Adapting Simulation Environments for Emergency Response Planning and Training

Bruce Donald Campbell

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2010

Program Authorized to Offer Degree: Industrial and Systems Engineering

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Bruce Donald Campbell

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

____
Thomas Furness III

Reading Committee:

____
Thomas Furness III

____
Miles G. Logsdon

____
Zelda Zabinsky

Date:______________________________

In presenting this dissertation in partial fulfillment of the requirements for the doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with “fair use” as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, 1-800-521-0600, to whom the author has granted “the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform.”

Signature ______________________

Date __________________________

University of Washington

Abstract

Adapting Simulation Environments for Emergency Response Planning and Training

By Bruce Donald Campbell

Chair of Supervisory Committee:

Associate Professor Thomas Furness III
Department of Industrial and Systems Engineering

Communities are preparing diligently for potential community-wide crises arising from natural and man-made causes. *First responders* are those people who train to fulfill emergency response roles on behalf of community residents, seeking to limit loss of life, protect property, and reduce the cost of long-term recovery periods associated with crisis scenarios. The cost of providing physical drills to train for participation in community-wide crises is exorbitant and the 24/7 demands for first responders can preclude participating in training even if a physical drill is made available. *Simulation environments* are computer programs with specialized interfaces that can expose humans to simulated crises in order to gain insight as to how they should respond in an actual crisis situation. *Role-play* allows for a live player to simulate the performance of activities independently as well as with other agents, all coordinated with simulation software to provide feedback as to their performance. The emergent field of *serious games* has attracted researchers who want to contribute to a distributed process of improving the experience and increasing the usefulness of such simulation environments.

This research develops and tests a software architecture named *RimSim* as a serious game for emergency response planning and training. The software design facilitates manipulation of various design issues such as the human interface and representational constructs for rapid assimilation and decision-making. Various implementations and testing of the *RimSim* within hospital evacuation teams for a specific community-wide hospital evacuation scenario demonstrates that the approach is viable and useful for further development and implementation. Appropriate metrics to evaluate the success of

emergency response team players comes from a wide variety of fields including distributed cognition, distributed intelligence, situation awareness, and insight generation, each of which is described and integrated into the evaluation of subject experiments. In this research, metrics are calculated and discussed in terms of applicability and relevance to future work.

ACKNOWLEDGMENTS

For nine years, I have been supported and encouraged by a fascinating group of people. Besides acknowledging my committee members, acknowledging others individually is one of the more enjoyable tasks of the doctoral endeavor.

Tom Furness III is a magnet for attracting compassionate, forward thinking individuals with expansive imaginations who are willing to collaborate generously. I felt his pull from across the whole North American continent and he brought me into his circle of Human Interface Technology Laboratory researchers where I first gained an enlightened point of view for considering the relationship of technology to human capacity and ability. Thank you Tom for persuading me away from pursuing the computer science degree I came to Seattle expecting to pursue.

In the HIT Lab, Suzanne Weghorst and Toni Emerson gave hours of their time to help me get up to speed from day one. Every researcher I worked with during the Seattle years and my six months in Christchurch, New Zealand made a contribution to my professional development and perspective. Konrad Schroder helped me navigate the vast range of systems and networking protocols we had available to us thanks to many generous funders and consortium organizations I worked with and for. Konrad also helped me improve my code and algorithms and offered up many code modules of his own for incorporation into the work contained in the pages that follow.

Tom's persistence finally paid off in getting me to start this PhD journey, but I doubt I would have agreed to dive in or made it through the gauntlet without the examples set by other friends pursuing the advanced degree before me. Nick Hedley convinced me how creative and joyful the process could be. Amlan Mukherjee and Chris Weaver demonstrated that there was value and honor in perfecting the role of an academic. Mark Billinghurst assured me the journey would take me many interesting places and introduce me to many interesting people. Richard May proved to me that the skills refined through the process could be used to help others.

My sister, Carolyn Campbell, showed me how it was possible to do many other things simultaneously by effectively leveraging doctoral work for multiple life goals. Ruth Fruland's kindness, as a PhD student *and* landlord who provided morning oatmeal, went a long way to getting me to finishing the document — as did all the other tenants that passed through during the three years I came and went as I pleased.

As Tom worked to bring our virtual and augmented reality work culture to new HIT Lab collaborators around the world, I found myself being adopted by the scientific visualization, information visualization, and visual analytics communities during what will likely turn out to be a heyday for pioneering those fields, as the technological support and emerging perception and cognitive theories made great progress to drive fields dependent on their intersection. I thank everyone who participated in scholarly discussions on plane, train, and road trips to Purdue, U Indiana, Stanford, Penn State, Michigan Tech, U Utah, U Nebraska, U Oklahoma, UTEP, Harvard, Wayne State, Cal Davis, U Otago, ICU, URI, and U Canterbury — as there is no better feeling than being part of a community making significant progress in understanding our world in order to make software-based tools to explore and improve it. I thank every author of every publication I read and virtual world I entered.

I thank the Seattle coast guard, Seattle police department, UW police, and UW medical center for providing me access to their world and Tamlyn Thomas especially for providing such direct doctoral study support. I thank Joanne Perez and Carol Lucey for their contribution as thoughtful hospital logicians.

Last and most, I want to thank my wife Irma Campbell, and my parents Ian and Shirley Campbell, for their unconditional love before, during, and after this graduate school epoch. I derive such pleasure through sharing this life with them, helping when I can, and always receiving back more than I give, which makes giving feel important and worthwhile in any life journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Number Page

Chapter 1 – Introduction

There exists widespread concern that community emergency response systems are inadequately prepared to respond to large-scale catastrophes of either man-made or natural origin. Both the 9-11 terrorist attack in New York City and Washington, DC and the Katrina hurricane event in the northern Gulf of Mexico region raised our concern about preparedness rather than diminishing it. While there is agreement that an optimal emergency response effort to community-wide catastrophic events provides an opportunity to save lives and property while mitigating short-term response costs and long-term recovery costs [1], such an effort requires the coordination of a complex system of people, materials, and supplies that cannot be expected to respond optimally on the first try.

Steven Bailey, a typical Director of Emergency Management Department for a community of a half-million semi-urban residents in Pierce County, Washington, warns that the general public is still generally unaware of the large expectation gap between amount of services available and amount of service required for a community-wide crisis response to reach and aid effected parties [2]. The severe windstorms that paralyzed Pierce County in November 2006 emphasize the point: 220,000 homes lost power in Bailey's jurisdiction and an all out effort by available emergency response workers still left 5,000 residents without power ten days after the event. Public outrage aimed at the delay in restoration of power appears unfair if a traditional time and motion study of the response effort is visualized. Why hasn't a visualization of resources mapped upon resource needs been widely spread in order to educate communities about realistic expectations? Technically, of course, visualizations of response efforts could be placed on the Web and viewed by those who were affected. But is the requisite visual literacy truly available in our communities to process such content?

Like most American mid-size and large urban counties in the United States, Pierce County is expected to build an organizational structure in anticipation of emergency response efforts through the guidance of the National Incident Management Structure (NIMS) handbook [3]. This handbook provides advice on how to organize people into planning, operations, logistics, and finance teams that scale up based on the size of the event. As defined in the handbook, an incident commander takes on the primary responsibility for the coordination of emergency response activities and tasks attempted. This model aligns well with a military model that the United States has used extensively in the past [4]. The incident commander makes decisions based on a situation awareness, whether implicitly or explicitly sought out, which influences how he or she proceeds throughout the response.

As Mica Endsley has investigated in years of highly cited research, the *situation awareness* needed for supporting decision-making in a complex and time-critical environment is difficult to obtain and, even if gained, may not suggest the proper course of action to pursue [5]. Situation awareness, defined as the perception of environmental elements within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future, can be improved through the collection and integration of each piece of data that validly sheds light on what is going on, where it is going on, who is involved, when it starts and ends, and what are its causes. Every person, whether a trained emergency responder or member of the general public, can help collect data that can be combined to provide situation awareness - our public 911 telephone emergency service has proven that over time when responding to smaller localized emergencies.

On the whole, situation awareness relies heavily on the distributed perception and cognition of humans located within the geospatial and temporal scope of the crisis event. Situation awareness also requires that humans ascertain how valuable perceived data is to the development of situation awareness and escalate or de-escalate their data reporting as a result. The value of data is highly dependent on whether it has already been reported and verified. Shared visualization can greatly assist with providing insight as to what has been reported and considered in building situation awareness, even if the situation awareness itself is not ready to be exposed to the public. Visualizing both crisis awareness and the response effort are just two examples of how real-time visualizations can be built to enhance *distributed cognition* (d-cog), when defined as the process in which cognitive resources are shared socially in order to extend individual cognitive resources or to accomplish something that an individual agent could not achieve alone.

P.D. Magnus asserts that distributed cognition is the perfect framework for characterizing the process “by which ordinary people do collectively what they could not do alone” [6]. When evaluated in this light, the emergency response effort to the Katrina catastrophic hurricane event appears to demonstrate the overall consequences of suboptimal distributed cognition while also demonstrating the power of strong distributed cognition in sub-tasks associated with the overall effort. Six months after Katrina reached New Orleans, emergency responders, governments, and city residents still publicly disputed each other’s version of what exactly took place in the city during the emergency response effort. Retrospective reviews of the Katrina emergency response are full of could haves and should haves that did not happen because situation awareness was inferior and distributed cognition was not coordinated into a coherent, emergent whole appropriate for supporting necessary decision-making. Accordingly, these retrospective reports lack a comprehensive presentation of the Katrina response effort. Alternatively, simulation technologies may suggest various presentations that would have been useful for gaining a better understanding of the dynamics of the Katrina situation.

The post-event evaluation of Katrina distresses many a citizen who becomes aware of a growing list of potential catastrophes that might occur in their community. Even if they wish to be proactive in helping participate in promoting distributed cognition to help prepare for possible threat scenarios they’ve become aware of, they aren’t sure how to proceed. Unfortunately, the complexity of the whole response is too large to contemplate without being overwhelmed. As a default, society identifies roles and trains individuals to participate in an emergency response effort with a limited, and perhaps unrealistic, set of tasks they can perform. Police officers are trained to keep order and lawfulness at all times. Firefighters are trained to limit property damage and save lives from the threat of fire. Medics are trained to administer medical aid to injured people. The general public recognizes these roles based on uniforms worn, tools held, vehicles driven and behavior protocols portrayed often in our culture. These trained roles have been successful in improving response to emergencies that only require a handful of participants. However, a response event that scales up to requiring hundreds of emergency responders becomes too complex to organize as simply an extension of individual roles. If our trained emergency response professionals are not able to maximize their distributed cognition and help build a useful situation awareness for response decision-making across a broad cadre of emergency responders, how can we expect the untrained public to best participate in their own right? Simulation technologies already are being used to train military, police, and firefighters individually in their tasks [7]. Perhaps we can successfully extend simulation technologies to the full emergency response effort across roles, authority, and responsibilities in order to support distributed cognition and help build useful situation awareness for response decision-making.

Many complex phenomena are regularly studied through software-based simulation providing deeper insights and mediated intellectual discussion. Weather researchers simulate environmental conditions and known physical principles in order to predict future conditions [8]. Supply chain developers simulate the movement of goods through value adding organizations and distribution centers in order to understand how flow can meet demand while minimizing distribution costs [9]. Construction management teams simulate the building of a structure in order to verify their plan works in the physical space available and can be completed within the time and resource constraints promised to a customer [10]. As a result of these successful practices, we contend that simulating emergency response efforts is likely to provide a useful tool for studying appropriate emergency response plans — more useful and cost-effective than any other method currently in use. A properly built emergency response simulator also enables emergency responders to train for their roles on their own asynchronous schedule. And, by properly simplifying and yet representing the complexity of the emergency response effort in an interactive visual simulation, simulation provides a tool for emergency responders and their protected citizens to gain an understanding of the nature of collaborative human effort in response to a wide array of potential catastrophic events.

CHAPTER 2 OMITTED FOR REDUCED DOWNLOAD SIZE (BACKGROUND INFORMATION)

Chapter 3 - Thesis, Objectives, and Hypotheses

Our research is unique in attempting to meld the best practices suggested by theory from all five of the domains in the previous chapter into a single framework for improving training and planning for emergency response – taking some direction from the work that has come before us.

Simulations, if built correctly, provide the opportunity to encode knowledge in software that can be interacted with by a learner to explore a complex phenomenon to build understanding over time. Researchers documenting and evaluating successful learning activities through the constructivist school of thought have consistently demonstrated how students learn with a greater contextualized understanding by experiencing the world directly or indirectly through a virtual simulation of the world [113]. Joseph Novak highly correlated such methods of learning by personal construction to the curriculum apprentices are exposed to in professional training programs outside of classroom learning exercises [114].

An agent-based simulation approach to emergency response modularizes first responder knowledge into a hierarchy of software objects whereby each simulated agent encodes the emergency response information and actions the agent is responsible for performing during a response. Generic unskilled behaviors of all human beings in an emergency response scenario can be inherited as well as the simulator grows in scope. An expert system-like, rule-driven database of all the agents available during an emergency response provides the opportunity for the action of one agent to effect the action of another. Other software modules that represent the state of all objects outside of those encapsulated as agents can trigger agent actions as well. Since many of the actions agents make can be expressed by geospatial movements over time, a geospatial visualization of software agents as a by-product of the simulation can inform a viewer as to the nature of emergency response. The success of an agent-based simulation approach to improving emergency response planning and training is heavily influenced by the appropriate encoding of agent behavior in software modules that reflect realistic behavior when combined in the simulation.

Other simulation frameworks and prototype implementations have been developed that are relevant to emergency response simulation. We aim to build upon the preliminary results obtained by Rojas and Mukherjee through building a Virtual Coach application aimed at improving the construction management role associated with a complex building activity. That work has tested and demonstrated the value of an agent-based approach to simulation that includes probabilistic response of agents and probabilistic inclusion of new environmental injects that require response from agents performing the construction [115]. The aim of our dissertation is to explore a similar hypothesis to Mukherjee's dissertation, whereby we could envision directly substituting the two words *emergency response* for the two words *construction management* in his hypothesis that states:

> *A situational simulation environment can be used as an educational environment for construction management personnel while providing a test bed to collect and analyze information in construction scenarios, thus allowing us to study construction management as a dynamic system, consisting of human and resource interactions.*

Beyond verifying that Mukherjee's construction management results are relevant to the emergency response domain, we hypothesize that an agent-based simulation environment can improve emergency response situation awareness through improving the distributed cognition among emergency response personnel. Our main objectives associated with this work being pursued have overlap with Mukherjee's objectives listed on page six of his dissertation in that they all need to be verified as applicable to the emergency response domain [116]. While construction management decision-making can be evaluated on a day-by-day basis, emergency response decision-making often requires split-second decisions that need to be evaluated in that light. To be more specific in relation to our application area, we state our thesis, objectives, and hypotheses succinctly in the next subsection.

3.1 Thesis

For our doctoral work, we propose the following thesis:

> *Emergency Response Performance is significantly improved by participation in visual distributed training tools that increase capacity for distributed cognition through improved insight generation and situation awareness.*

3.2 Objectives

Because this thesis exists within a broad area of research with subjective metrics, we propose four objectives to be researched while testing the thesis above. We intend to:

- Develop a methodology for encoding emergency response scenarios into one or more environmental modules that capture the significant variables associated with each scenario and allows relevant agent-based rules to be triggered by changes in the state of the environment.
- Develop a methodology for encoding specific emergency response roles into agent modules that capture the essence of that agent's behavior in the real world.
- Identify a realistic interface that allows one or more role players to perform an agent's responsibilities within a running simulation that involves environmental and agent-based modules culminating in a realistic experience from which to learn to consider and improve performance.
- Encode derived metrics into a comprehensive assessment tool that enables a visual analytics process defined by both the *number of insights found per time unit spent* using the tool and *level of situation awareness attained* during tool use.

Testing and validating our thesis through the application of our objectives is the primary result to be documented herein.

3.3 Hypotheses

We propose two hypotheses to be tested in association with the thesis:

- *A multi-user situational simulation environment can be effectively used as a* *training tool for generating insight among emergency response personnel.*

- *A multi-user situational simulation environment can be effectively used as a training tool for improving situation awareness among emergency response personnel.*

3.4 Relevance of Hypotheses to Work Performed To Date

Many of the debacles associated with emergency response have resulted from a lack of organization, inferior allocation of resources, and ineffective deployment of external assets (which of themselves appeared adequate) as the complex system did not work effectively as a coherent whole. Based on our review of the literature, observation of teams working on complex tasks, and our personal experience with performing a role within a team performing a complex activity, we believe the distributed cognition of any team can be improved through exposure to external artifacts and social communications that stimulate thinking about team activities. Ideally, to gain experience as a participant in a team emergency response effort, we would like to consider all aspects of distributed cognition in improving the individual's ability to participate: internal cognition, external cognition, abstract cognition, and social cognition.

Hutchins, Varela, Pea, and other's arguments convince us that the body of research work on cognition has already spent too much emphasis on attempting to determine the nature of internal cognition with neither the ability to agree on most conclusions nor how to improve it. Our discussions with emergency responders has convinced us that most external objects used in emergency response are already given adequate training emphasis and existing methods just need adequate repetition in isolation or among a subset of the team to gain full competence. As a result, we look to methods for improving abstract cognition and social cognition as the most likely place to make a significant impact in improving emergency response team cognition. We turn to simulation as an opportunity to provide emergency responders adequate time and place to practice and improve abstract cognition and social cognition in conjunction with other team member roles. In human history, large-scale emergencies have not occurred often enough for emergency responders to practice solely in the physical world. Due to the cost of personnel time and materials, emergency responders have not come close to finding a cost-effective method to run an additional magnitude's worth of emergency response drills in order to plan and train for contingencies. We do not wish for the rate of community-wide emergencies to rise in the future and so we can look to simulation strategies as a potential cost-effective way to prepare roles for first responders in a wide variety of scenarios.

3.4.1 Enabling Abstract Cognition with Artifacts

Based on our literature review, we have concluded that abstraction is best served for the advanced thinker who can already chunk concepts in memory to the degree that makes the abstraction effective. But, we believe map reading is an important abstraction that all human beings can benefit from and must become better at given the explosion of geospatial visualization tools made available to humanity for applying abstract, visual thinking to large-scale problems. The map abstraction, when maps are generated from aerial and satellite photography, requires one major skill to perfect: the ability to visualize a first-person location on Earth from a bird's eye view above that location. In the case of a large-scale community-wide emergency crisis, it seems like a necessary skill in order to understand where resources, incidents, and people are located throughout the response effort.

Once the map abstraction is familiar to an emergency responder, he or she can mentally, or physically through dynamic visualization, manipulate symbols on the map to express thinking to others and absorb thoughts expressed by others. Given sensors and/or in-the-sky observers like the media or military aircraft, updated representations on the map can enhance both spatial and temporal thinking. At that point, everything on the map has an opportunity to contribute to situation awareness. Through repetition, an emergency responder can train to better perceive those symbols on the map that are most relevant to his or her situation awareness. Through exposure, an emergency responder can begin to comprehend the significance of different temporal and spatial combinations of those symbols. Through practice, an emergency responder can become better at projecting the future state of the crisis and suggest how their behavior should be modified to fit the emerging situation. Even if a physical dynamic visualization is not available to the emergency responder at the time of the crisis, the responder can gain the ability to maintain an abstract dynamic visualization with a simple piece of paper and pencil (or internally should that skill be developed reliably).

The realities of emergency response suggest that an up-to-date dynamic visualization (whether physical or mental) is unlikely to stay concurrent with reality during an emergency response scenario. Social cognition is likely to be very important in informing the changing state of the crisis and discussing contingencies, priorities, and tasks. Roger Pea's evidence that social cognition fills a basic human need in affixing value to cognition seems clear in relation to how we have observed cognition in our ten years as a classroom instructor. We have experienced more concrete examples of knowledge affixing itself in a higher state in our mind during teaching in the classroom than in any other facet of our life. The fact we are sharing information with interested learners in a highly social environment, as we first and foremost attempt to set up with a new class, crystallizes knowledge into abstract models of a higher order on which we can then rely upon to teach better going forward.

