MA: AAAI Press/The MIT Press, 1997.

[18] Suri, N., J.M. Bradshaw, M.R. Breedy, P.T. Groth, G.A. Hill, R. Jeffers, T.R. Mitrovich, B.R. Pouliot, and D.S. Smith. "NOMADS: Toward an environment for strong and safe agent mobility." *Proceedings of Autonomous Agents 2000*, Barcelona, Spain 2000.

[19] Bradshaw, J.M. "An introduction to software agents." In *Software Agents,* edited by J.M. Bradshaw, 3-46. Cambridge, MA: AAAI Press/The MIT Press, 1997.

[20] Kay, A. "User interface: A personal view." In *The Art of Human-Computer Interface Design,* edited by B. Laurel, 191-208. Reading, MA: Addison-Wesley, 1990.

[21] Bradshaw, J.M., V. Dignum, C. Jonker, and M. Sierhuis. "Introduction to Special Issue on Human-Agent-Robot Teamwork," *IEEE Intelligent Systems 27,* no. 2 (March/April 2012), 8-13.

[22] Hoffman, R.R., Bradshaw, J.M., and Ford, K.M. "Introduction." In Hoffman, R.R. (Ed., Au). with Hayes, P., Ford, K.M. & Bradshaw, J.M. (Eds.) (2012) *Collected Essays on Human-Centered Computing.* New York: IEEE Computer Society Press, in press.

[23] Bradshaw, J.M., Feltovich, P.J., and Johnson, M. (2011). Human-Agent Interaction. In G.A. Boy (ed.), *The Handbook of Human-Machine Interaction*. Farnham, Surrey, England: Ashgate, 283-302.

[24] Christofferson, K., and David D. Woods. "How to make automated systems team players." In *Advances in Human Performance and Cognitive Engineering Research, Vol. 2,* edited by E. Salas. JAI Press, Elsevier, 2002.

[25] Carroll, J.M., W.A. Kellogg, and M.B. Rosson. "The task-artifact cycle." In *Designing Interaction: Psychology at the Human-Computer Interface,* edited by J.M. Carroll. New York: Cambridge University Press, 1991, 74-102.

[26] Ford, K.M., J.M. Bradshaw, J.R. Adams-Webber, and N.M. Agnew. "Knowledge acquisition as a constructive modeling activity." In *Knowledge Acquisition as Modeling,* edited by K.M. Ford and J.M. Bradshaw, 9-32. New York: John Wiley, 1993.

[27] Johnson, N.E. "Mediating representations in knowledge elicitation." In *Knowledge Elicitation: Principles, Techniques and Applications,* edited by D. Diaper. New York: John Wiley, 1989.

[28] Holland, J.H. *Emergence: From Chaos to Order*. Reading, MA: Addison-Wesley, 1998.

[29] Langton, C. *Artificial Life: Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, Addison-Wesley, 1989.

[30] Hoffman, R.R., P.J. Feltovich, K.M. Ford, D.D. Woods, G. Klein, and A. Feltovich. "A rose by any other name… would probably be given an acronym." *IEEE Intelligent Systems*, July-August 2002, 72-80.

[31] Uszok, Andrzej, Jeffrey M. Bradshaw, Maggie R. Breedy, Larry Bunch, Paul Feltovich, Matthew Johnson, and Hyuckchul Jung. "New developments in ontology-based policy management: Increasing the practicality and comprehensiveness of KAoS." In *Proceedings of the 2008 IEEE Conference on Policy*. Palisades, NY, 2008.

[32] Uszok, A., Bradshaw, J. M., Lott, J., Johnson, M., Breedy, M., Vignati, M., Whittaker, K., Jakubowski, K., & Bowcock, J. Toward a flexible ontology-based approach for network operations using the KAoS framework. *Proceedings of the 2011 Military Communications Conference (MILCOM 2011)*. New York City, NY: IEEE Press, November 2011, pp. 1108-1114.

[33] Klein, G., P.J. Feltovich, J.M. Bradshaw, and D.D. Woods. "Common ground and coordination in joint activity." In *Organizational Simulation,* edited by W.B. Rouse and K.R. Boff, 139-84. New York City, NY: John Wiley, 2004.