3.4.2 Enabling Social Cognition

Hutchins elegantly demonstrates how important social cognition becomes in military operations, aircraft piloting, and large ship navigation. By building a sense of team and a responsibility to others, participants in a joint activity pursue knowledge acquisition and focus on key perceptions when they know another person's role is dependent on that data. The social nature of human beings drives us to consider other people when performing our own tasks – the more we are concerned about letting the team down, the more acutely we cognate in order to avoid that outcome. Hutchins demonstrates how evolved roles in the military, aircraft piloting, and ship navigating have lead to clear social responsibilities in fulfilling those roles, and how those trained upon responsibilities provide a performance buffer for the team as a whole – a buffer where mistakes can be overcome more flexibly and more often.

In the case of a large-scale emergency response effort, social bonds can also limit the response effort as the effort grows to necessitate new social interactions between groups that have no experience nor trust in working with each other. Performing the social pleasantries typical of building trust, under the typical time pressures of an emergency response effort, can be awkward and uncomfortable compared to sticking to social interactions with the usual team a responder knows well. As a result, human nature pushes human beings to place emphasis on communication with known social acquaintances inappropriately over those communications with those who most need to be communicated with for the overall situation awareness to be maintained and to improve the overall emergency response effort.

Social cognition needs to be both planned and trained for as much as abstract cognition. A simulator that provides new communication channels that are realistic of potential communication channels at response time can expand the social thinking of a simulation participant. Exposure to thinking about a large community crisis can suggest new relationships that are important to form before a crisis occurs. Exposure can help a participant better understand the social network within which he does his work, as well as the greater social network associated with the community at large. Having the ability to adjust communications channels to simulate any quality from an ideal channel to a strongly impaired channel affords the opportunity for an emergency responder to practice their communication skills under the wide range of realistic situations he or she may encounter.

3.4.3 Enabling Recognition-primed Decision-making

When considering cognition, we must revisit the evidence provided by Klein and his associates. Their evidence suggests that human beings use recognition-primed decision-making during firefighting efforts and nuclear power plant crises to look for solutions based on prior knowledge of those decisions that have worked well in the past [12]. The built-in ability of humans to find patterns in complex data suggests we have a built-in penchant for matching patterns of complex perceptions to thoughts about the significance of our current state. Fires occur often enough that a senior firefighter has the opportunity to perceive enough fires under a wide-enough variety of conditions in order to connect his pattern-recognition ability to his comprehension of what is going on and to his projection of what to do about it. And, if fires are not occurring regularly, training regimens allow for minimally-controlled fires to be set in order to gain the requisite experience.

We are willing to accept the setting off of bombs, creation of earthquakes, and unleashing of tsunamis on a community in order to realistically understand the patterns of human beings behaving in emergency response efforts under severe stress. A simulator, while training abstract and social cognition, appears to have a side-benefit of providing exposure to patterns of community crisis – even the opportunity to replay crises that have been recorded in history. Following logic based on Klein's findings, we may be able to build better recognition-primed behavior in our emergency responders as well.

3.4.4 Dynamic Visualization for Sense-making

While there is often little time for reflection during an emergency response effort, there is plenty of time for reflection after the fact. The ideal training tool for simulating an emergency response effort is unlikely to be the ideal tool for dissecting performance and considering behavior modification for next time similar conditions arise. Thankfully, there is an emergent field of dynamic visualization that can advise us how to best build a review tool with which to provide an in-depth evaluation of a simulated emergency response session. The literature is full of hundreds of examples of how visual analytics can enlighten a team of individuals through sense-making activities spurred on by an interactive, dynamic, visualization tool. Building an appropriate dynamic visualization tool and demonstrating its worth independent of the simulation tool should shed light on additional tools to improve our abstract cognition when thinking about emergency response in general.

Chapter 4 – Preliminary Development and Testing of a Role Simulator

We believe that our first challenge was to verify that a role simulation tool would show potential for improving a first responder's depth of understanding when considering their individual effort within the emergency response effort as a whole. We concluded the tool's potential was best verified through demonstration of a tool to an emergency response team. We attended two emergency response drills that took place at two Emergency Operation Centers (EOC): The University of Washington (UW) EOC drilling a delivery truck accident scenario that released a chlorine gas plume into the environment, and the UW Hospital EOC drilling a tampered water supply scenario. Both EOCs follow the National Incident Management Structure (NIMS) recommendation for organizational structure during incident response – organizing the effort among four key response teams within the EOC: planning, operations, logistics, and finance/accounting. We watched the drills looking for an acceptable pilot team of personnel with which to present our ideas on role improvement through simulation.

Based on their identifiable tasks and their willingness to work with us, we chose the medical logistics team within the UW Hospital EOC and contacted the emergency response coordinator, Tamlyn Thomas, to coordinate working with the University of Washington hospital emergency response medical logics team. Through interview and observation, we identified five roles the medical logistics team needs to perform during a community-wide crisis response, having watched similar tasks in action during the water contamination scenario played out in the EOC command room in the center of the UW Hospital. Our emergency hospital evacuation scenario includes medical logistics roles comprising of five overlapping tasks:

- Hospital Transportation Coordinator coordinates patient evacuation through close work with the hospital evacuation and floor coordinators.
- Hospital Control coordinator provides an external facilitation role with those at the evacuating and receiving hospitals.
- Fire Department Transport Coordinator coordinates fire department personnel in the removal of patients from the evacuating hospital.

- Receiving Hospital Coordinator One coordinates patient deliveries to hospital one.

- Receiving Hospital Coordinator Two coordinates patient deliveries to hospital two.

We concluded that these five roles all have clear and distinct goals with overlapping data needs that require a shared common operating picture (a key base requirement for requiring distributed cognition). Specifically, data we would need to represent in a role simulator includes:

- Supplies and materials
- Transportation routes and vehicles
- Patients in the evacuating hospital
- Emergency response personnel

We built a training simulator for the first role listed above, the Hospital Transportation Coordinator, who coordinates patient evacuation through close work with the hospital evacuation and floor coordinators and other key roles in the emergency evacuation scenario. In building the training simulator, we implemented the system architecture we had been promoting and refining for a year though interaction with the National Visualization and Analytics Center (NVAC) community. Our resultant architecture, as used in our first simulation runs, is shown in Figure 3.

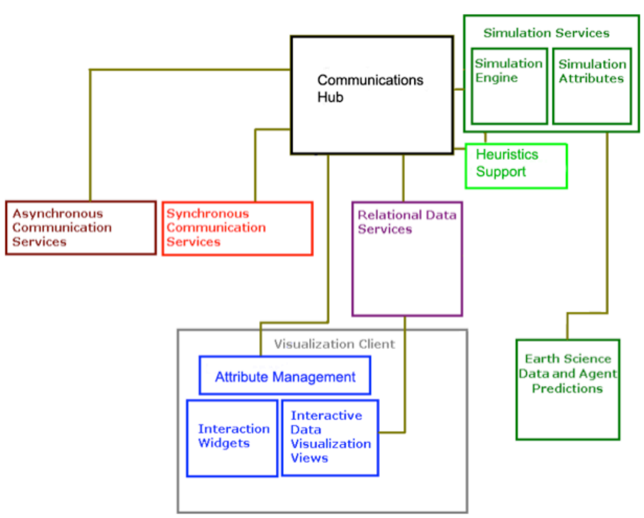


Figure 3 – RimSim computer-mediated simulation support architecture

We introduced the word *RimSim* as a simple name for grouping together software-supported services that could support visual, thoughtful interaction with a simulation of any emergency response crisis scenario anywhere around the Pacific Rim. In this regard, we use the word RimSim to describe any system where we are using the architecture presented in Figure 3 to support planning and training activities of one or more first responders considering an emergency response scenario.

We attended the Visual Analytics Science and Technology conference in Sacramento, California on October 31-November 2, 2007 (VAST 2007), to present the RimSim architecture and a plan for its use in pursuing a research agenda to a Doctoral Colloquium review panel of five well-known visual analytics specialists and the general visualization specialists in the audience who were invited to attend. The RimSim architecture provides a modular software development process whereby any of the software modules identified as diagram boxes in Figure 3 can be improved and iterated upon independently of the others. The interface between any two modules can be negotiated as new characteristics and needs of any particular module require a change in message passing to other modules in a system built using the RimSim architecture as a blueprint. Each of the modules provides well-defined services to the rest of the modules as a coordinated system.

Because many of the emergency threats around the Pacific Rim exist due to natural threats (e.g. earthquake, volcano eruption, tsunami, and wind storm), the RimSim architecture provides a module that can drive simulation attributes based on physical earth science models (e.g. tectonic, hydrological, and weather). These earth science models exist as forcing agents in predictive models around the Pacific Rim, such as exists within the PRISM community in the Puget Sound watershed of King County, Washington in the United States [117].

Manmade simulation services suggest a tighter coupling with the simulation attributes as many of the simulation attributes can effect the simulation whereas simulation attributes are not able to significantly affect the forces of earth science processes. As a result, the RimSim architecture shows simulation services containing closely knit simulation engine and simulation attributes modules. The simulation engine module runs a time step code loop that keeps the simulation coherent over time. The simulation attributes module keeps track of the state of the simulation through value pairs of attributes and values of those attributes. For example, if a truck is deployed to deliver supplies to a location, the truck is tracked through an array of simulation attributes such as truck1_latitude, truck1_longitude, truck1_supply1_allocation, etc. Simulation attributes can be added and removed by the simulation engine, but are only added through the earth science data and agent predictions module.

A simulation scenario developer can extend the heuristics support module to calculate heuristics being used in an emergency response scenario. Heuristics are often documented by subgroups of an emergency response team and are agreed upon as support algorithms for first responder behavior. Potential heuristics that wish to be considered can also be encoded within the heuristics support module and implemented to support agents being run in code.

The communications hub communicates the state of simulation services and the heuristics support to the role players who are using the simulation services to play a scenario. Both asynchronous communication services and synchronous communication services can be encoded in models that match a messaging protocol for influencing the simulation or receiving communications from the simulation as to the state of simulation attributes at any time step.

Asynchronous communication services allow software developers to add popular computer-mediated communications features to the simulation session. For example, a bulletin board can be provided that allows participants to leave messages and check messages during the course of a simulation session. Other potential asynchronous features include e-mail, video posting, presentation services, and spreadsheets.

Synchronous communication services allow software developers to integrate real-time communications links to the simulation. Text chat, instant messaging, video and audio streams are just some of the possible synchronous communication services in which messages to and from the simulation services modules can be incorporated.

To help with coherent message passing between the communications hub and the visualization client that presents the state of the simulation to a participant, relational database services are available to provide a proper data model in which to connect data visualization views. The participant can change these views through the interaction widgets module and the participant can maintain the overall appearance of the views through the attribute management module. The visualization client architecture follows the *Improvise* modular model described previously.

When we presented the architecture to the doctoral colloquium, the panel suggested no significant updates to this presented architecture, and neither did the broader audience during the ten-minute question and answer period. Instead, useful suggestions were all directed towards possible implementation of both the synchronous and asynchronous communication services. The panel's review focus certainly cemented our already anticipated growing emphasis on social cognition as critical when evaluating the effectiveness of group analytical visualization tools.

Although we had already presented at the colloquium before attending the lunchtime awards ceremony, we see considerable alignment between Card's augmented cognition features in Figure 2 and the features of our RimSim architecture in Figure 3. To align the two figures, we suggest synchronous and asynchronous modules emphasize social cognition process in tools created with the architecture; we suggest the attribute management, interaction widgets, and interactive data visualization views provide dynamic visualization services to our tools; and we suggest all other modules assist in augmenting the human's computing capacity through useful machine algorithms (such as simulation services, heuristics computation services, earth science prediction services, and agent behavior prediction services). Adding a communications hub and a relational database, for quick and efficient data sharing among multiple tool users, adds the connectivity Card suggests is so critical to optimally amplify cognition. The correlation of our suggested architecture to a wise and well-traveled researcher's words of wisdom to the entire visualization community present at its annual conference reinforced out belief our architecture was ready for use.

Focusing on a single role let us explore a typical emergency response role within the context of visualizing a community-wide event. Figure 4 shows our interface with red icons representing hospital locations and blue icons representing warehouse locations (where medical supplies are inventoried in King County). We use blue-purple shaded circles to represent current supply levels at each warehouse and yellow shaded circles to identify demand for materials at each hospital. We chose the blue-yellow scheme to avoid common color-blindness troubles for our users. Truck icons show delivery truck locations. Supply trucks are queued at each warehouse location at the start of the simulation. Medical logicians determine routes between hospitals in advance based on driver training and the routes the medical logistics team identifies as known and easily used without likely driver confusion. Tabular output for supply, demand, and route combinations are provided in the right pane of the single role simulator interface shown in Figure 4.

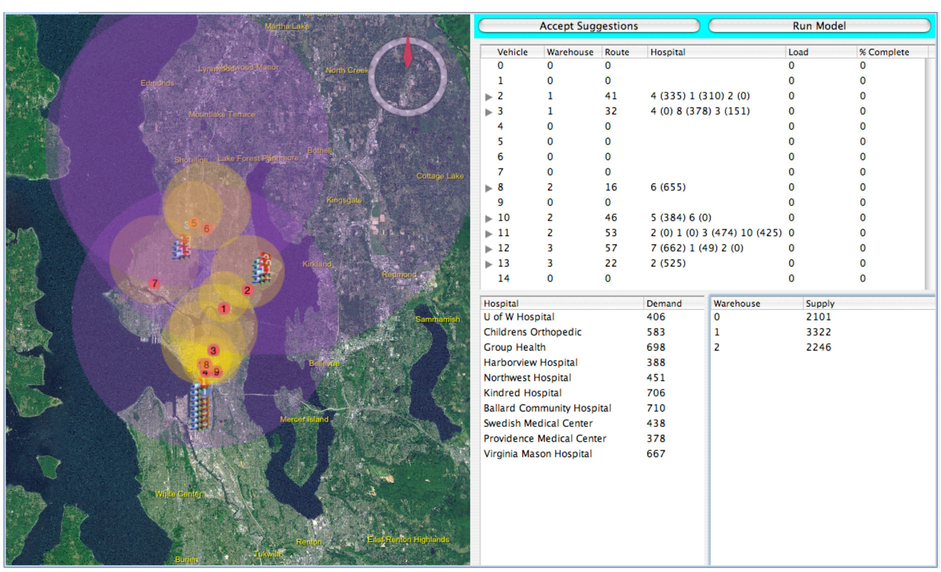


Figure 4 – An allocation of medical resources planning and training tool

Using the tool, routes can be evaluated visually using a mouse rollover response mechanism whereby routes then appear as seen in Figure 5. As a route is selected, the current route duration is presented so the medical logistics role player can consider that route. In all, 72 different routes were available for evaluation and selection during our pilot tests. The medical supply logician role-player loads supplies on a delivery truck according to the role-player’s desired delivery amount by hospital on the route. The truck attempts to make the deliveries as requested, extending or contracting the estimated duration based on real-time conditions within the simulation.

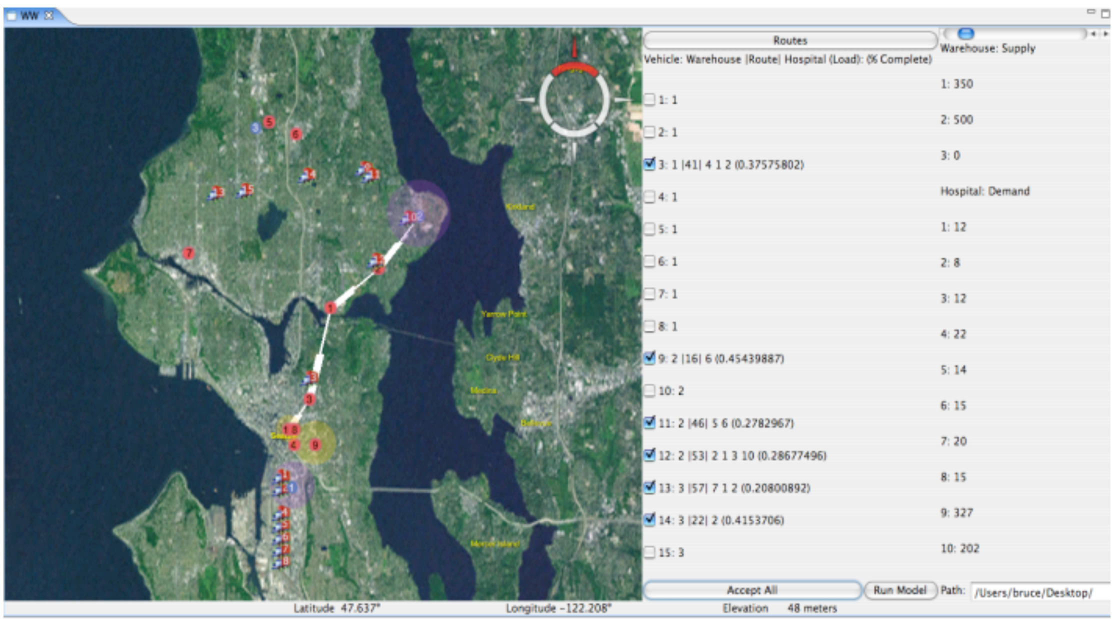


Figure 5 – Evaluation of a supply truck route for potential assignment

We chose 72 routes as representative and sufficient to capture the essence of the allocation role as the team explained it to us. We worked with Onur Mete, a University of Washington graduate student focusing on optimization, to solve the allocation problem using a linear programming model. By solving the optimal allocations mathematically, we could compare the role player’s attempt to an optimized ideal and provide feedback to the player. We could then listen to justifications we might have not thought about and use that feedback to improve our interface to make it more authentic to the success of the task.

Our preliminary results showed that a medical supply allocation role player could use the interface to repeatedly practice their allocation task. Of the six emergency responders we asked to use our tool, their ability to satisfy hospital demand ran at approximately 80% of what the optimization model suggested should be possible. Many times, the user's hesitation to allocate resources fully suggested a concern to us about anticipated future conditions that were not programmed into the optimization model. Because our simulator provided stochastic demand increases and new supply restocking arrivals at warehouse locations, we observed the role player hesitate at various times in releasing a truck when a new supply amount seemed eminent to arrive at a warehouse that would then significantly change her strategy.

Demonstration of our role simulator to the hospital emergency response coordinator gave us hope such training simulators could provide great benefit as both a planning and training tool, but needed to be played in conjunction with other team responsibilities to better represent the cognitive skill needed within an EOC at crisis time. Having gained confidence with work on a specific role, and realizing how many potential different roles we might be able to help train with our tool, we began working on a generic base simulator framework that could be rapidly modified to support a wide range of emergency response roles. We also made note of how time consuming testing a role simulator could be with emergency personnel that don't often have the luxury of providing their attention to researchers during their work shift. As a result we committed to making the simulator run on as many computing platforms as possible.

Chapter 5 – Developing Computer-based Agents to Facilitate Planning and Training

We implemented our base emergency response planning and training framework in software in order to be able to test its merits and iterate its design based on feedback from emergency response personnel. Our software implementation follows the architectural design originally published in *the 2008 IEEE International Conference on Technologies for Homeland Security* proceedings [118]. Using that software, we ran studies of distributed heuristics simulated by agents in a software-based emergency simulation tool we call *RimSim:Response* (RSR) that lets us study the effectiveness of emergency response heuristics while at the same time lets us verify our approach to implementing agents that can simulate any heuristic we want to involve in a distributed cognition emergency response scenario.

5.1 RimSim: Response for Emergency Response Simulation

We ran weekly tests of our software implementation to verify smooth and coherent multi-player use and iterate upon our design for a better player experience. The RSR software lets us:

- Build a scenario anywhere on the planet through a drag-and-drop interface on top of a virtual Earth-based globe.
- Generate multiple roles based on jurisdictions within the geospatial extent of the scenario.
- Apply an agent heuristic and a communication strategy to a role in preparation for a simulation session using that agent.
- Delegate a role to a live player who performs that role within the simulation session – using the graphical interface to aid in her performance.

Various parameters are available to vary the scenario in which the emergency response simulation takes place. Incidents that demand resources in order to administer response services can be set up to trigger at geospatial locations over a specific timeline or time distribution. Resources can be allocated to players with geospatial starting positions.

In the course of our research, the RSR simulator became a test bed for planning and training for emergency response scenarios. Test plans can be run with live players or computer-based agents in either local or remote-over-the-Internet mode. We spent hundreds of hours developing RSR to be flexible for testing a wide range of scenarios. Scenarios can be developed with a scenario developer tool that allows for a visual scenario build on top of the *NASA World Wind* whole Earth drill-down visualization system. Seven scenarios have been built to-date to look at four location-based communities with different characteristics of interest:

- Seattle, WA for a focus on a water barrier environment with unique geographical characteristics.
- Vancouver, BC for a focus on a large center metropolitan island with surrounding suburban communities.
- Christchurch, NZ for a focus on further distributed communities with natural mountainous barriers between.
- Detroit, MI for a focus on an international border for multi-team organization based on nationality.

Three thematic scenarios have been built with differing characteristics in terms of the incidents and resources required to respond effectively:

- Earthquake, with spread out incidents but with many intense incidents occurring within close proximity.
- Tsunami, with incidents skewed closer to water sources than Earthquake and requiring help from an inland jurisdiction.
- Man-made bomb, with a single major epicenter for incidents – requiring help from neighboring jurisdictions to keep up.

Currently, as a ubiquitous tool, emergency professionals can edit these scenarios interactively within the scenario configuration tool. The tool enables its user to iteratively change:

- jurisdiction boundaries between players.
- off-limits areas within the community (such as water and mountainous areas).
- incident locations, quantities, resource demands, and trigger timings.
- starting resource levels and locations.

Since the scenario configuration tool is highly visual and interactive, a demonstration of the tool is warranted in lieu of a long and inefficient written description and is included in this document's accompanying video made available online at *http://bdcampbell.net/thesis_video.html*. Once a scenario has been created, it can be played many times with agents or live players to look for improvements in strategy and then be practiced for plan execution by one or more human players.

Upon attending tens of emergency response and visual analytics conferences, live exercises and business meetings among emergency response personnel, we learned without question how important it is to polish any simulation tool before requesting precious time from emergency response personnel who are burnt out mentally from being provided so many technology support tools for their jobs. To be considerate, we have focused heavily on related emergency response published literature and disclosures made at emergency response meetings, conferences, and live exercises in order to design our simulator. Accordingly, it is essential that we be well organized when requesting emergency response teams to participate in the simulation sessions that will be key tests of the hypotheses of this doctoral thesis.

Consequently, our approach has been to simulate entire participant sessions with heuristic agents to remove any kinks from the emergency response simulation process. We simulate various agent behaviors through our agent code that varies agent behavior in three core facets, the agent's:

- willingness to cross jurisdictional boundaries;
- communication frequency with other agents and EOC personnel;
- response behavior to requests for help from other players or agents.

As a result, some agents are more willing to travel long distances to participate in incident management, others only stay close to home, some agents are highly communicative, others rarely communicate, and some agents are highly responsive to requests for help versus others that are more reluctant.

We have packaged these characteristics into agents to develop profiles of emergency responders that match those of published literature on human behavior. In building them, we have focused on rapid construction so that we can interview first responders and generate new agents based on their strategies for specific scenarios we present to them in table-top exercises.

We believe there is value in presenting the result of our agent-based emergency response runs in their own right. These results perform useful sensitivity analyses of our simulator when looking for a reasonable ability to inform possible behavior changes for better emergency response among teams of responders. Because the simulations are run in code, we are able to interface our simulator to a genetic algorithm to quantify the potential opportunity for improvement among team heuristics. To use a genetic algorithm, we generated a population of encoded strings of heuristic behaviors (called chromosomes or the genotype of the genome in the terminology of a genetic algorithm), which encoded a random set of candidate solutions (called individuals, creatures, or phenotypes) to first responder behavior, and then used chromosome mating and mutation techniques to see how the combination of heuristic behaviors evolved toward better solutions (through the evaluation of those combinations as simulation session runs).

Figure 6 shows the RSR in action for the Seattle-based scenario where agents drive the behavior of all four roles designed in the scenario editor (north, east, south, and west). The quad-colored diamond icons show current outstanding demand for resources at an incident location. The resources that can satisfy the demand are shown as smaller circular icons with type represented by colors that match the respective incident icon quadrants. Yellow lines show the path resources en route are taking to an incident location while white lines show a potential path being evaluated.

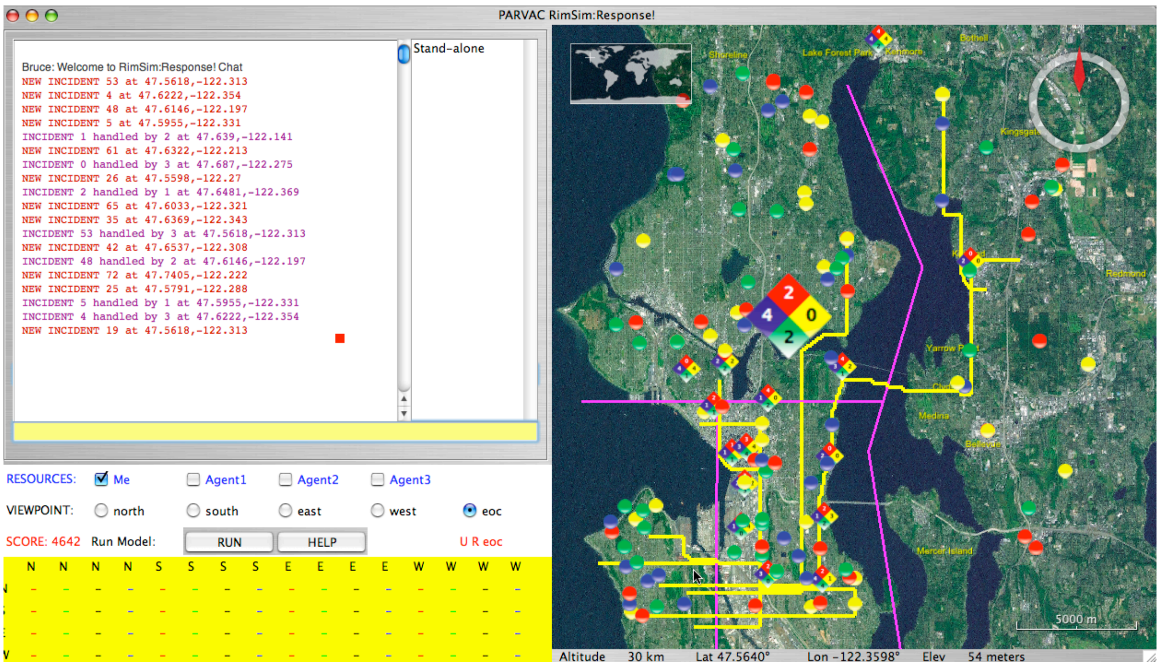


Figure 6 – An RSR Session in Action

A resource reduces incident demand as soon as it is identified for allocation, but can be redirected at any point in time whereby the demand returns to pre-allocation state. Once all required resources reach an incident location, the incident is removed from its visible location and resources are available for redistribution to other incidents. The left-hand pane shows the messaging traffic between players or agents and resource layers can be toggled for visibility. The viewpoint is pre-defined based on the bounding box for each role's home jurisdiction with the EOC role viewpoint providing a full view of all jurisdictions involved in the simulation. The simulation ends when all incidents have been triggered and resolved by the allocation of demanded resources.

Our team ran eighty-nine simulation sessions using both computer-based agents and live players. To play as a live player, a team member used his or her mouse to drag resources from their current location to the incident they wished to resolve. The response within the visualization to all other players appeared identical to the agent-based mode.

Once we felt comfortable that both live players and agents were consistently applying resources as intended, we built facilities to run emergency response scenario sessions automatically via a configuration file. To seed configuration files for multiple sessions, we incorporated and extended the popular Genetic Algorithm Java Implementation Toolkit (GAJIT) to encode each role on a genome that we designed (such that each possible behavioral heuristic the role could implement was represented by a series of 0s and 1s) and to produce a configuration file by interpreting its genes, thus implementing the heuristics encoded in the genes for each role, for each simulation session. Interpreting each genome as a blueprint to seed simulation first responder behavior conditions let us run emergency response scenarios with multiple agent heuristics and optimize group behavior based on any evaluation metric that could be calculated in code.

We performed a simulator sensitivity study by applying the genetic algorithm (GA) to a scenario with four first responder roles, making four different response heuristics available to each role and four different inter-agent communication strategies available to each role as well. With such genomic encoding, each scenario role could thus simulate 16 different responder profiles (4 times 4) and each profile could be expressed zero, one, two, three, or four times in a simulator session.

In the experiment shown in Figure 7, we chose an evaluation metric based on resource effectiveness and responder performance solution time that would evaluate the effectiveness of the emergency response effort throughout the simulation. We set the incidents to occur every seven seconds and an intermediate score to be calculated as a fraction of total resource demand met divided by total resource demand (calculated immediately when the new incident was announced, but not including its new resource demand burden on the session). We then weighted the fractional value evenly by each intermediate score to get a single quantitative evaluation for that emergency response scenario session. The GA was able to significantly improve team success as identified by the plot of quantitative evaluation as seen in Figure 7. Each series of simulation sessions improved over time with time represented on the x-axis from left to right and the level of first response efficiency vertical on the y-axis.

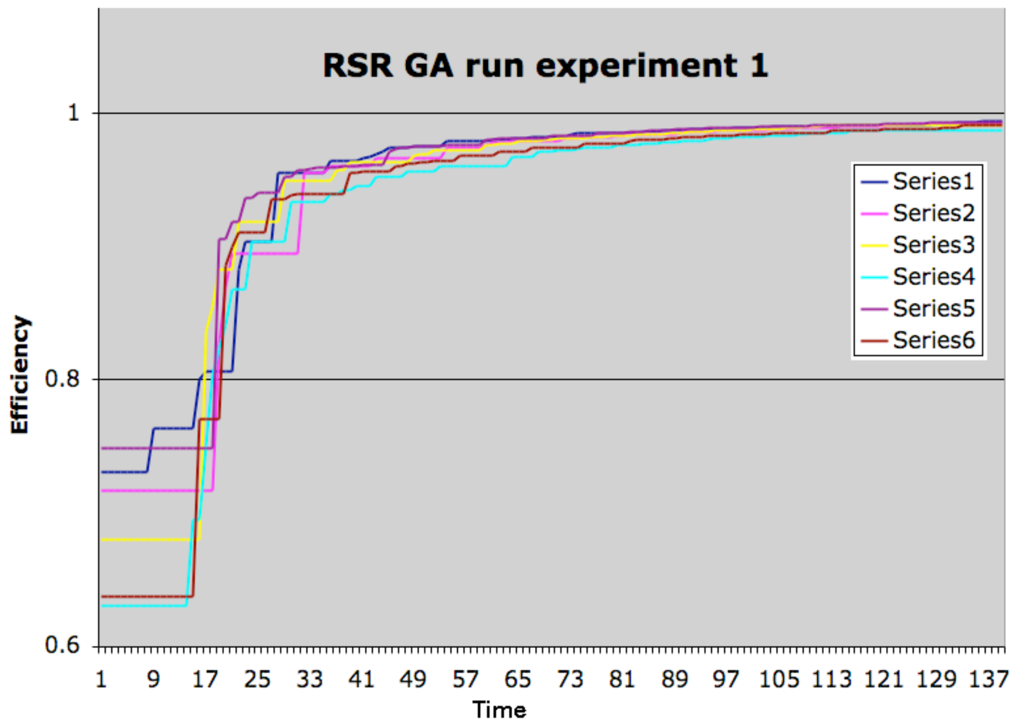


Figure 7 – Representative result of GA-driven RSR runs

Figure 7 shows the improvement in first response evaluation over time of six different simulated chromosomal populations that were mutated (average of one bit per chromosome per population) and cross-bred (average of two cross-over points) over time with a 15% elite rate and 40% cull rate to produce the next generation. Each population has twenty chromosomes and is bred six times to produce seven generations. All 140 resultant chromosomes (20 x 7) are evaluated using the average intermediary value described above. The results for that population are then ordered an assigned a number from lowest score to highest.

In all six starting populations of twenty chromosomes, no procedurally-derived simulation session response effort scored higher than 0.8. In all six ending populations, no procedurally-derived simulation session response effort scored lower than 0.96. Conservatively, that is at least a 20% increase in meeting demand, just by letting the GA attempt to optimize the emergency response effort. Not only that, but all six populations converged on a similar mix of agent profiles for the roles that performed best.

This preliminary result showed promise in suggesting the genetic algorithm as a vigorous way to test out our simulation framework as working properly, while at the same time providing feedback on the merit of allocating different response heuristics in different team combinations for man-made and natural scenarios. The experiment showed how the GA could improve first-responder effort over time through simulation feedback. We began to expect it would do the same for live role-players as it would do for agents or a mix of agents and live players.

In order to verify that these results were not exclusive to our initial generic classes of scenarios, we built our scenario editor with help from Konrad Schroder of the University of Washington Human Interface Technology Laboratory in order to generate thousands of different scenarios through both random and GA-based design. We provided the tool seen in Figure 8 to let anyone handcraft a scenario they wanted to provide for use in the RSR simulator (including those scenarios that were meant to represent known event classes of incidents like earthquake, tsunami, and man-made bomb).

We found a consistent pattern for finding improved team heuristics across all the scenarios we created using our scenario editor and, as a result, felt motivated to add new features to further vary scenario session play in our future work.

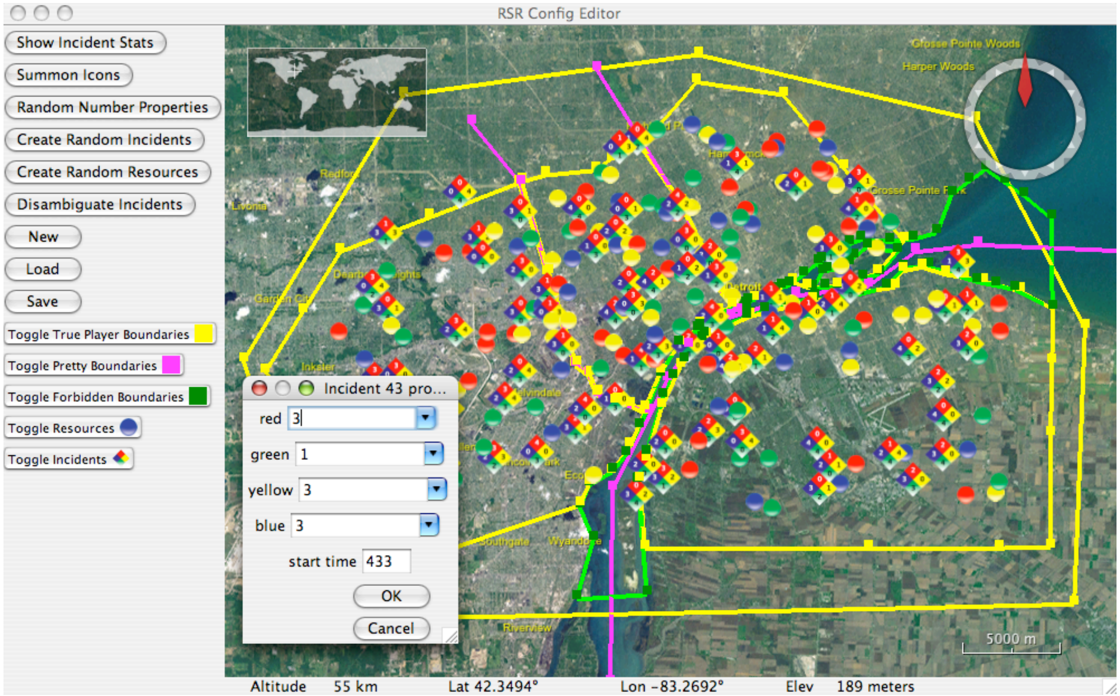


Figure 8 – The RSR configuration editor modifying a Detroit-based Scenario

Upon completing the above range of live player and computer code agent-driven simulation sessions, we learned that our approach could be flexible in allowing a simulation session developer to code an emergency response scenario using the RimSim architecture as a guide. The RSR rendition allows for matching available first responder participation wishes to the simulator: Those who wish to participate can have their roles set up as live participant roles while those who wish not to participate can have their roles simulated in code based on modeling their role as best as possible using expert system techniques.

In our eighty-nine sessions among members of the RSR testing team, we allowed each participant to choose to run live or as an agent within their geographical jurisdiction. The software behaved stably and responsively to allow each participant to reflect upon their role's participation and consider how they might improve their participation next time (either though playing the role in a real-time simulation with others differently or through suggesting changes to be made to the underlying agent behavior algorithms).

Our various runs of 140 scenario sessions to support our genetic algorithm investigation ran without crashing suggesting that the software was ready for intense use if five roles were to be supported in a simulation of similar magnitude to our pilot scenarios. We needed to find a specific scenario that could provide value to a team of emergency responders and also allow us to encode the scenario following the RimSim architecture blueprint. Our pilot period suggested we were ready to use the RSR in focused user subject experiments.

As we aimed to use software to support emergency response team training, and the RSR software was built to focus on enabled real-time simulated role-play, we felt the need to also provide a visual analytics tool that would allow a team of RSR session players to review their simulation session in an interactive querying approach that might suggest reflective, thoughtful analysis. We called such a tool *RimSim: Visualization* (RSV)

5.2 Dynamic Visualization for Sense-making

To develop a tool to help evaluators perform sense-making analyses of emergency response team role-play sessions, we formed a relationship with Chris Weaver at the Penn State GeoVista Center [119]. Weaver maintains an interactive, real-time visual analytics platform he calls *Improvise*. Through his dissertation and various academic papers, Weaver has demonstrated that his toolkit is very powerful for building visual analytic tools for reviewing complex processes and data intensive phenomena in many different fields of study [98]. We spent considerable time working with Weaver to build a tool that could let first responder teams evaluate their performance in an emergency response effort. Our instincts told us that understanding how best to evaluate emergency response scenarios would shed much light on the necessary design of the simulator tool so that the evaluation data could be generated for review sessions.

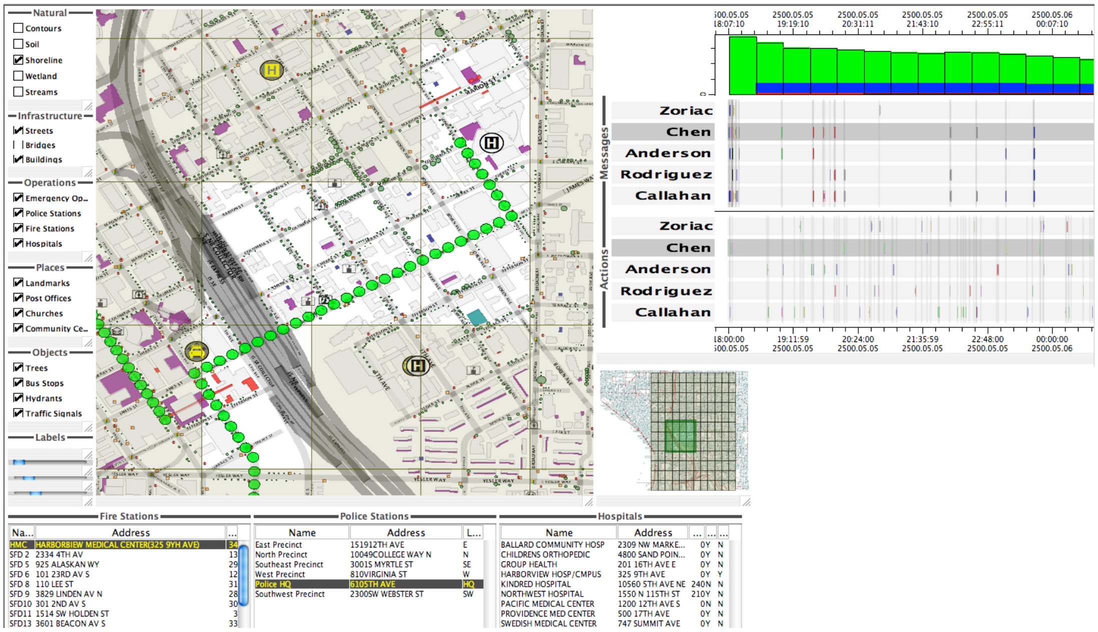


Figure 9 – Improvise Visualization of Resource Movement

A community-wide emergency response effort would suggest both a geospatial and temporal evaluation of key variable states throughout the length of the effort. Starting from that core evaluation tool requirement, we added additional visual query features and integrated them with existing widgets one by one. The result was the two-tabbed visual tool presented in Figures 9 and 10.