[34] Klein, G., D.D. Woods, J.M. Bradshaw, R.R. Hoffman, and P.J. Feltovich. "Ten challenges for making automation a "team player" in joint human-agent activity." *IEEE Intelligent Systems 19*, no. 6 (November-December 2004): 91-95.

[35] Bunch, L., J.M. Bradshaw, M. Carvalho, T. Eskridge, P. Feltovich, J. Lott and A. Uszok. Human-Agent Teamwork in Cyber Operations: Supporting Co-Evolution of Tasks and Artifacts with Luna. Invited Paper in the *Proceedings of the Tenth German Conference on Multiagent System Technologies (MATES 2012)*, Trier, Germany, 10-12 October 2012. Berlin, Germany: Springer, LNAI 7598, pp. 53-67.

[36] EsperTech. http://esper.codehaus.org/ (accessed 18 July 2012).

[37] Carvalho, M. A distributed reinforcement learning approach to mission survivability in tactical MANETs. In *CSIIRW '09: Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research*, Pages 1-4, New York, NY, USA, 2009.

[38] VanderHorn, N., B. Haan, M. Carvalho, C. Perez. Distributed Policy Learning for the Cognitive Network Management System. In *The 2010 Military Communications Conference - Unclassified Program - Cyber Security and Network Management (MILCOM 2010-CSNM),* San Jose, California, USA, November 2010.

[39] Still, D.L. and Temme, L.A. (2003). OZ: A human-centered computing cockpit display, in *Interservice/Industry Training, Simulation & Education Conference (I/ITSEC),* Orlando, FL.

[40] Still, D.L., Eskridge, T.C. and Temme, L.A. (2004). Interface for non-pilot uav control. In N. J. Cooke (Eds.) *Human Factors of UAVs Workshop*, Mesa, AZ.

[41] Temme, L.A., Still, D.L. and Acromite, M. (2003). OZ: A human-centered computing cockpit display, in *45th Annual Conference of the International Military Testing Association*, (pp. 70-90) Pensacola, FL.

[42] Thibos, L. N., Still, D. L. and Bradley, A. (1996). "Characterization of spatial aliasing and contrast sensitivity in peripheral vision." *Vision Research 36*: 249-258.

[43] Leibowitz, H., & C. L. Shupert (1984)**.** Low luminance and spatial orientation: In *Proceedings of the Tri-Service Aeromedical Research Panel Fall Technical Meeting*, NAMRL Monograph 33, pp. 97-104. Pensacola, FL: Naval Aerospace Medical Research Laboratory.

[44] Leibowitz, H., Shupert, C.L. and Post (1984). The two modes of visual processing: Implications for spatial orientation. In *Peripheral Vision Horizon Display (PVHD),* NASA Conference Publication 2306 (pp. 41-44). Dryden Flight Research Facility, NASA Ames Research Center, Edwards Air Force Base, CA.

[45] Matthews, T. L. (2007). *Designing and Evaluating Glanceable Peripheral Displays*. EECS Department, University of California, Berkeley, Ph.D. Dissertation.

[46] C. Ware, *Information Visualization: Perception for Design*, Second Ed. San Francisco: Morgan Kaufmann, 2004.

[47] G. Caputo, "The role of the background: texture segregation and figure-ground segmentation.," *Vision Research*, vol. 36, no. 18, pp. 2815-26, Sep. 1996.

[48] Lind, M. (1996)**.** Perceiving motion and rigid structure from optic flow: A combined weak-perspective and polar-perspective approach. *Perception and Psychophysics*, 1996, 58, 1085-1102.

[49] Pollick, F.E. (1997)**.** The perception of motion and structure in structure-from-motion: comparisons of affine and Euclidean formulations**.** *Vision Research*, 37, 447-466.

[50] S. Kastner, H.C. Nothdurft, and I.N. Pigarev, "Neuronal correlates of pop-out in cat striate cortex.," *Vision Research*, vol. 37, no. 4, pp. 371-6, Feb. 1997.

[51] C.D. Wickens and H.G. Hollands, *Engineering Psychology and Human Performance*, Third Ed. New Jersey: Prentice-Hall Inc, 2000.

[52] Kaplan, A. *The Conduct of Inquiry*. New York: Harper and Row, 1963.

[53] Yin, X., W. Yurcik, *et al.* (2004). VisFlowConnect: NetFlow visualizations of link relationships for security situational awareness. *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. Washington DC, USA, ACM.

[54] Eskridge, T.C., D. Lecoutre, M. Johnson, and J.M. Bradshaw. "Network situation awareness: A representative study." *Proceedings of the Fourth Workshop on Human-Computer Interaction and Visualization (HCIV 2009),* Kaiserslautern, Germany, 2 March, 2009.

[55] Feltovich, P.J., J.M. Bradshaw, W.J. Clancey, M. Johnson, and L. Bunch. "Progress Appraisal as a Challenging Element of Coordination in Human and Machine Joint Activity." In *Engineering Societies in the Agents World VIII* edited by A. Artikis, G.M.P. O'Hare, K. Stathis and G. Vouros. Vol. Lecture Notes in Computer Science Series, 124-41. Heidelberg, Germany: Springer, 2008.

[56] Cabri, G., L. Leonardi, and F. Zambonelli (2000). "Weak and strong mobility in mobile agent applications." *Second International Conference and Exhibition on The Practical Application of Java*.

[57] Khronos Group, 2011. http://www.khronos.org/opencl/

[58] Carvalho, M., T. Lamkin, C. Perez. Organic Resilience for Tactical Environments. In *5th International ICST Confernece on Bio-Inspired Models of Network, Information, and Computing Systems (Bionetics),* Boston, MA, December 2010.

[59] Müller-Schloer, C. Organic computing " on the feasibility of controlled emergence. In CODES+ISSS '04: *Proceedings of the International Conference on Hardware/Software Codesign and System Synthesis* (Washington, DC, USA, 2004), IEEE Computer Society, pp. 2–5.

[60] Carvalho, M., D. Dasgupta, M. Grimaila, & C. Perez. Mission resilience in cloud computing: A biologically inspired approach. In *Proceedings of the Sixth*

International Conference on Information Warfare and Security (2011).

[61] Byrski, A. and M. Carvalho. Agent-based immunological intrusion detection system for mobile ad-hoc networks. In *Proceedings of the 8th International Conference on Computational Science, Part III* (Berlin, Heidelberg, 2008), ICCS '08, Springer-Verlag, pp. 584–593.

[62] Carvalho, M., M. Rebeschini, J. Horsley, N. Suri, T. Cowin, and M. Breedy. MAST: Intelligent Roaming Guards for Network and Host Security. *Scientia*, 16(2):125-138, December 2005.

[63] Carvalho, M. and C. Perez. An evolutionary multi-agent approach to anomaly an evolutionary multi-agent approach to anomaly detection and cyber defense. In CSIIRW '11: *Proceedings of the 7th Annual Workshop on Cyber Security and Information Intelligence Research* (New York, NY, USA, September 2011), ACM.

[64] Lott, J., J.M. Bradshaw, A. Uszok, and R. Jeffers. "Using KAoS policy and domain services within Cougaar." *Proceedings of the Open Cougaar Conference 2004*, New York City, NY, 20 July, 2004, 89-95.

[65] Loyall, J., M. Gillen, A. Paulos, L. Bunch, M. Carvalho, J. Edmondson, D. Schmidt, A. Martignoni III, A. Sinclair, "Dynamic Policy-Driven Quality of Service in Service-Oriented Information Management Systems". *Journal of Software: Practice and Experience 41,* issue 12, 2011, 1459-1489.