Figure 9 shows the physical movement of resources over time and the routes taken. The spatial visualization takes advantage of many King County emergency response data sets. Each visual glyph on the map in the upper-left is drawn on a layer that can be toggled on and off visually. Example layers include the location of hospitals, fire hydrants, bus lanes, police stations, fire stations, other public buildings of interest, streams, lakes, roads, etc. In the middle of the interface is a miniature community coverage map with a green rectangle that can be moved, grown, or shrunk to change the larger city map view interactively.

Tabular lists of key strategic glyphs are hyper-linked to locations on the map and provide direct movement to their location for consideration of response activity within that area. In the visualization's upper right, timelines of all actions made by role players (releasing a resource for allocation to an incident, for example) are shown as tick marks for an overall view of the players' temporal pattern of response. Each tick mark is hyperlinked to the spatial location where that resource was located at the time of the decision in order to quickly analyze other variables at that time and place.

The visual component that appears directly below the action tick marks shows the timing of communiqués made by role players either to other players or computer-based agents involved in the simulation session. These can be correlated with the decisions made for the length of the simulation run by locking the two timelines and scrolling them in lock step.

Figure 10 shows intra-player communications and decisions in a manner that visually exposes relationships between players over time. Again, both messaging and action details can be locked to represent the same time period and both can be scrolled in unison to visually evaluate characteristics of player interactivity over long or short periods.

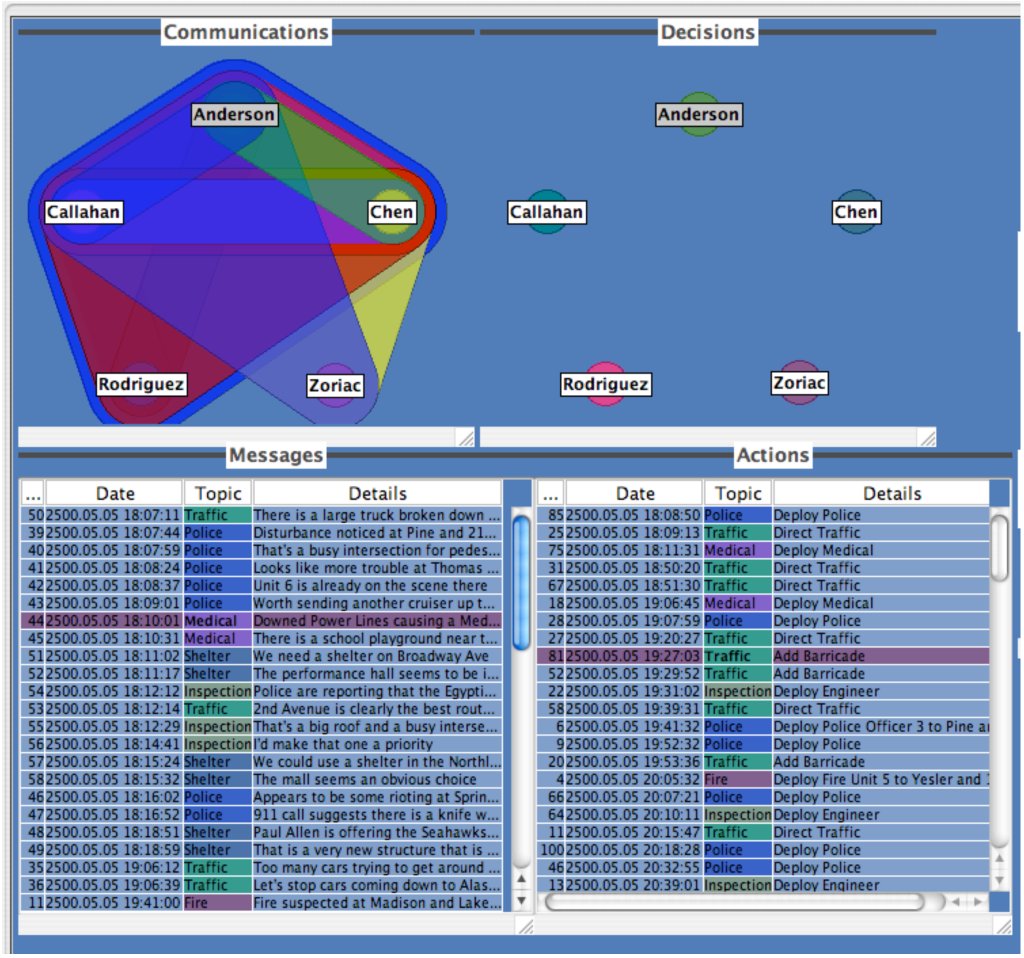


Figure 10 – Second Tab RSR Evaluation Tool Components

By using this RSV prototypical emergency response effort evaluation tool occasionally to review our RSR team play sessions, we were able to gain insight into the nature of emergency response and provide ourselves with a rich inner dialog of thought on insights represented in the data. We demonstrated the tool to both Tamlyn Thomas, the UW Hospital emergency response coordinator and an experienced FEMA emergency response coordinator and addressed their review comments in order to produce better iterations of our visual analytics tool.

To be able to seed data in the RSV visual analytics tool for evaluation, we coded logging statements into the RSR tool in order to generate a text-based file that logs the same key session variables that can be imported and visualized with our RSV *Improvise*-based tool. In all, we spent hundreds of hours investigating and implementing visual analysis configurations for potentially reviewing RSR simulation team efforts. Since the tool is highly interactive and responsive, demonstration of our interactive RSV tool is best done via software demonstration. A text-based description is unable to do the tool justice as it is intended to engage the dynamic visual cortex more than the fixed image or verbal centers of the brain. As a result, we have posted a Web-based, QuickTime-encoded video of our tools online at *http://bdcampbell.net/thesis_video.html*.

Chapter 6 - Lessons Learned from Pilot Tests of Simulation Tools

The most important lesson we learned through a rigorous pilot testing period was that emergency response personnel have demanding jobs that often require 24 hours a day, seven days a week, coverage among multiple people who perform community-critical roles. Because their jobs are so demanding, first responders have very little extra time to waste on co-developing "yet another" tool – including a planning and training simulator. Live exercises coordinated in EOCs are precious and although planned in advance, often lose attendance at the last-minute as the rest of the world requires a scheduled participant's attention. Because first responders work in shifts around the clock, most emergency response exercises, like ours, run in parallel to regular job duties.

A gained respect for the cost of first responder personnel time justified us spending even more time on our base system development without co-development by the first responders we targeted for eventual experimentation. We believe that developing the tool and iterating its evolution among a team of developers who have interviewed first responders and attended emergency response exercises is a reasonable alternative given the realities of the first responder occupation. We may only have one opportunity at gaining the trust of an emergency-response team who, in turn, agrees to participate in a computer simulation-aided test trial. Every representative group with whom we talked complained of being inundated with software solutions that show little respect for the first responder's existing tools, culture, and collaborative process.

By simulating popular first response heuristics in software agents, we have learned that jurisdiction is very important as heuristics that work globally in a community contribute vary differently than heuristics that are applied to local regions only. This finding supports the reality of a jurisdiction approach to emergency response groups traditionally defined as police, fire, and medical.

We have found that communication behaviors affect team response significantly as well. As we often heard from interviews with first responders that communication success is often the most significant variable in an emergency response event. Our software agents show that to be true, as we have learned that team response effort success is sensitive to the messaging buffer level variable for communicating help between players.

The communication model that we had used for our pilot tests included two simplifying attributes that could continue to be adapted in order to better represent the reality of emergency response scenarios. First, the communication model assumed that a message recipient always receives messages sent by another player clearly and with only a minor delay. When watching emergency responders in action, we notice there is a lot of *echoing* where one communicator repeats back what they think they have heard from the initiator of the conversation: This echoing step takes time that our computer-based agents should reflect to be more realistic. Second, some communication messages never make it to the sender's intended destination (especially technology-mediated ones). As a result our communications model could benefit from including a method for degrading communication messages and/or the communication channel quality between responders. We consider communication channel degradation an important consideration for future work. To compensate in our experiments, we decided to use known poor phone connections on the UW campus to connect participants. During the drills we experienced significant static on the lines as expected from line testing prior to the drills taking place.

We provided four resource types for resource allocation tasks in our pilot studies, but differences between resource types consisted solely in the color attribute by which each resource was identified. We used other simplifying features in our pilot studies: All resources took the same amount of time to reach an incident from another location; all resources immediately provided full value upon arrival at the incident; and all resources awaited the end of the incident before being reassigned to other incidents.

More often in real emergency response scenarios, resources have different characteristics that require different behaviors of use. Some resources, like a fire hose, take a long time to extinguish the incident that made the demand. Other resources, like food, expire after a certain period of time when they are no longer effective towards satisfying demand. A police officer in a police car arrives at an incident faster than a medic on foot. A key goal of our simulator was to be able to represent resources faithfully to the scenario our first responder teams wish to use for planning and testing. As a result, our response class needed to have a process by where a subclass could be created for unique resource representations. To prepare for our hospital resource allocation experiments, we generated floor nurse, human patient-assistant, wheelchair, monitor, and ventilator classes as available resources.

In our pilot tests, we used a single incident type, but varied the resource demand each incident requests in order to satisfy the incident's requirements. In reality, incidents have different characteristics that require different behaviors of expression. Some incidents, like a fire burning outdoors in a windy and dry environment, magnify in resource demand as time passes without resolution. Some incidents, like a contained house fire, conclude even if no resources ever arrive to resolve them. Some incidents, like those spawned by a knife-wielding madman, move over time. Our simulator goal is to be able to represent incidents faithfully to the scenario our first responder teams wish to use for planning and testing. As a result, our incident class needed to have a process by where a subclass could be created for unique resource representations.

Through our pilot study period, we were surprised by our inability to predict which combination of first response heuristics would best meet a scenario response effort, no matter what metric we used to determine success. Choosing a specific success metric has been very difficult as we find issues with every success metric we have chosen to date. As a result, we have not found a single metric that can predict the best combination of agent characteristics for multiple agents within a team of agents. We agree with Endsley's suggestion that we use many different evaluation methods [36].

In anticipation of the general exam process by which our thesis, hypothesis, objectives, and work schedule eventually would be appraised, we sponsored a brainstorming session at the UW Human Interface Technology Laboratory to receive feedback regarding what the RSR project team members thought were our most important tasks to accomplish before suggesting that more formal emergency response teams test the emergency response tool with us. We seeded the discussion with the work we believed would be most relevant in pursuing and received feedback as to priorities and additional thoughts. We learned that collaborative project teams that watched a user group in action and then worked together closely for a significant time period (18 months in this case) would all gain a natural consensus of next steps associated with the project. The project team's brainstormed suggestions were nearly identical to our own work plan but together independently for our doctoral experiments.

As we continued to improve the tool for rapid generation of scenarios, heuristic support policies, and inter-agent communication strategies, we made it possible to encode a wider range of scenarios into the tool for emergency response effort study and training role-play.

Because we found that inter-agent communication is a sensitive variable to response effort success, we continued to discuss and consider better communication features between agents to better simulate communications between humans who participate in a simulated emergency response session. We also needed to build a better interface for live simulation role players to communicate with simulated software agents.

Through the pilot testing, we were able to iterate upon our first responder design in order to improve the:

- Ability of our tool to model an emergency response scenario in a manner appropriate for simulating a realistic scenario for first responder training.

- Ability of our tool to provide an interface that faithfully represents the cognitive load of performing the emergency response roles identified by a hospital evacuation scenario.

- Ability of our visual sense-making analytics tool to provide insight on any run simulation session to suggest role-play improvements.

- Ability of our tool to simulate degradation in the communications channel between response agents.

As a result, we decided that our simulation tool was ready for empirical evaluation in more robust experiments. The core of our experimental design reflects the use of our simulator with specific emergency response scenarios suggested by existing teams of emergency response personnel with whom we worked extensively including:

- The University of Washington hospital medical logistics team.
- The University of Washington police department.
- The Seattle area coast guard logistics team at the Joint Harbor Operations Center

To test the hypotheses of this dissertation it was determined to investigate a significant emergency hospital evacuation scenario with at least one of the above emergency response teams. Since we had developed good relations with team leaders, we discussed our goals of experimentation with the hopes of finding willing participants in our experiments. Due to the lack of availability of police and Coast Guard personnel we placed our emphasis on the UW medical logistics team.

Our various pilot studies of the various incarnations of the emergency response simulator were conducted over a period from June 2007 to September 2009. At the end of our pilot period, we identified five major task categories associated with the work needed to finish our software implement. We had to:

- iterate on our simulator code to improve the emergency response scenario session experience and adapt it to a hospital evacuation scenario.

- encode the hospital evacuation scenario we intended to use for our major experiments with domain specialist groups – this required consulting with those who have knowledge in all aspects of such scenarios.

- run informal pilot tests with the UW Medical Logistics team to make sure the software was usable without requiring undue attention to the interface.

- encode our data needs from the hospital evacuation scenario into our data model and encode situation awareness performance metrics from which we can evaluate performance, and then

- iterate on our visual analytics approach to evaluating emergency response scenario sessions in order to run future sessions of the hospital evacuation scenario looking for improved performance.

Although we had implemented a strong framework for running emergency response planning and training sessions, we continued to consider the development of new environmental modules to include specific variables for the scenarios our experimental subjects require. We continued to demonstrate the competent extension of core agent types to support specific roles identified by our target emergency response team: the medical logistics team with whom we had been working. Before we ran each scenario with our target domain expert teams, we continued to run pilot tests with the *RimSim: Response* software to find glaring errors in our interface design and simulation play. After running the scenario often with knowledgeable role players, we were able to test each role module with an agent in order to reach acceptable behavior to the *RimSim: Response* team's satisfaction. We used our agent-driven simulation sessions to visualize agent behavior and share those visualizations with a coordinator of our target domain expert team in order to verify their efficacy.

Chapter 7 – Hospital Evacuation Scenario Development

Integrating the tools developed and distilling the insight gained in the research described previously, we were ready to develop a simulation tool tailored to the needs of hospital emergency first responders. We worked with Tamlyn Thomas of the UW Medical Center and her colleagues at the King County Healthcare Coalition (KCHC) to develop our simulator roles for the specific hospital evacuation scenario they had identified as a critical skills development and assessment scenario. Through various meetings and publications, an example being the overall flowchart in Appendix I, the KCHC convinced us that they strongly believed training for a major hospital evacuation scenario would build critical emergency response competence within the King County community in Washington State. In parallel with our general scenario software simulator development, the coalition reviewed and adjusted scenario roles, responsibilities, and cross-communication details associated with hospital evacuation. While they worked on the human aspects of the scenario, we researched artifact development for computer-based support as a potential training tool that could provide long-term benefit to hospital evacuation preparedness.

As a result of a focused brainstorming session with the RimSim:Response team, we ascertained a strong vision of what needed to be done to modify our software in order to test our two hypotheses:

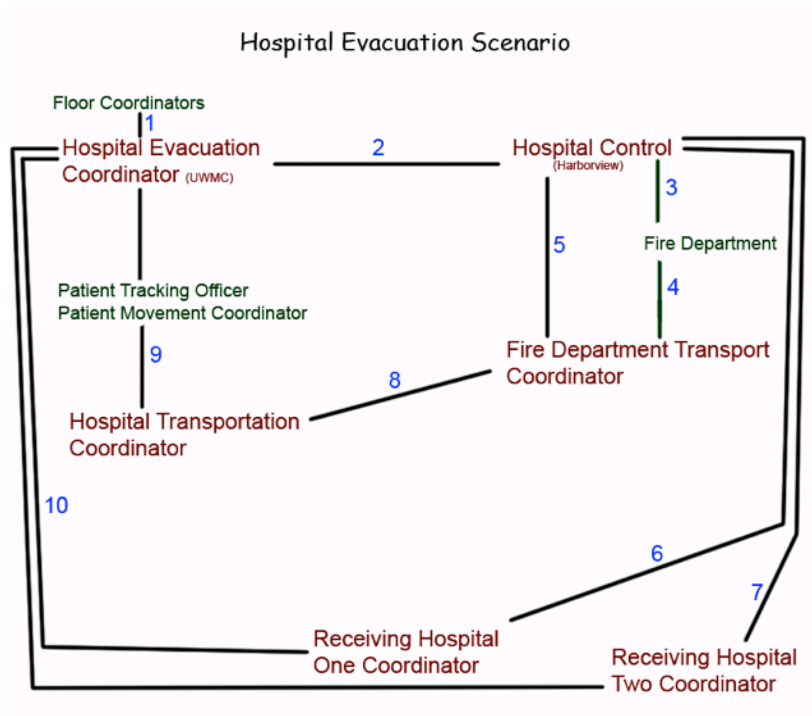
- *A multi-user situational simulation environment can be effectively used as a training tool for generating insight among emergency response personnel.*
- *A multi-user situational simulation environment can be effectively used as a training tool for improving situation awareness among emergency response personnel.*

To be able to test these hypotheses, we made changes to the simulator software components. We added new resource subclasses to include equipment resources that patients would need to have access to throughout their stay in a hospital. We added incident subclasses that were specific to response needs of patients in a hospital. We added the ability for incidents to be dependent on one another such that one incident had to be responded to completely before the next incident could be considered. And, we implemented a situation awareness scoring algorithm that would give players a rough estimate of how well they were responding to the current crisis as a team.

Upon outlining the code work, we approached the KCHC to help us effectively integrate simulator use into training drill support and research experiments in order to test our first hypothesis. We were able to modify and use an emergency hospital evacuation scenario developed in conjunction with the emergency response drill coordinator at the UW Medical Center, Tamlyn Thomas. The KCHC met independently to document the hospital evacuation scenario. As a result of their work, the team produced, and provided to me, the written documents shown in Appendix I.

Hospital evacuation is performed by defined roles identified in a variety of manuals and specifications maintained by the King County Emergency Response committee. A Hospital Evacuation Coordinator (HEC) in the evacuating hospital begins the emergency evacuation process by contacting all evacuating hospital floor coordinators who then provide a patient status report for all patients on each floor. The HEC contacts the prearranged Hospital Control (HC) contact at an external location to report on the current situation. The HC contacts the Fire Department who selects a Fire Department Transport Coordinator (FDTC) to be in charge of all physical patient removal performed by Fire Department staff. The HEC also contacts the evacuating hospital's Hospital Transportation Coordinator (HTC) who is responsible for coordinating patient transfer with the FDTC. A Patient Tracking Officer and/or Patient Movement Coordinator may be involved in the communications between the HEC and HTC.

To negotiate patient allocation away from the evacuating hospital, the HC communicates with each Receiving Hospital Coordinators (RHC) to prepare the receiving hospitals for the receipt of evacuated patients and gain agreement for transfer. The flow of communications between emergency hospital evacuation scenario roles is shown in Figure 11.



1. Floor coordinators turn in count sheets to HEC.
2. HEC asks for assistance from Hospital Control and faxes count sheets.
3. HC asks for help from Fire Department
4. who then determines an FD Transport Coordinator.
5. FDTC and HC begin discussions
6. HC works to find a receiving hospital for evacuated patients
7. HC works to find another receiving hospital for evacuated patients
8. FDTC (movement of bodies - if fire, horizontal movement out of fire areas & fight fire - too complicated for our scenario) and HTC (release of patients with medications, accompanying staff, and oxygen) co-locate to make communications easier.
9. HTC coordinates with the HEC (potentially using a PTC and/or PMC as intermediaries)
10. Receiving Hospital Coordinators secure electronic documents from HEC.

Figure 11 – Emergency Hospital Evacuation Roles and Communications

7.1 The Hospital Evacuation Scenario Preparations

With help from the KCHC, we were able to generate different patient incident classes that each required creative response unique to incident type. We were also able to generate different vehicle resource classes that would require creative matching of patient needs to vehicular capabilities. These incident and resource classes continued to be iterated upon for scenario performance training and improvement up to the time we performed the experiment that would test our second hypothesis.

In order to be able to evaluate emergency response participation, we enhanced the RimSim software so that it would generate and store the more relevant data we believed would be most important to visualizing emergency response team behavior during patient evacuation of the UW Medical Hospital to other area hospitals. We needed to capture data for the hospital evacuation scenario that would at least provide similar analysis to each pilot session we had run.

We further developed our skills in building our evolving data model into visual analytics widgets that could best provide insights into role player performance in light of overall team performance. As expected from our literature review, we found that the *Improvise* visual analytics framework allowed us to extend the *Improvise* visualization widget library to include our own widgets that were unique to emergency response in general and hospital evacuation in particular. In parallel to simulator development, we continued to develop useful visual analytics widgets based on feedback from our first response experiment participants.

With satisfactory versions of those features in place, we fine-tuned the tool to be ready for a variety of experiments that could perform first as agent-based simulations and then as live players and mixed live players with agents. To prove the merit of such work, we knew our simulator experiments should take place with emergency response personnel who could evaluate a simulator and/or analysis tool and demonstrate improved insight into better designing their role within the hospital evacuation scenario.

As the final step in preparing our software for hypothesis testing, we chose the date of Tuesday, April 20^{th} 2010 as the base date for providing starting conditions for our scenario. We obtained the complete list of bed allocations for the UW Medical hospital, and seeded our hospital floor plan component with visual representation of all 235 patients. The KCHC helped us encode patient incident types into visual glyphs that contained color coding comprised of two parts: A semi-circular solid color pattern on the left-hand side of the glyph would represent the mobility of the patient while a vertical rectangle solid color pattern on the right-hand side of the glyph would represent health class designation. The specific color allocations were as follows:

Left-hand mobility encoding color	**Significance**
Green	Able to move without assistance
Yellow	Able to move with wheelchair
Red	Requires human assistance to move

Right-hand health encoding color	**Significance**
Yellow	Stable without monitor
Pink	Monitored
Orange	Unstable
Purple	Critical

Upon placing visual glyphs on the five floors of the hospital to represent patient incident types, the patient tracking widget appeared with the five floor starting layouts seen in Figure 12.

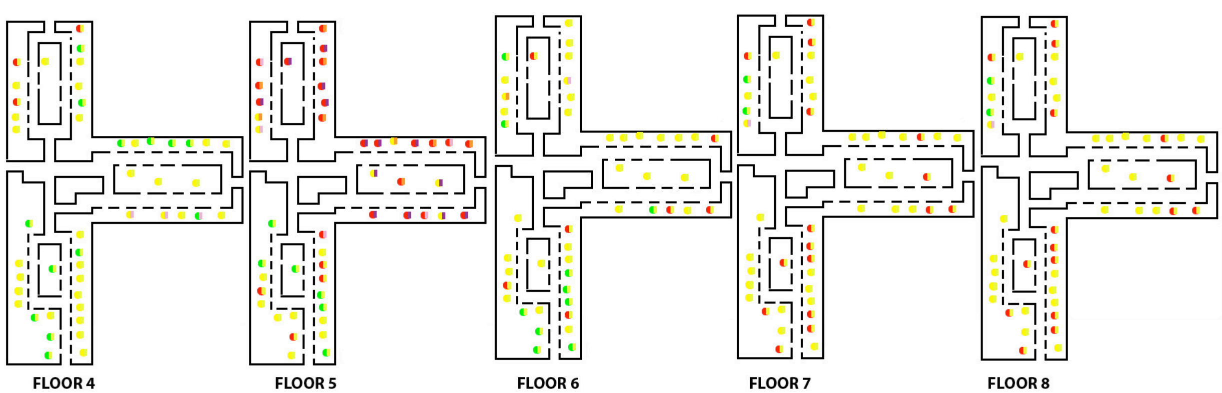


Figure 12 – Patient visual encoding and distribution by patient floor at startup

In preparation for our experiments, we also updated our hospital list and visual glyphs (including transportation routes from the UW Medical Center). The KCHC provided us with the nineteen hospital care locations that participate in the KCHC and would be contacted by Hospital Control in times of countywide emergency. The final list and presentation at software launch time appear in Figure 13.

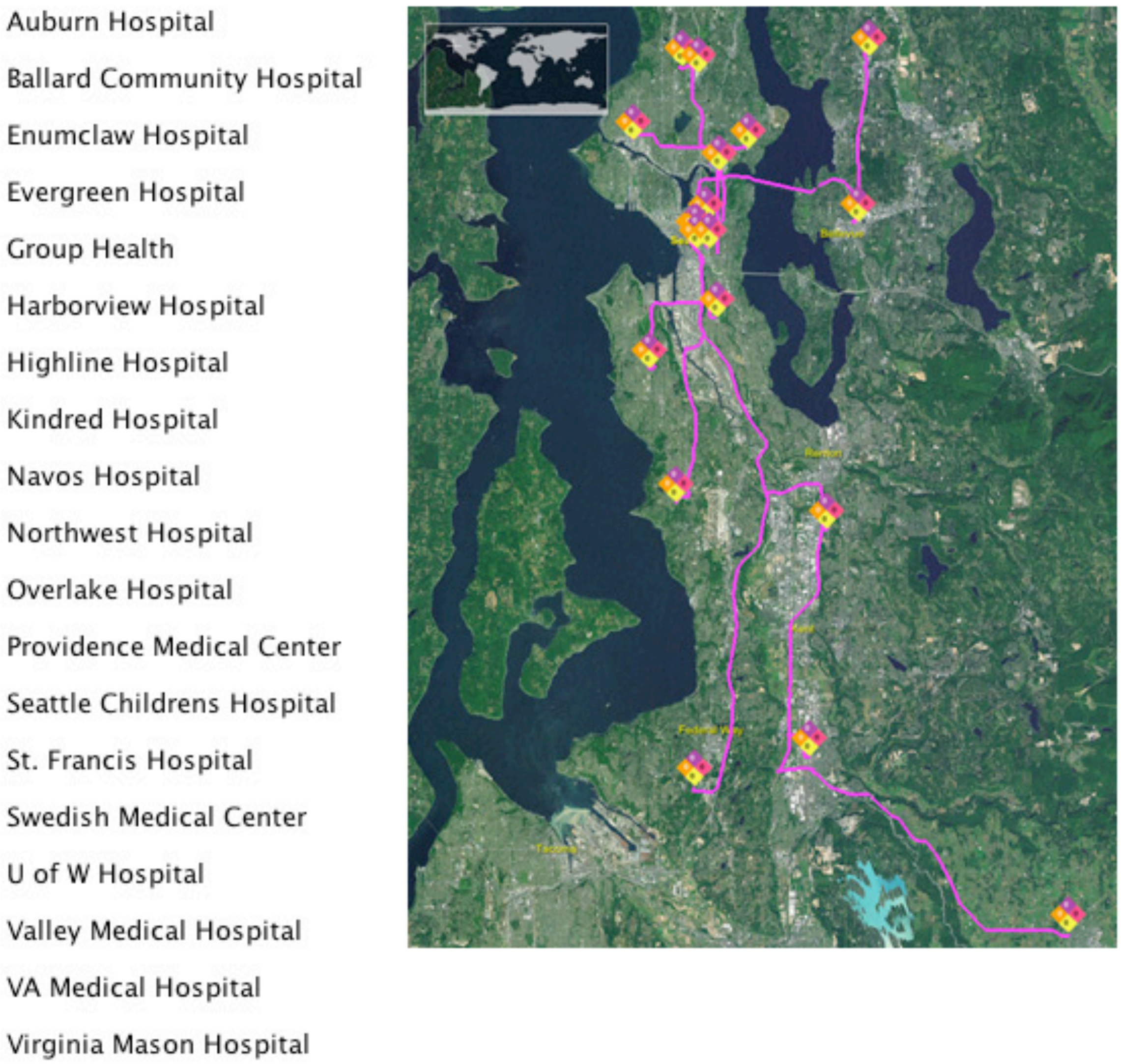


Figure 13 – Participating hospital list and visual map layout of hospital locations

To prepare for our experiments, we color-coded the vehicle resource types available for the scenario using the following color scheme:

Vehicle circular glyph color	**Significance**
Green	Ambulance
Yellow	CCT Medic

Red	High-end EMS Medic
Cyan	Eight-Seat Aid Car
Purple	King County Metro Bus

Because the literature consistently identifies interface as a key experimental design factor for software-augmented training tools, we continued to refine and verify our hospital evacuation interface with emergency response teams. The interface had been developed to test out the usability of our code with general emergency response agents instead of specific emergency response roles. This approach had great value for extensibility to a wide variety of emergency response scenarios, but we chose a specific hospital evacuation scenario to test the flexibility of our approach to encode new emergency response scenarios in a timely manner. Much of this work has been incidental, and yet necessary, to the hypotheses testing we pursued through formal subject experiments.

7.2 Experiment Design

Since we needed to focus our experiment design on testing the two hypotheses, we designed experiments with the expectation that both hypotheses could be tested simultaneously through dual emergency response drills: the first drill would be run using role play without the use of a computer interface for emergency response and the second drill would be run with the use of our computer interface. Our agent simulator is used to report the state of the hospital evacuation scenario from start until finish.

In the first scenario three key role players would attempt to fulfill their responsibilities in a simulated hospital evacuation scenario using the case study we had developed with the KCHC. We would use the state of the UW Medical Center on April 20th, 2010 and the current agreement between KCHC constituent hospital members to identify the incidents and resources involved in the simulation. As agreed upon in role negotiation, the Hospital Control, Hospital Transportation Coordinator, and Fire Department Transport Coordinator would play the simulation live and all other roles would be simulated by software agents. The simulated crisis would not involve any mitigating community crises occurring simultaneously that would require additional environmental modules to be included in the scenario.

We refined basic performance measurement metrics to be specific to the stated case study of an emergency hospital evacuation scenario. Time and motion parameters used for the experiment would include averages determined through literature review, personal observation, and metrics provided by respected authorities. For example, to determine stairwell evacuation times, we considered minimum, maximum, and average evacuation times provided by number of steps in the Glasgow Hospital study and adjusted them based on timed personal experience of walking the UW Medical staircases personally. To determine vehicle travel times to and from UW Medical to other hospitals, we used estimated travel times from Google routes provided by the Google Maps Application Programming Interface (API) at 2:00pm in the afternoon. Since we would use the same time and motion parameters in both experiment sessions, we were not concerned about them being exactly precise, but instead realistic.

Prior to running the study at the UW Medical center, the RSR development team in Seattle and a team in New Britain, Connecticut (to be introduced later in this document) repeatedly ran the RSR simulator in role-play sessions that would indentify performance bottlenecks. Upon removing bottlenecks to our satisfaction, we focused our attention on developing the appropriate hypothesis metrics to be certain that we collected all necessary data, during role-play with the software, for testing our hypotheses.

We attempted to develop the best appropriate metrics regarding insight and learning to test our first hypothesis and metrics regarding situation awareness in order to test our second hypothesis. Based upon prior review of the literature and the pilot studies above, we decided to test our second hypothesis (situation awareness) with in-simulation methods and test our first hypothesis (insight generation) with post-performance review metrics (that review all participant communications to agree on insight value as a contributor to shared knowledge base).

Given that we were working with well-trained, specialized emergency responders, we included representative role players in the development of a real-time freeze probe questionnaire for testing situation awareness and a post-activity self-rating technique for evaluating insight generation. The data collection details of our hypothesis-testing metrics are provided further below, but because we anticipated that our findings be highly dependent upon our choice of metrics, we first provide the following two sections to explain and support our choices.

7.3 Measuring Insight from Emergency Response Drills

To test the first hypothesis, we investigated responsive metrics for measuring insight. Much had been written about insight as a metric during the years of our simulator preparation. Insight has varying definitions that appear to focus on two concrete definitions pursued by two distinct research groups — computer scientists and cognitive scientists [120]. Computer scientists investigate insight as a contribution to knowledge building whereby each insight contributes to a relationally semantic knowledge base that enables problem solving and reasoning heuristics. In this regard, each insight is a describable incremental piece that adds value to the whole knowledge base — insight as a noun. Cognitive scientists investigate insight as a neurological function of the brain's left hemisphere where a new perspective on a problem is gained through a burst of brain activity — insight as a verb.

Experiments in the cognitive science realm consistently show a general pattern whereby left-brain activity spontaneously erupts in the pattern described as insight after the right-brain has been active grasping with a problem domain for a period of time. As a result, Chang et al. make a strong argument for using the knowledge-based insight metric for evaluating the value of external artifacts over the spontaneous thought metric because the former appears to be a necessary precedent to the latter given enough focused thought time with a robust enough knowledge base [120]. Since we are interested in improving performance in emergency response scenarios, we are limited by rules, regulations, and legal actions that can be taken by first responders under the urgency and duress of a first response crisis event. Spontaneous insight as "aha moment" may be very useful for long-term planning processes, but insight that leads to better actions in response to the current crisis conditions suggests that insight to build a better knowledge base for improving distributed cognition is a useful metric in its own right.

Developing the hospital evacuation scenario as a realistic activity to build emergency response skills in the community had already required insight from all role players involved in the scenario. The KCHC agreed to a five-step process for training participants in their responsibilities during a hospital evacuation scenario. The five steps, listed chronologically, involve discussions to scope the scenario, a tabletop exercise to run through the scenario in front of all KCHC members, a paper-based drill to role-play the scenario, a computer-based drill to role-play the scenario, and a physical drill with actors to represent patients in need of evacuation.

We anticipated that the five-steps would likely provide insight at each step of the process. We have described the insights that the scenario development team gained through the various discussions they shared and documented through the scenario-scoping period and consider that part of our experiment preparation period. Below we also describe the insights that were gained during the regional evacuation tabletop exercise that occurred at Evergreen Hospital on March 3, 2010. This also occurred before we ran our experiments to compare paper-based drills with computer-based drills in order to evaluate the ability of each to provide insight to drill participants. In those experiments, we would identify insights and evaluate them in terms of their incremental value to the knowledge base associated with emergency response to a hospital evacuation scenario.

We would use our post-session visual analytics tool to review all communications between players and earmark communications that the team agreed represented insight into the knowledge-base necessary to succeed at the hospital evacuation activity. The team would then provide their own rating scale, which we would document and then challenge the team to use consistently between drill session evaluations.

Insights come from many people and places when developing relevant and representative scenarios for training first responders. In the case of the KCHC hospital evacuation scenario, the council met many times to build a scenario that would refine and test developed guidelines and procedures. Scenario developers met often in groups of twos and threes, with Tamlyn Thomas often in attendance. Document review sessions revisited existing forms and procedures mandated or recommended by national and regional standards. Roles and responsibilities were reviewed over and over in order to make sure the roles and responsibilities associated with a hospital evacuation were appropriately in line with document completion and successful first response as identified by the National Incident Management System (NIMS).

As stakeholders began to reach a consensus regarding regional hospital evacuation roles and procedures, the council determined enough progress had been made to warrant a tabletop exercise be run with all key stakeholders in attendance. In the past, tabletop exercises had provided new insight and challenged previous insights through rigorous discussion of regional first response activities through focus on a particular first response scenario. Nothing in the nature of a hospital evacuation scenario suggested similar insight would not be gained through a regional medical evacuation tabletop exercise. The KCHC convened on March 3rd, 2009 , in for a "Regional Medical Evacuation Tabletop Exercise and Pediatric Annex Education Session". The pediatric annex was chosen as a special sub-interest component of the broader tabletop activity since the hosting hospital could easily provide a pediatric team to attend the session.

The session ran for three hours with the first hour spent with a review of the current rendition first response documents developed for emergency response purposes (these documents continue to be iterated upon today and will continue to be iterated for the foreseeable future). As the KCHC already felt confident about the documents that had been provided and had discussed them heavily before the organizing the tabletop session, no insights were raised during the first hour of the March 3rd session. As invited observers, we noticed a clear consensus and appreciation for the state of the documents being used to advise and report upon the tabletop exercise that was to be run during hours two and three of the March 3rd KCHC meeting.

Before beginning the tabletop exercise, the session coordinator reiterated the goal of the exercise explicitly via PowerPoint slide presentation and verbal acknowledgement:

> To orient hospitals and response partners to the Regional Evacuation and Patient Tracking Mutual Aid Plan as well as the newly completed Pediatric Annex and test components of the plan, such as transportation coordination, patient tracking, and identified roles and responsibilities.

Twenty-nine people participated in the tabletop exercise, including multiple representatives for each key role identified in our simulator. Both the primary and backup Hospital Coordinators attended, four Hospital Transportation Coordinators attended, two fire departments with their Fire Department Transport Coordinators attended, and eight Receiving Hospital Coordinators attended. The role players who would participate in our software-supported drills were among these attendees. In addition, two overall domain expert tabletop exercise coordinators were brought in to facilitate the discussion associated with the drill.

Other special experts with relevant pieces of domain knowledge attended including Hospital Emergency Managers, Nursing Managers (from surgery, pediatrics, intensive care, labor, and deliver), Emergency Medical Service Technicians, King County Metro Employees, and Local Emergency Management. All attendees were invited to interject in the tabletop exercise in order to provide insight to the KCHC as to how first response for a hospital evacuation scenario could be improved. For those unwilling to interject comments during the exercise, an after exercise debriefing known as a hot wash to participants was scheduled for the end of the exercise.

We attended the tabletop exercise to perform two duties:

1. To review the hospital evacuation scenario in order to assess the appropriateness

of our adaptation of the RimSim simulator for supporting a hospital evacuation training session.

2. To record the insight detail, timing, and contribution associated with each insight identified through the tabletop exercise.

The first duty pushed us to gain our own insights as to the nature of the scenario while the second duty allowed us to record tabletop exercise insights for comparison and contrast to future software-supported sessions.

The following table records the results of duty number two above:

Table 1 - Insights During March 3rd KCHC Hospital Evacuation Tabletop Exercise

Insight	Discussion Time	Contribution
Media management provides a heavy workload for the evacuating hospital – preferably by people not associated with key evacuation roles	9:50 – 9:54	Reduction of Hospital Transportation Coordinator's responsibilities in regard to the media.
The media is omnipresent and likely to broadcast information on a hospital evacuation event soon after the event begins	9:58 – 10:03	If possible, each patient should have a media management strategy for getting information to concerned friends and family of the patient. This will likely interfere with the physical relocation of patients.
Supplies for evacuating patients are often as important as tracking the patients themselves. And, yet, communications with suppliers have not been coordinated as well as communications with patient providers	10:12-10:20	Improvement of supplier relationships and coordination of supplier relations among KCHC hospital staff.
Hospital Control should not allocate patients to nearby hospitals as a first choice of action since those hospitals will have to take over new intakes and could still provide beds on the tail end if absolutely necessary.	10:32-10:34	Change in expectations of regional hospitals, especially receiving coordinators who are located far from evacuating hospital.