[66] Westerinen, A., Digital Policy Management: Policy Language Overview. Presentation at the DPM Meeting, Jan 19, 2011 / Updated Mar 27, 2011. (restricted access)

[67] Westerinen, A., *et al*., Digital Policy Management Ontology Discussion. Presentation at the DPM Meeting, January 25, 2012. (restricted access)

[68] van Diggelen, J., Bradshaw, J. M., Johnson, M., Uszok, A., and Feltovich, P. Implementing collective obligations in human-agent teams using KAoS policies. *Proceedings of Workshop on Coordination, Organization, Institutions and Norms (COIN), IEEE/ACM Conference on Autonomous Agents and Multi-Agent Systems*, Budapest, Hungary, 12 May 2009.

[69] van Diggelen, J., Johnson, M., Bradshaw, J. M., Neerincx, M., and Grant, T. Policy-based design of human-machine collaboration in manned space missions. *Proceedings of the Third IEEE International Conference on Space Mission Challenges for Information Technology (SMC-IT)*, Pasadena, CA, 19-23 July 2009.

[70] Radvansky, G. A., D'Mello, S. D., Abbott, R., Morgan, B., Fike, K., Tamplin A. K., & Villano, M. Rum Runner: A Model of Strategy Switching. *Cognitive Science*, manuscript under review.

[71] S. Jajodia, A. K. Ghosh, V. Swarup, C. Wang, and X. S. Wang, *Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats*: Springer Publishing Company, Incorporated, 2011.

[72] Carvalho, M., J.M. Bradshaw, L. Bunch, T. Eskridge, P.J. Feltovich, R.R. Hoffman, and D. Kidwell. Command and Control Requirements for Moving Target Defense. *IEEE Intelligent Systems*, May/June 2012 (vol. 27 iss. 3), pp. 79-85.

[73] Carvalho, M., A. Granados, C. Perez, M. Arguedas, R. Winkler, J. Kovach, Steve Choy. A Cross-Layer Communications Susbtrate for Tactical Environments. Patricia McDermott, Laurel Allender (eds.), Chap. 5, *Collaborative Technologies Alliance, Advanced Decisions Architecture*, 2009.

[74] Carvalho, M., A. Granados, K. Usbeck, J. Loyall, M. Gillen, A. Sinclair, & J. P. Hanna. Integrated information and network management for end-to-end Quality of Service. *Proceedings of MILCOM 2011.*

[75] Hoffman, R. R., Hayes, P. J. and Ford, K. M. (Human-Centered Computing: Thinking in and outside the box. *IEEE Intelligent Systems*, September-October 2001, pp. 76-78.

[76] Hoffman, R. R. and Elm, W. C. HCC implications for the procurement process. *IEEE Intelligent Systems*, January/February) 2006. pp. 74-81.

[77] Hoffman, R.R., Lee, J.D., Woods, D.D., Shadbolt, N., Miller, J. and Bradshaw, J.M. The dynamics of trust in cyberdomains. *IEEE Intelligent Systems*, November/December 2009, pp. 5-11.

[78] Ballas, J. A. (2007). Human-centered computing for tactical weather forecasting: An example of the "Moving Target Rule." In R. R. Hoffman (Ed.), *Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Making* (pp. 317-326). Mahwah, NJ: Erlbaum.

[79] Cisco Systems (2007). Netflow Services Solution Guide. http://www.cisco.com/en/US/docs/ios/solutions_docs/netflow/nfwhite.pdf (accessed 18 September 2012).

[80] Salas, E., S. M. Fiore, and M. P. Letsky. *Theories of Team Cognition: Cross-Disciplinary Perspectives.* New York City, NY: Routledge Academic, 2011.

[81] Ostrom, E. Polycentric systems as one approach for solving collective-action problems. Social Science Research Network, SSRN-id130469, 2008. http://ssrn.com/abstract=1304697 (accessed 18 September 2012).

[82] Woods, D.D. (1995a). Towards a Theoretical Base for Represenation Design in the Computer Medium: Ecological Perception and Aiding Human Cognition. In J. Flach, P. Hancock, J. Caird, and K. Vicente, editors, An Ecological Approach to Human Machine Systems I: A Global Perspective, Erlbaum, 1995.

[83] Woods, D. D. (1995b). The alarm problem and directed attention in dynamic fault management. Ergonomics, 38(11), 2371-2393.