In fact, evacuating hospital should not recruit any first response aid workers from any regional hospital since personnel will be needed at receiving hospitals to make creative solutions for intake	10:35-10:36	Change in expectations of Hospital Transportation Control as to who will be available to perform the duties associated with decisions made.
Receiving hospitals that commit to accepting relocated patients should cancel all surgical procedures for the day in case crisis escalates (contentious issue).	10:39-10:45	Change in awareness of regional hospitals in terms of effect of any hospital evacuation on a regional hospital's day-to-day operations.
Categorization of patients in order to make transport decisions is critical yet not fully optimal yet	10:45-10:54	Additional work required by Hospital Control and Hospital Transportation Control roles in order to be ready to decide and communicate transport needs.
Initial short-term surge numbers did not generate enough capacity for the patient evacuation numbers given the number of beds available on the WA Trac system.	11:01-11:07	Awareness of Receiving Hospital Coordinators of the criticality of clearing beds and thinking creatively as to how to accept more evacuating patients up front.
Differences in equipment types among hospitals exist but are called the same name across hospitals.	11:15-11:18	Awareness of Hospital Control, Hospital Transportation Coordinator, and Fire Transport Coordinator as to potential for miscommunication in arranging equipment for patients.
Concern related to equipment availability and tracking so that equipment gets back to lending hospital	11:22-11:26	Awareness of Hospital Control, Hospital Transportation Coordinator, and Fire Transport Coordinator as to potential for hesitation in sharing equipment
Supplies staging is still in the air but planned for stability that will support this scenario	11:28-11:29	Awareness of Hospital Transportation Coordinator and Fire Transport Coordinator as to unpredictable delays associated with getting needed supplies to the evacuating hospital

We observed direct discussion of the eleven insights above take up thirty minutes of the two-hour tabletop exercise.

7.4 Measuring Performance-based Situation Awareness

To test our second hypothesis and evaluate changes in situation awareness levels when comparing two emergency hospital evacuation drill sessions, we developed a measurement process after evaluating the existing successful situation awareness measurements that have been found in the literature as documented in section 2.1.2 above.

At first glance, good team situation awareness might appear similar to good distributed cognition, but using Hutchins definition of distributed cognition from section 2.3.1, we see how interwoven distributed cognition is to the process measurement and performance measurement techniques of situation awareness reviewed above. A well-designed process performs well when the distributed cognition is embedded in the external world such that the participants naturally perform the process better through expertise captured in the external world. But we are not trying to be designers of the process. We are attempting to measure the current process to suggest the level of situation awareness provided by the participants and, ideally, the team as a whole.

Many team skills are difficult to measure using the techniques mentioned in section 2.1.2. Because we want to measure team situation awareness levels in an existing process, we don't want to design or engineer the process as part of our body of work. Although that work could likely be fruitful in improving team performance and improving the overall resiliency of the team to potential variables in a real crisis, it is not the aim of this work. We aim to measure the current process that may shed light on needs to re-engineer the process. That would be a side effect of our work.

Although we wish not to engineer the process, we may wish to engineer the simulation to be able to measure team situation awareness. Through designing the specifics of the emergency, we can manipulate the scenario to raise conditions by which the skills that lead to and suggest having good team situation awareness are demonstrable. For example, in considering Prince and Salas' list of recommended skills, we can create example situations in the hospital evacuation scenario when multiple participants should demonstrate their suggested skills:

- By impairing a key transportation route (creating a bridge outage), we can test the ability of the team to identify problems.
- By overcrowding an area of the hospital beyond the Fire Marshall's stated capacity, we can test the ability of the team to recognize the need for action.
- By sending different in-the-field updates to key roles in the hospital evacuation process, we can test the ability to determine root causes in discrepancies.
- By injecting events that are not relevant to hospital evacuation, we can test the ability of the team to demonstrate awareness of an overall goal (by ignoring superfluous information).

And although not all skills are strictly measurable in terms of response to the simulation, the mapping does suggest that each skill can be considered independently when looking for an appropriate metric to measure that skill. We can pick the best measurement technique for measuring each skill and then sum up the measurements for an overall measurement of goodness.

By suggesting we consider the skill of the team to *note deviations in situation awareness across members and between members*, Prince and Salas made a strong suggestion that we ascertain individual situation awareness in order to identify deviations so we can measure whether the team notes the deviation. We can choose between all the measurement techniques identified above and yet know that if we choose poorly on measuring just one skill, the measurement of other skills can help compensate.

By suggesting we consider the skill of the team *ability of team members to exchange information for prevention of errors*, Prince and Salas touch upon the basic requirement of perception of the current situation and the sharing of those findings. In our hospital evacuation scenario, we can measure the ability of team members to share information by tracking the entropy of the location of patient awareness as an aggregate across all roles, weighted by their responsibility to know that information in detail. We appreciate a mathematical model measure of situation awareness works very well for a hospital evacuation scenario in that we can track the movement of evacuated patients and define uncertainty as the possible distance the patient might be from where he or she is expected to be. Uncertainty is a complex metric and is based on probabilities of many factors. But, we can measure the uncertainty as the possible distance each patient is from expected location based on the most recent communicated update from a member of the response team. Such communication can come from a simulated responder or a simulation participant.

The skill-by-skill approach to measuring team situation awareness acknowledges the shortcomings of the methods used to identify individual situation awareness and attempts to water down each individual shortcoming by mixing the techniques for each skill measurement in order to make any one skill's measurement immaterial to the overall assessment. Our focus then becomes one of identifying all relevant skills instead of stressing over choosing the appropriate method. Spending time on identifying the skills sheds light on designing effective training methods as well. A natural way of identifying skills is by determining which skills lead to the best overall result for the shared activity. The metric for success in a hospital evacuation is highly weighted by the time it takes to safely evacuate all patients from the hospital at risk and have them properly cared for en route to receiving hospitals.

As described earlier, situation awareness quantification can occur by comparing an individual's perception, comprehension, and projection to some ground truth reality. In the team case, team situation awareness quantification can be measured by comparing each individual's perception, comprehension, and projection to each other participant's. In the case of a simulated hospital evacuation, the ground truth is simulated based on known hospital patient evacuation timelines [121] that represents reality in the context of a drill. The more concurrent the individuals' reported state of awareness with reality, the higher the value of situation awareness.

To quantify situation awareness, willing participants are interrupted at ten random times while performing their roles, including all simulated activity, in order to test their current level of situation awareness. Situation awareness is ascertained by asking open-ended questions and recording verbal responses that demonstrate the current state the participant experiences.

Our situation awareness questionnaire consists of five questions that subjects are expected to answer within thirty seconds to minimize interruption to the drill. The following questions were chosen upon discussion with the KCHC as being highly relevant to a hospital evacuation scenario:

1. How many patients are in a significant state of discomfort currently?
2. Where are these patients located?
3. How many patients are currently in transit between the evacuating and receiving hospital?
4. How much more time will it require to fully evacuate the existing hospital given ideal circumstances?
5. How much more time will it require to fully deliver all evacuating patients to their receiving hospital given ideal circumstances?

The answers to these questions would be objectively compared to the actual state of the drill to ascertain situation awareness. The closer the quantitative response reflected quantitative reality, the higher the level of situation awareness metric we would report in our conclusions.

We also embed quantitative measures of situation awareness into our software based on the success of Pritchett's demonstration of wide applicability. Unlike measurement techniques that attempt to ascertain the subject's mental model of the situation at different times throughout an experiment, performance-based testing focuses solely on the subject's outputs. This quality makes it ideal for comparing the desired and achieved performance of a human-machine system, and for ascertaining weak points of the subject's situation awareness. We inject conditions into the emergency hospital evacuation scenario that test situation awareness by setting up situations whereby if the subject has sufficient situation awareness, an action is required. By doing so, we aim to provide an unambiguous accounting of the types of tasks for which the hospital evacuation decision-makers had sufficient situation awareness.

7.5 Data Collection

Data collection took place concurrently with the research experiments and complied with the data collection document submitted with the UW Human Subjects Division as seen in Appendix VII. A pre-study questionnaire asked each subject the following three questions:

- What role would you perform if a hospital evacuation emergency response activity were required of you today?
- How many months have you been in that role?
- Do you have any specific personal characteristics that would make your performance in a hospital evacuation emergency response drill significantly different than someone else with your training? If so, what are they?

The first of the two emergency hospital evacuation drills, the paper-based condition, accumulated data throughout the drill including:

- Text and timestamp (nearest second of clock time) of any and all voice utterance(s) uttered and/or overheard by each subject during the drill (as transcribed from voice stream).
- Latitude, longitude, and altitude of each (anonymous) hospital patient being evacuated every minute (time stamped via the official drill clock) of the drill along with a conversion to known descriptive location when describable (e.g. second floor Pacific Tower lobby elevator landing).

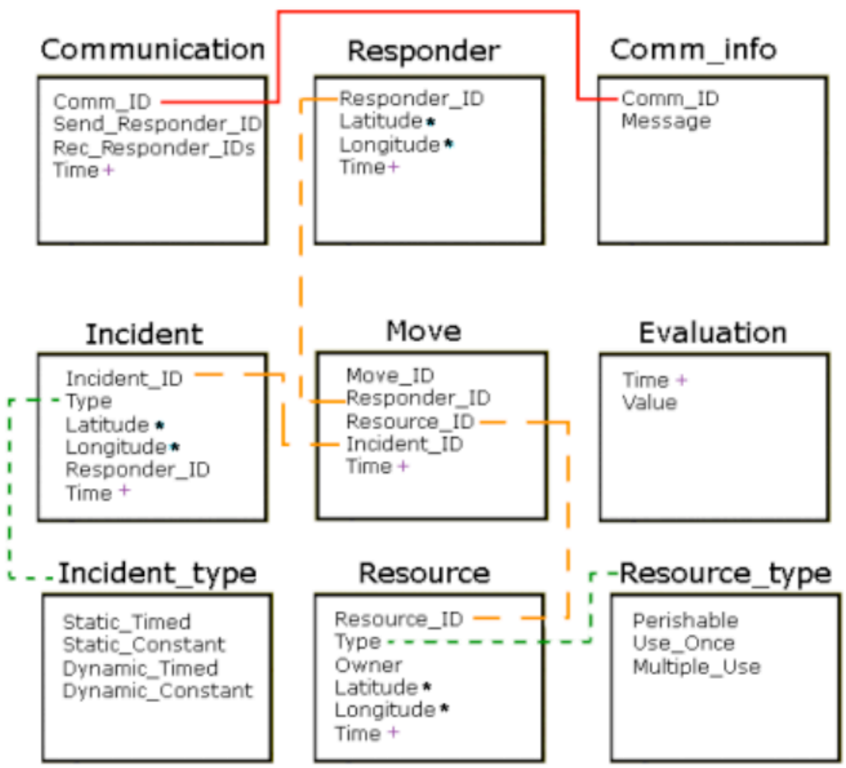
- Latitude, longitude, and altitude of each live and simulated responder personnel during evacuation every minute (time stamped via the official drill clock) of the drill along with a known descriptive location when describable (e.g. second floor Pacific Tower lobby elevator landing).
- Latitude, longitude, and altitude of each injected incident (e.g. Pacific Tower Elevator outage) during evacuation drill (time stamped with start and end times via the official drill clock) along with a known descriptive location when describable (e.g. second floor Pacific Tower lobby elevator landing).
- Latitude, longitude, and altitude of each medical supply (e.g. water bottle, ice bag, ambulance, etc.) during evacuation every minute (time stamped via the official drill clock) of the drill along with a known descriptive location when describable (e.g. second floor Pacific Tower lobby elevator landing).
- Responder ID for sender and recipient(s) for each command made in the drill – both simulated and live participant (along with timestamp of command start).
- Responder ID attached to patient ID for the duration when a responder is responsible and accountable for that patient during the drill.
- Answers to all questions in the situation awareness questionnaire seen in Appendix VI, along with timestamp to the nearest minute of the simulation clock for when the questionnaire is implemented (ten times per drill).

Data collection for the second emergency hospital evacuation drill was identical to the first drill with the added capture of continuous mouse cursor location on the screen of the visualization tool being tested, and the collection of all mouse button up, down, and drag events. Continuous mouse cursor location was captured every 30 milliseconds whenever the mouse cursor was moving on the screen along with a timestamp of start and end of each movement (in milliseconds via each participant's built-in computer clock which were synchronized before start and verified as being synchronous at drill end).

We asked one additional question of subjects before the second drill that was inapplicable to the first drill:

- How much time have you spent gaining a basic comfort with the visualization tool before this drill begins?

We verified that the simulator's data capture routines would allow the running of both experiment conditions to fill out the data model in Figure 14 as accurately as possible within ample time precision (as described above).



+ after data attribute means zero or more
* after data attribute means one or more of them

Figure 14 – Data Model for Experiment Data Collection

As our response team visual analytics evaluation tool uses this exact data model in visualizing the simulation session, we were able to test proper software data capture during software stress testing. In order to provide an opportunity for subjects to guide us in performing future emergency response simulated roll play experiments, we asked for open-ended feedback regarding our experiment design at the end of our contact time with them. Specifically, we asked:

- Please let us know any thoughts from participating in the experiments with which you don't mind going on record.

7.6 Experiment Schedule

We requested a Human Subjects Division (HSD) review of our proposed study on November 14, 2009. The HSD committee responsible for assigning study applications to personnel requested that we attend a meeting at their facility that took place on December 11, 2009. As a result, we made changes to the study to make the review process easier and an approval more likely. The amended study documents were approved on April 9, 2009 and contained the following agreed upon experiment timing.

The first stage of our experiments required our KCHC personnel to perform their usual roles within a two-hour long emergency hospital evacuation drill at the University of Washington Medical Center (UWMC) on April 26th, 2010. This drill was a drill that Tamlyn Thomas, the UWMC Emergency Management Coordinator, had wanted to run for some time with handpicked personnel she chose who would benefit from such a training exercise. She actively recruited participants for the drill during February 2010. As a result of her recruitment, there were five roles identified as participating in the drill (with one person in each role except two each for the receiving hospital roles):

1. Evacuating Hospital Control Coordinator
2. Fire Department Transport Coordinator
3. Evacuating Hospital Transportation Coordinator
4. Receiving Hospital One Coordinator
5. Receiving Hospital Two Coordinator

Upon determining the participants in the drill, we were to request voluntary participant inclusion in the research study aspects of our experiments. As explained below, we eventually determined that this first group of participants would not be ideal for our published experiment results and we recruited a second team of subjects based in New Britain, Connecticut, in the United States.

Based on the recruitment process results, we concluded that the roles of Evacuating Hospital Floor Coordinators (one per patient floor) and Evacuating Hospital Evacuation Coordinator would be simulated through the use of our simulation software (no live simulation role players for those roles). The movement of fictitious patients throughout the evacuating and receiving hospitals and the road networks between hospital locations were also to be simulated with software.

As a stipulation of the human subjects review, both the author and Thomas agreed to use the evacuation scenario software during the drill to provide drill participants with data they would normally gain access to in any drill or real life case should the hospital evacuation scenario happen in the future. The data coming from the simulation software were to be strictly related to simulated patient locations and obstacles (physical and time delays) encountered to desired patient movement. No real person patient data would be used for movement.

For those who would agree to be voluntary experiment subjects for our research purposes, we would measure situation awareness using the data collection and metrics described above while the drill took place. During the first drill that took place on April 26th between 10am and Noon, drill participants communicated with each other via voice to perform actions in the drill and we captured the voice utterances made and heard along with timestamps for subjects but not for non-subjects. As the first drill was a paper-based drill, we also provided updates to all participants as to the progression of the simulation whenever asked by any participant.

Without a computer interface at their disposal, participants used pen or pencil and paper to incorporate simulated scenario state into their decision-making process. The Hospital Transport Coordinator generated Exhibits 1 and 2 of Appendix I to keep track of patient encodings and patient release events. The Fire Transport Coordinator generated

Exhibits 3 and 4 to keep track of patient vehicle allocations and the status of vehicle movement. Hospital Control created a simple list of hospitals and placed numbers from receiving hospital discussions into columns that aggregated numbers into patient types according to the categories requested by the Hospital Transport Coordinator in Exhibit 1 of Appendix II.

Within an hour of the completion of the first drill, Tamlyn debriefed the first drill participants after they performed their drill in a manner consistent with all previous drills she has coordinated. As this is normal protocol without the investigators involvement, we were not present and the debriefing was outside of the scope of our research study.

A second two-hour long drill took place two days after the first drill, on April 28th, between Noon and 2pm. The same key participants as in the first drill, and who were the focus of our experiments, performed their tasks a second time, but with a different hospital evacuation scenario. The participants performed their roles with the addition of the role support software that we were testing for providing superior situation awareness and insight generation.

The role-play software required a conventional personal computer for each participant, with a keyboard, monitor and mouse. Our agreed-upon protocol suggested that subjects would have ample opportunity to gain familiarity with the software at their own leisure via a Web-based process. In the second drill case, situation awareness metrics were potentially available through data collection in the software in addition to the same situation awareness questionnaires we used in the first drill.

As a final opportunity for participants to provide feedback, the KCHC coordinator, Tamlyn Thomas, debriefed the KCHC participants after they performed the second drill in a manner consistent with all previous drills she coordinated. This debriefing was intended for KCHC use only and was outside the scope of the documented research study.

Chapter 8 – Results

We followed the experiment schedule reported in chapter 7 strictly except for the expectation of providing ample time for drill participants to practice using our simulation software via the Web before participating in the second drill. We froze software development on April 21, 2010 and provided software to all participants at least three days before the second drill began but the participants did not have any time to practice the use of the software.

Unfortunately, due to other scheduling constraints, we were not able to complete the Web-based version of our simulation and have it work satisfactorily for participants to practice leisurely on their own computers. As a result, we provided all the computers that were used in the drill. We recorded all the data for the simulated drills, with the exception that we did not conduct the freeze-probe situation awareness questionnaires since our protocol was not ready for situation awareness testing, as we were uncertain the software would be transparent for participants.

As a consequence of the lack of subject pre-training on our simulation software, we decided to use the drills to evaluate insight generation and perfect the simulator with what was obviously a highly experienced group of role players. We would use the individual player’s results to evaluate internally without publishing nor sharing with others. We did not pursue formal subject participation requests from the KCHC pool of potential subjects, and accordingly, we cannot provide individual data for role players in the KCHC drills. Instead we report only the aggregate team results.

Even though additional work would be required and we’d have to do our own recruitment to find qualified first responder personnel who could competently role-play with our simulator, we believe this decision was not only the right decision to make, but significantly motivated us to be better prepared for the next willing and appropriate participant group we could find for our formally documented experiments.

8.1 Full Protocol Experiments

Upon performing our own recruitment process using the forms and process approved by the University of Washington's Human Subjects Division, we recruited and received consent to participate in our research experiments from two employees of a Veteran's Administration Hospital (VAH) and a recently retired fireman in Hartford County, Connecticut. Our consent letter appears in Appendix V. The two VAH employees had jobs that required significant hospital logistics work and had interacted with doctors and nurses on site for multiple decades. They both reported having thought of hospital evacuation in their careers and were enthusiastic about gaining insight into such an activity through paper and computer-based role-play. The retired fireman had performed community-wide fireman services for over thirty years and was highly trained in community-wide first response activities that were easily adapted to the hospital evacuation scenario. To include him in our subject pool meant we would not be interfering with any potential emergency response tasks on the dates we chose for our two simulated role-playing sessions.

We performed the complete list of steps in the experiment schedule described in section 8.6, with the exception that the first paper-based session took place on June 12th, 2010 and the second session took place two weeks later on June 26th, 2010. Role-play took place for two hours each session with the addition of ten, approximately two-minute, simulation breaks to administer the freeze-probe situation awareness questionnaires.

We asked our pre-study questionnaire on June 12th, 2010. Respondents' answers to these questions are contained in Appendix VIII. Compared with the results from KCHC drill participants, the Hartford County participants' roles were less defined. We anticipated this difference, as the hospital evacuation scenario had not been part of any organized Hartford County hospitals development. The hospital participants had similar years of experience, and neither had much experience in the roles they were to play in the simulation. But the KCHC personnel had been through the scenario-generation discussion where their roles were defined and had participated in a KCHC-wide paper-based simulation where they did not perform their roles but watched them being performed by the collective. KCHC personnel had less experience with the software before participating in the second simulation session.

Again our strategy in these investigative procedures was configured to test the insight generation and situation awareness hypotheses presented earlier in this document. Metrics for both insight generation and distributed situational awareness were tested through drill support and experimentation. As a result, each metric, supporting a different, yet likely correlated, hypothesis is described in its own this section that follows.

8.2 Insight Generation

To calculate our insight generation metric for the two emergency response drill role-playing sessions, we transcribed the communications and actions made by role players into data tables that could be imported into our post-simulation Improvise-based visual analytics tool. The tool allowed the team to review their performance, identify insights made and score the insights on a scale of one to a hundred as to the significance of that insight to their shared knowledge base for performing an emergency hospital evacuation.

An example configuration of our visual analytics tool is shown in Figure 15. Participants reviewed the geospatial relationships of all incidents and resources and reviewed their communications by utterance, timing, and relationship to actions taken. They then extracted the list of insights they acquired in the order they gained them, created a visual list of the insights, and then scored them on a scale of one to one hundred where one hundred meant critical to the success of a hospital evacuation activity and lesser numbers represented the contribution they thought the insight had to success of a hospital evacuation activity.

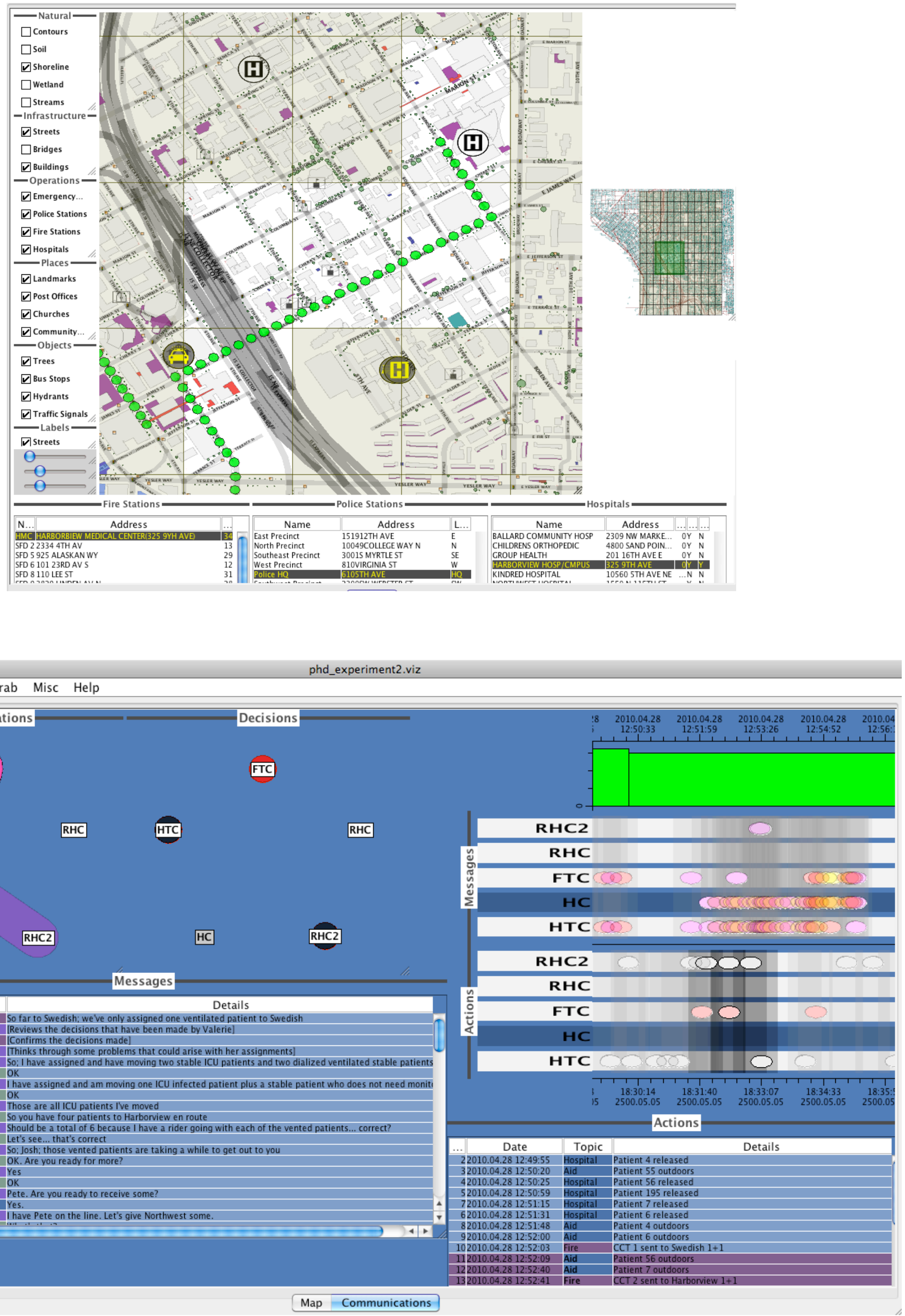


Figure 15 – State of the Visual Analytics Tool During An Insight Analysis Session

Upon sharing a visual analysis session, all participants listed all significant insights identified for the two drills and reached a consensus score for each. The results for the paper-based session follow with the role that first gained the insight listed and the time of that insight for all insights that scored higher than 10:

Table 2 - Insights and Team Score During Paper-based Drill

Role	Time	Insight	Value
FTC	4:12	My task is highly dependent on my knowing the timing of routes to receiving hospitals.	25
HTC	5:45	The hospital has many critical patients who need assistance on the top floor.	40
HTC	5:57	I care about the discomfort of patients and thus don't want to start releasing critical patients until they can get out.	60
HTC	6:12	The hospital has many ambulatory patients who can walk out on their own on the fourth floor.	25
FTC	8:28	Monitored patients are important to track in order to do transport allocations	60
FTC	10:18	I have a lot of time to help the Hospital Control role before patients are actually available for transporting	72
HTC	10:45	The ambulatory patients from the lower floors could already be out and on a bus ready to go to a receiving hospital.	50
HTC	11:29	The success of a hospital evacuation is highly dependent on my prep work with the patient tracking sheets	90
FTC	12:04	The ambulatory patients, especially stable ones, can wait longer outside before getting on transport	60
HTC	12:26	My task is highly dependent on the time it takes to clear assisted patients from the floor.	80
HTC	13:14	My task is highly dependent on getting information from Hospital Control	72
HTC	13:58	My task requires me to think about the location of and distance to the stairwells.	15
FTC	14:21	When considering the eight car and the metro bus, I need to make a rule of thumb based on wheelchair capacities	30
HTC	14:23	My task requires me to think about the availability of trained people to accompany assisted patients.	40
HC	15:33	I could help out much better in choosing my receiving hospital negotiation strategies if I knew the plan better	80
FTC	15:49	I really need to pay attention to critical patients as to when they become available for transport allocation	55
HTC	16:50	The most important consideration I need to focus on is tracking ambulatory versus assisted patient designations.	75
HTC	18:30	ICU bed availability at receiving hospitals is highly critical to my initial patient release planning	45
FTC	20:51	The choices I make are important for Hospital Control to consider in negotiating transfers with receiving hospitals	70
HTC	21:13	It takes longer to get the assisted patients prepped and out of	35

		the hospital than I anticipated	
HTC	25:34	It really helps the Fire Transport Coordinator when I verbalize the patient number of those I release.	50
FTC	26:28	Hospital Transport Control prep time has taken more time than expected and reduced that time for patient movements.	70
HTC	26:35	To avoid tying up the stairwells, I need to understand how wheelchair patients will get down and how long it takes	35
HC	29:09	I need a lot more information from the Hospital Transport Control role in order to do my task effectively	80
FTC	37:00	I am having a hard time deciding when I should send the Metro Bus to a receiving hospital	35
HTC	41:43	I need to be able to provide better summary data in order to answer questions from Hospital Control and Fire Transport Control	80
FTC	49:38	I need to create better visual aids for keeping track of vehicle availability – especially when assigned vehicles return	60
FTC	57:29	Some assisted patients take significantly more time to get outdoors than I had anticipated	40
FTC	59:04	I could be helping and planning better if I asked for thorough verbalization and confirmation of patient numbers	50
FTC	1:07:53	The tabular summary data I have been keeping is helping me make decisions	60
FTC	1:12:30	Now that I better understand the flow of patients from the hospital, I realize I could have had more time to help the HTC with the initial plan	50
HTC	1:22:06	I really could use help internalizing the wheelchair patient evacuation time versus ambulatory	60
FTC	1:31:17	Anticipating bus and eight car travel times are critical to my task	40
FTC	1:42:16	I am really struggling with keeping up with the ventilators and monitor equipment needs of patients	55
HC	1:48:37	Once I get good at receiving hospital negotiations, I have lots of time to help the other roles.	45

In addition to the insights listed above, there were seventeen other insights generated but deemed to be less significant than the above chronological list.

The results for the computer-based session follow:

Table 3 - Insights and Team Score During Computer-based Drill

Role	Time	Insight	Value

HTC	2:17	I care about the discomfort of patients and thus don't want to start releasing critical patients until they can get out.	60
HTC	2:42	The hospital has many critical patients who need extra time and assistance on the top floor.	40
HTC	4:52	The hospital has many ambulatory patients who can walk out on their own on the fourth floor.	25
HC	4:53	Watching the patient allocations is going to help me anticipate needs when negotiating with receiving hospitals	60
FTC	4:55	There are many ambulatory patients on the lower floors who can be evacuated quickly and put on a Metro bus or eight car	50
HTC	5:19	I can plan my use of the floor nurses before I begin allocating assisted patients	40
FTC	5:22	I have a sense of basic priorities of which hospitals I'd like to allocate to based on the map	40
HTC	6:08	The patient hospital interface does a lot of the prep work for me if I can trust these encodings	80
FTC	7:00	I can help Hospital Control make overall sense of the patient encodings while Hospital Transport works on the details	75
HTC	7:29	The success of a hospital evacuation is highly dependent my ability to internalize these patient color encodings	60
HTC	8:16	My task is highly dependent on the time it takes to clear assisted patients from the floor.	60
HTC	8:32	The ambulatory patients are almost all stable	25
HTC	8:57	The ambulatory patients, especially stable ones, can wait longer outside before getting on transport	60
FTC	9:09	The HTC task is highly dependent on the time it takes to clear assisted patients from the floor.	80
HTC	10:21	To be successful, I need to review the allocations from Hospital Control each time I make significant patient releases	70
FTC	11:43	When considering the eight car and the metro bus, I need to make a rule of thumb based on wheelchair capacities	30
FTC	14:28	The ICU bed availability at receiving hospitals is a critical discussion point between me, HC, and the HTC	50
FTC	16:29	The ambulatory patients from the lower floors could already be out and on a bus ready to go to a receiving hospital.	50
FTC	20:13	I need to verbalize the transportation choices I will make so HC can anticipate my needs when discussing with receiving hospitals	62
HC	21:44	I can spend more time verifying HTC decisions instead of always asking for the plan with this interface	70
HTC	21:48	There are some summary data values that take me too much time to calculate when asked by the FTC or HC	42
HTC	23:07	The most important consideration I need to focus on is tracking ambulatory versus assisted patient designations.	75

HTC	25:16	I am getting a good sense of how the flow of patients unfolds in the simulation	40
HC	25:42	Equipment planning might be just as important as bed availability planning	90
HTC	26:17	Although tempting to use color descriptors of patients, patient number is still a better identifier although the redundancy is often helpful for finding patients visually	50
FTC	27:40	Anticipating bus and eight car travel times and how often I can expect them to return in time is critical to my task	50
HC	27:58	Equipment planning might be just as important as bed availability planning	90
FTC	29:29	The Hospital Transport Control patient releasing process still takes longer than I need or want it to take.	70
FTC	29:43	I forget to look at the vehicle availability panel and yet don't think return distance when looking at the map	40
FTC	40:35	Selecting multiple patients at once helps me chunk my planning so that the task feels less complex	40
HC	40:51	I probably should try to convince a larger hospital to take a full bus load of patients at once	40
HTC	41:23	To be successful, I need to pay better attention to the behavior of patients in the stairwells	60
FTC	42:50	I am better off focusing on the right half of patient icons than the left as the right is a better identifier of transport needs	58
FTC	44:27	I waited too long to send the Metro Bus out on a patient delivery run	40
HC	46:35	We all could use some more communications about the plan as we are doing are individual actions	70
FTC	49:03	I realize I should be more concerned about how many patients are coming down the stairwells currently	55
HC	50:23	Now that I see the flow of patients and vehicles, I realize I have more time to work on detailed receiving hospital negotiations once I get the first few going	60
FTC	51:17	Creating a tabular presentation of the iconic data helps me organize my thoughts	50
FTC	56:17	I make better decisions when I focus on the orange and pink patient health types	40
FTC	1:00:15	I can track critical patients much better with a visual color coding scheme and bother HTC less as a result	55
FTC	1:04:47	Now that I have a sense of the flow of patients from the hospital, I realize I could have spent more time helping the HTC	60
HC	1:05:46	Tracking the vehicular movement myself helps me make better requests of the receiving hospitals	80
HTC	1:12:45	I get the sense that wheelchair patients take approximately twice as long to get out of the hospital as ambulatory	45

FTC	1:26:03	Once I finish my plan for patient evacuations, I have time to spend pointing out problems with other's potential actions	60
FTC	1:40:52	I am spending time implementing a plan that could be implemented by another process if I could help out with others	40

In addition to the insights listed above, there were forty-nine other insights generated but deemed to be less significant than the above chronological list.

8.3 Distributed Situation Awareness

For both the paper-based and computer-based hospital evacuation session conditions, we used a random number generator to choose ten numbers between 0 and 120, the two hours of our experiment timeline. For the paper-based drill, the generator produced the numbers 2, 119, 94, 103, 47, 21, 89, 51, 35, and 104. For the computer-based drill, the random-number generator produced the numbers 109, 7, 54, 25, 44, 10, 63, 39, 83, and 82. We asked our situation awareness questionnaire at these equivalent minutes from the start of the drill.

The questions were discussed in advance with the subjects to let them agree on a common definition of what the question was asking. As an observer, we listened for common ground with the KCHC's discussion of each question as well. We also interviewed each participant to ascertain that they agreed with the significance of each question to their belief as to what they would consider a successful evacuation.

We recorded the following responses to the five questionnaire questions for both treatments:

1. **How many patients are in a significant state of discomfort currently?**

As a result of a consensus-building discussion, the subjects agreed that any critical patient would be considered experiencing a significant state of discomfort at all times and all assisted critical, unstable, or assisted patients would be deemed experiencing a significant state of discomfort when moving.

In the paper-based session, the hospital transport coordinator had information on all the patient encodings from patient tracking sheets but did not share that information with the hospital control nor transport coordinator roles until deemed relevant. As a result, the awareness unfolded over time, as those patients were part of the individual's role responsibilities. The answers to question one on the freeze-probe questionnaire for the paper-based session follow:

Table 4 – Freeze-probe Questionnaire Answers to Question One by Role for Paper-based Trial

Question Time	Hosp Control	Hosp Transport	Transport	Actual
2	unknown	20	unknown	15
21	8	15	unknown	15
35	10	15	12	15
47	14	25	17	27
51	14	24	22	26
89	30	30	28	32
94	30	29	26	28
103	30	33	30	32
104	30	33	30	32
119	25	25	26	26

In the computer-based session, visual representation of patient-related data was available for the needs of all three roles to all participants. As a result, awareness did not require direct input from other participants, as the data upon which their roles were acted out were available irrespective of any one individual's role responsibilities.

Table 5 – Freeze-probe Questionnaire Answers to Question One by Role for Computer-based Trial

Question Time	Hosp Control	Hosp Transport	Transport	Actual
7	15	15	15	15
10	15	15	15	15
25	15	15	15	15
39	22	21	20	22
44	25	24	22	24
54	24	25	24	24

63	27	26	27	26
82	33	31	31	31
83	33	31	31	31
109	31	30	30	30

A paired t-test was performed to determine if situation awareness was improved in the computer-based trial. The mean number of improvement (M=2.6, SD =3.28, N= 30) was significantly greater than zero, t(29)=4.347, two-tail p = .0002, providing evidence that the situation awareness of discomforted patients is higher. The calculation of the statistics are provided in Appendix III.

2. **Where are these patients located?**

As a result of a consensus-building discussion, the subjects agreed that relevant location information would include their status (lying in a bed, moving in a stairwell, moving from indoors to outdoors, or in transit to receiving hospital), and general location (which floor, which stairwell, which waiting location, and which transport).

We compared their answers to the actual location of patients and recorded the number of locations they got correct for patients that met their *experiencing significant discomfort* definition.

In the paper-based session, the hospital transport coordinator had information on all the patient encodings from patient tracking sheets but did not share that information with the hospital control nor transport coordinator roles until deemed relevant. As a result, the awareness unfolded over time as those patients were part of the individual's role responsibilities. The answers to question one on the freeze-probe questionnaire for the paper-based session follow:

Table 6 – Freeze-probe Questionnaire Answers to Question Two by Role for Paper-based Trial

Question Time	Hosp Control	Hosp Transport	Transport	Actual
2	0	15	0	15
21	0	15	0	15

35	0	15	15	15
47	7	20	22	27
51	4	24	22	26
89	8	30	28	32
94	10	29	26	28
103	12	33	30	32
104	12	33	30	32
119	15	25	26	26

In the computer-based session, visual representation of patient-related data was available for the needs of all three roles to all participants. As a result, awareness did not require direct input from other participants as the data upon which their roles were acted out were available irrespective of any one individual's role responsibilities. The answers to question one on the freeze-probe questionnaire for the computer-based session follow:

Table 7 – Freeze-probe Questionnaire Answers to Question Two by Role for Computer-based Trial

Question Time	Hosp Control	Hosp Transport	Transport	Actual
7	10	15	15	15
10	12	15	15	15
25	15	15	15	15
39	20	20	20	22
44	18	21	23	24
54	16	20	22	24
63	14	18	24	26
82	20	22	30	31
83	22	22	30	31
109	16	18	28	30

A paired t-test was performed to determine if situation awareness was improved in the computer-based trial. The mean number ($M=3.867$, $SD=8.114$, $N=30$) was significantly greater than zero, $t(29)=2.610$, two-tail $p = .0142$, providing evidence that the situation awareness of discomfort patient location is higher. The calculation of the statistics are provided in Appendix III.

3. How many patients are currently in transit between the evacuating and receiving hospital?

The subjects agreed immediately that the appropriate answer to this question is the total number of patients associated with vehicles in transit to or from receiving hospitals.

In the paper-based session, the transport coordinator verbally communicated what he thought was relevant in regards to patient transport allocations and timings. No other role player had transportation included as part of their individual role responsibilities. The answers to question one on the freeze-probe questionnaire for the paper-based session follow:

Table 8 – Freeze-probe Questionnaire Answers to Question Three by Role for Paper-based Trial

Question Time	Hosp Control	Hosp Transport	Transport	Actual
2	0	0	0	0
21	0	0	0	0
35	0	0	0	0
47	8	4	8	8
51	12	4	12	12
89	34	24	34	32
94	32	20	30	28
103	42	26	34	32
104	42	26	34	32
119	26	22	26	26

In the computer-based session, visual representation of patient and vehicle data was available for the needs of all three roles to all participants. As a result, awareness did not require direct input from other participants as the data upon which their roles were acted out were available irrespective of any one individual's role responsibilities. The answers to question one on the freeze-probe questionnaire for the computer-based session follow:

Table 9 – Freeze-probe Questionnaire Answers to Question Three by Role for Computer-based Trial

Question Time	Hosp Control	Hosp Transport	Transport	Actual
7	0	0	0	0
10	0	0	0	0
25	0	0	0	0
39	0	0	0	0

44	6	6	6	6
54	12	12	12	12
63	38	36	36	36
82	42	40	40	38
83	42	38	40	38
109	24	24	24	24

A paired t-test was performed to determine if situation awareness was improved in the computer-based trial. The mean number (M=1.933, SD =32.99, N= 30) was significantly greater than zero, t(29)=3.537, two-tail p = .0014, providing evidence that the situation awareness of in-transit patients is higher. The calculation of the statistics are provided in Appendix III.

4. How much more time will it require to fully evacuate the existing hospital given ideal circumstances?

As a result of a consensus-building discussion, the subjects agreed that the relevant answer to this question was the number of hours remaining to completely get all patients out of the hospital from the current point in time.

Since the role-playing session by design only lasted two hours, we had no actual data for how long it would actually have taken to evacuate the hospital if the subjects continued to evacuate the hospital until the last patient was out of the hospital and ready for transport. But, as we had calibrated the simulation based on KCHC role-play, we were able to run the simulation from each point in time to generate a simulated time to finish up from the current point in time. We compare subject's answers to simulated time in the last four tables in this section.

The answers to question four on the freeze-probe questionnaire for the paper-based session follow:

Table 10 – Freeze-probe Questionnaire Answers to Question Four by Role for Paper-based Trial

Question Time	Hosp Control	Hosp Transport	Transport	Simulated
2	10	12	8	6
21	10	12	8	6
35	11	12	9	7

47	11	12	9	8
51	11	12	9	9
89	9	11	9	8
94	9	11	8	8
103	9	11	8	7
104	9	11	8	7
119	9	10	7	7

The answers to question four on the freeze-probe questionnaire for the computer-based session follow:

Table 11 – Freeze-probe Questionnaire Answers to Question Four by Role for Computer-based Trial

Question Time	Hosp Control	Hosp Transport	Transport	Simulated
7	8	10	8	6
10	8	10	8	6
25	8	10	9	8
39	8	9	9	8
44	7	9	9	9
54	7	9	9	8
63	7	8	8	8
82	7	8	8	7
83	7	8	8	7
109	7	7	7	6

A paired t-test was performed to determine if situation awareness was improved in the computer-based trial. The mean number (M=1.4, SD =1.380, N= 30) was significantly greater than zero, t(29)=5.558, two-tail p = <.0001, providing evidence that the situation awareness of time to complete the scenario is higher. The calculation of the statistics are provided in Appendix III.

5. **How much more time will it require to fully deliver all evacuating patients to their receiving hospital given ideal circumstances?**

As a result of a consensus-building discussion, the subjects agreed that the relevant answer to this question was the number of minutes it would take to deliver all patients to their receiving hospital after all patients had been evacuated from the sending hospital.

The answers to question five on the freeze-probe questionnaire for the paper-based session follow:

Table 12 – Freeze-probe Questionnaire Answers to Question Five by Role for Paper-based Trial

Question Time	Hosp Control	Hosp Transport	Transport	Simulated
2	60	90	45	42
21	60	90	45	42
35	60	90	45	41
47	60	75	45	39
51	60	75	45	36
89	60	60	45	37
94	60	60	45	37
103	60	60	45	39
104	60	60	45	39
119	60	60	45	38

The answers to question five on the freeze-probe questionnaire for the computer-based session follow:

Table 13 – Freeze-probe Questionnaire Answers to Question Five by Role for Computer-based Trial

Question Time	Hosp Control	Hosp Transport	Transport	Simulated
7	60	60	45	42
10	60	60	45	42
25	60	60	45	41
39	60	60	45	40
44	60	60	45	39
54	60	60	45	38
63	60	60	45	37
82	60	60	45	40
83	60	60	45	40
109	60	60	45	40

A paired t-test was performed to determine if situation awareness was improved in the computer-based trial. The mean number (M=3.767, SD =10.040, N= 30) was significantly greater than zero, t(29)=2.055, two-tail p = 0.049, providing evidence that the situation awareness of remaining patient delivery time is higher. The calculation of the statistics are provided in Appendix III.

Chapter 9 – Discussion

We began applying a RimSim architecture to an emergency hospital evacuation response scenario because we were convinced the KCHC could benefit from our participation in their efforts to have first responders gain skills and training through the use of such a scenario. We tested our application for utility using two metrics that were well supported by both the literature and advisement from experts in the field. We would like to be able to conclude that a RimSim architecture-driven approach would be valuable to train responders for a scenario before having to be confronted by such a scenario without training.

Having run our experiments with two different teams (one a team that helped define the roles for a hospital evacuation scenario and another that attempted to train on those roles as defined by the first team), we generated data that supported an analysis of the contribution that a geospatial interface provides a hospital evacuation team when performing a hospital evacuation scenario. Although the experiments with the role designing team were short of an ideal protocol, we can make some broad conclusions about the contribution the visual interface provides that team, based on the role-play sessions captured by our software and observations.

Having learned from the interface and procedural shortcomings identified by the first team, we were able to implement a more complete experimental protocol with a second team of hospital staff and fire department trained personnel as described in the previous chapter. We received their approval in sharing the results of the formal experiment, which were also reported in chapter 8. We now evaluate those results in the context of our doctoral hypotheses.

We then end this chapter with some insights into the quality of the emergency response game play by the teams. We also compare and contrast generalities between the team that spent time generating and refining the roles.

9.1 Evaluating Insight Generation Hypothesis

When comparing insights generated of the paper-based role simulation session and computer-based role simulation session, we look at both the number of insights and the sum of the value of all scores on a 1 to 100 scale arrived at by a team discussion process. In analyzing the data, we noticed a clustering effect of those insights rated less than or equal to ten versus those insights rated more than ten. As a result, we broke those two clusters out separately. We found the following aggregate results:

Table 14 – Insight Comparison Between Paper-based and Computer-based Trials

Trial	Number of Insights	Number valued <=10	Number values >10	Average Score <=10	Average Score >10	Total Score <=10	Total Score >10
1.Paper	52	17	35	7.41	53.97	126	1889
2.Computer	94	49	45	5.51	55.27	270	2487

Looking at both the number of insights and the total score of insights, we found a substantial increase in the number of insights in the computer-based trial over the paper-based one. The average score of those insights scored greater than ten by team consensus is not significantly different between the two trials. The average score of those insights that were deemed less significant, i.e. with a score less than ten, was significantly higher for the paper-based trial. But since there were more computer insights than paper insights, the total score for the computer-based insights is substantially greater.

Because the goals of the computer-based trial were the same as the goals of the paper-based one, we were able to compare the times at which similar insights were made within the simulation session. Thirteen of the insights were deemed identical and took an average of 54 seconds earlier to be made by our computer-based simulation participants. Of these thirteen, four came later in the simulation with the computer-based scenario, with the longest later timestamp being seven minutes and 27 seconds after the insight timestamp in the paper-based scenario. Each of the differences provides potential insight for us into the possible change in role thought-processes brought on by the change of tool used to facilitate role-play.

We also considered which role player first had which insight in each scenario and found three situations where a different role player had the insight first in the computer-based trial versus the paper-based one. All three suggest there was added distributed cognition enabled by our computer-based interface.

In the paper-based drill, at ten minutes and forty-five seconds into the simulation, the Hospital Transport Coordinator realized she could have already processed and released many ambulatory patients from the lowest patient floor. By releasing them earlier, the Metro Bus could have been better utilized by making a trip to a receiving hospital with ambulatory patients earlier than when the bus was finally released to do so. With the addition of a computer-based interface to the simulation, the Fire Transport Coordinator was able to have that insight just four minutes and fifty-five seconds into the simulation. The HTC and the FTC we able to have a tactical conversation that helped the two coordinate patient releases and transport releases in an improved manner.

In the paper-based drill, eighteen minutes and thirty seconds into the simulation, the Hospital Transport Coordinator realized that ICU bed availability was a highly critical data point to monitor in order to improve initial patient release planning. With the addition of a computer-based interface to the simulation, the Fire Transport Coordinator was able to realize the significance of the ICU bed availability numbers only fourteen minutes and twenty-eight seconds into the simulation. As a result, the FTC began a conversation between the HTC, HC, and himself that led to better planning of patient release timings.

These three examples reflect the benefit of having a computer-based interface for role players to consider the conditions of the incidents and resources in a way that they can contribute to overall decision-making through improved distributed cognition. Not only were insights made earlier in the simulation, but they were made by a different participant who had time to assist in role planning in order to improve their contribution and the contribution of a fellow first responder.

9.2 Evaluating Situation Awareness

When comparing insights generated of the paper-based role simulation session and computer-based role simulation session, we report the following aggregate results for the five freeze-probe survey questionnaire questions asked of the Hospital Control role:

Table 15 – Analysis of Hospital Control Insights for Freeze-Probe Questionnaire by Question

Question Number	HC Variance Paper	HC Variance Computer	HC Variance as % Paper	HC Variance as % Computer
1	61	7	24.6	3.0
2	180	70	72.6	30.0
3	26	10	15.3	6.5
4	25	9	34.2	12.3
5	210	201	53.8	50.4

When comparing insights generated of the paper-based role simulation session and computer-based role simulation session, we report the following aggregate results for the five freeze-probe survey questionnaire questions asked of the Hospital Transport Coordinator role:

Table 16 – Analysis of Hospital Transport Coordinator Insights for Freeze-Probe Questionnaire by Question

Question Number	HTC Variance Paper	HTC Variance Computer	HTC Variance as % Paper	HTC Variance as % Computer
1	15	2	6.0	0.8
2	15	47	6.0	2.0
3	44	2	25.9	1.3
4	41	15	56.2	20.6
5	330	210	84.6	50.4

When comparing insights generated of the paper-based role simulation session and computer-based role simulation session, we report the following aggregate results for the five freeze-probe survey questionnaire questions asked of the Fire Transport Coordinator role:

Table 17 – Analysis of Fire Transport Coordinator Insights for Freeze-Probe Questionnaire by Question

Question Number	FTC Variance Paper	FTC Variance Computer	FTC Variance as % Paper	FTC Variance as % Computer
1	51	51	20.6	21.9
2	41	11	16.5	4.7
3	8	4	0.8	0.4
4	10	10	13.7	13.7
5	60	51	15.4	12.8

When comparing insights generated of the paper-based role simulation session and computer-based role simulation session, we report the following average results for all role players who participated in the hospital evacuation sessions:

Table 18 – Analysis of Insights for Freeze-Probe Questionnaire by Question

Question Number	Avg Variance Paper	Avg Variance Computer	Avg Variance as % Paper	Avg Variance as % Computer
1	42.3	20	17.1	8.6
2	78.7	42.7	31.7	12.2
3	26.0	5.3	14.0	2.7
4	25.3	11.3	34.7	15.5
5	200.0	154.0	51.3	37.9

For all five questions, the average variance is lower for the computer-based trial than the paper-based trial. In fact, the variance from actual for the computer-based trial is less than half the variance for the paper-based trial, except for the estimate of residual time that patients would still be in transit after all patients were evacuated from the hospital. Our findings suggest there is improvement in situation awareness, for initial training purposes with our Hartford County team, when using our computer-based simulation interface compared to the paper-based one currently being used by the KCHC.

We performed a paired-t statistical analysis to compare situation awareness levels between paper-based and computer-based trials. We were interested in looking at the potential influence of computer-based interfaces during simulation sessions as provided in Appendix III to pinpoint the likely effect of our computer-based interface and reported results of significance with high p confidence levels.

We calculated a t-value of 4.347 for question one and compared it to the required p value necessary to refute the null hypothesis. Our null hypothesis suggested that the two drills had identical situation awareness for the tracking of patients experiencing significant discomfort. Our t-value of 4.347 was large enough for us to refute the null hypothesis with confidence, with a chance of refuting improperly at .0002.

We calculated a t-value of 2.610 for question two and compared it to the required p value necessary to refute the null hypothesis. Our null hypothesis suggested that the two drills had identical situation awareness for the tracking of patient locations of those experiencing significant discomfort. Our t-value of 2.610 was large enough for us to refute the null hypothesis with confidence, with a chance of refuting improperly at .0142.

We calculated a t-value of 3.537 for question three and compared it to the required p value necessary to refute the null hypothesis. Our null hypothesis suggested that the two drills had identical situation awareness for the tracking of in-transit patients. Our t-value of 3.537 was large enough for us to refute the null hypothesis with confidence, with a chance of refuting improperly at .0014.

We calculated a t-value of 5.558 for question four and compared it to the required p value necessary to refute the null hypothesis. Our null hypothesis suggested that the two drills had identical situation awareness for the amount of time it would take to finish the hospital evacuation. Our t-value of 5.558 was large enough for us to refute the null hypothesis with confidence, with a chance of refuting improperly at less than .0001.

We calculated a t-value of 2.055 for question one and compared it to the required p value necessary to refute the null hypothesis. Our null hypothesis suggested that the two drills had identical situation awareness for the tracking of patients experiencing significant discomfort. Our t-value of 2.055 was large enough for us to refute the null hypothesis with confidence, with a chance of refuting improperly at less than .0490.

As a result, we were able to refute three of the five situation awareness null hypotheses with at least 95% confidence and suggest our participants retained a higher level of situation awareness by using a computer-based interface to evaluate the simulation. Our other two null hypotheses showed some evidence of also being refutable, but not to the level of significance we were willing to be comfortable with given the potential of natural inter-session variability not relevant to situation awareness.

As a result of working with just one team of first responders in evaluating situation awareness, we consider our contribution to be more one of demonstrating a viable and useful approach to considering team situation awareness instead of being able to be definitive in how the approach makes a difference. The approach can be applied to many teams of first responders participating in many first response scenario sessions in order to gather a book of evidence as to how well the use of a simulator helps prepare first responders for their roles in conjunction with others.

9.3 Contribution to Team Preparedness

Our results suggest that our approach to using a geospatial interface for simulated role-play can help a team prepare for working with each other when responding to a community-wide crisis. As we added the interface, we observed at least two indicators that the team was improving distributed cognition so as to overlap in a shared mental model of the activity. These suggested here require more rigorous verification, but at first glance suggest a better preparedness for the team as a whole:

1. Role players had insights from the perspective of other roles as they became familiar with visualizing the causes and effects of other role player actions.

2. The team as a whole appeared to be significantly faster at performing actions with the computer-based interface that would get them to their stated goals faster.

As we build our simulation play interface to provide a different tool for each role player and yet with common shared visual components to foster efficient communication, we offer each team member the opportunity to play other roles to gain first-hand knowledge of how that role considers the joint activity and decides upon actions to take. Consequences of actions can also be evaluated upon running the simulation repeatedly while playing a role. Our challenge is one of building the best-shared components that then trigger memories of other role player perspectives without distracting from the actual role expected of the participant come an actual emergency response event.

9.4 Contribution to Individual Training

We found more insights and better situation awareness for all three of our roles in both the KCHC drill support and the Hartford County team training trials The more often we iterate our design and run simulations using our latest interface designs, the more we can design, develop, and deliver ideal interfaces for training these individual roles. This dissertation has tracked an approach to building emergency response training tools in conjunction with existing emergency response training programs while at the same time making the resultant tools available to any team that wishes to then train using the simulator. With the added step of incorporating role player agents into the mix, we can anticipate providing a software environment where any person anywhere can attempt to play any emergency response role with a scenario that is made accessible. We believe the Web is an ideal medium through which to provide such opportunities.

We imagine a day when community preparedness for emergencies can go beyond just training the first responders to play important emergency response roles and can instead become accessible to all residents in the community who wish to be better prepared for assisting and anticipating emergency response efforts by gaining a cognition as to how they likely will be enacted. Geospatial-temporal visualizations are an opportunity to provide an overall structure in which to think about an emergency response crisis as it relates to all facets of the community. Human pattern recognition can help individuals picture potential states of incidents and resources and how they move about within a community over time when responding to one or more events. Taking on the perspective of an actual role that analysts have scoped and engineered to be performed well by a trained professional is potentially a perspective that lets any citizen makes sense of emergency response crises in general. Adding additional perspectives over time through simulated role-play would perhaps provide opportunity to iterate upon a better mental model in a logical manner.

9.5 Relationship to Hypotheses

We presented our subject experiment results in this chapter in order to support conclusions made about testing our hypotheses. Although our sample sizes are small, we have shown experimental findings that support our insight generation hypothesis for both of our teams of hospital evacuation scenario simulation participants. Participants from the KCHC participated more fully in the design and implementation of the RSR and RSV tools and as a result developed insights into the nature of the scenario they were training on *before* they participated in the two simulation drills. And yet, they identified 44% more insights that they attributed to better performance during the computer-based drill than the paper-based drill. They helped us iterate the design of the RSR tool so that participants would more effectively gain insight through interaction with the interface. The results of the second team of hospital evacuation scenario role-players presented above showed an 80% increase in the number of insights deemed significant to overall team performance. Even if we cluster the specific insights identified into major (scoring more than ten points) and minor (scoring less than or equal to ten points) groups, we find a 29% increase in the number of specific insights generated by the computer-based RSR tool.

The situation awareness two-tailed paired-t test results support the situation awareness hypothesis. For all five critical success factor questions on the freeze probe situation awareness questionnaire, participants scored higher with the computer-based RSR tool than without. We found the results to be significant beyond a 95% confidence level for all five questions.

Chapter 10 - *Future Work*

In our readings, we found that there were many areas of research that required collaboration among experts in order to coalesce a useful shared vision of phenomena that could be explained and applied for human understanding and use. In many cases, an expert with tangential or seemingly environmental knowledge was able to shed additional light, by becoming invited into the existing vision of a group of people. The cybernetics field of the 1950s is one example of a field that brought in one science group after another to shed light on fundamental principles. The chaos and complexity field involved experts from a wide variety of fields and continues today through such organizations as the Santa Fe Institute. The virtual reality field of the 1980s is another field that made progress through engineers, computer scientists, cognitive scientists, sociologists, and others. Most recently, the social networking field has brought together physicists and sociologists. There are many other research domain examples we could cite.

Because research is a human pursuit, we noticed that progress among interdisciplinary teams has most often made significant leaps when a new member joined the community with a passion to explore the existing research with the existing community of researchers who were open to discuss and collaborate together on the research. The enthusiasm of a respected person with a fresh point of view, supported by a valued body of work in a different field, and with a different experience in community building with other researchers, seems invaluable to an existing community. Such a community allows expression for their work in terms that can bridge a gap between the two knowledge bases in which collaboration takes place. This association also allows participants to repackage their own knowledge for better expression within the greater interdisciplinary community, sharing new insights with all parties.

One ripe area for further research is the area of interface and interactivity design. We iterated the interface to our hospital evacuation simulator many, many times through perspectives we received from perception experts, cognition experts, and interaction design experts. But, our feedback from the six simulation session participants who actually used the interfaces to do analysis and interact with the simulation seemed the most immediately valuable. Because community-wide crisis events and response activities generate so much data ripe for visualization, we feel confident much of our time will be spent researching how to improve that aspect of the RimSim architecture.

Another area for future research is the emergent behavior of human groups in crisis. Anticipation of potential human behavior can be a significant part of training for community-wide emergency scenarios based on our readings of the aftermath of hurricane Katrina and our first-person experience of the emotions associated with two significant earthquakes in Seattle just on either side of the year 2000.

A third area research of we could enhance with the RimSim is decision support. We have noted the need for better decision support tools that are available for use during emergencies. Since our existing software already does simulation and provides us the ability to speed up or slow down simulated time, we feel there is a basis on which decision-support tools could be built. Decision-support requires a rigorous approach to development that we believe shows promise in pushing our iteration process towards improving RimSim rapidly. And, decisions of life and death are difficult to make when the human psyche has to deal with the responsibility and accountability for such decisions. Providing decision support tools can widen the ownership of responsibility and accountability of decisions to a wider community – those who create and test the decision support tools.

There are other areas of research we believe RimSim could help support. We believe the RimSim architecture is available to support the kind of collaborative research endeavor that can bring a wider understanding to researchers across fields of research. And, we believe progress in improving cross-field understanding comes from individuals who are committed to participating with their point of view while being open to having that point of view enhanced through the point of view of others. We believe enthusiastic RimSim participants will make the most important contributions by convincing others that sharing an architecture for shared research is valuable and that gaining the help from others to implement one's research openly will be a critical success factor to a RimSim approach.

We also believe RimSim is available to support a wide range of first response communities who wish to apply a simulation approach for better insight into potential community-wide crises of concern in their communities. We've shown that a RimSim approach has potential for training individual first responders in a scenario and in helping a team of first responders train in the collaborative aspects of first response activities. We know that first responders who engage in a RimSim-supported process can improve training potential. In turn, they can co-develop tools that help train better and train their colleagues. We believe RimSim is a terrible approach to force on a first responder or group of first responders who don't want to spend the time to engage in the process of developing a better tool for themselves from which to train. As we saw first-hand when working with the groupware industry in the early 1990s, computer-mediated solutions to support human behavior either succeed or fail through many complex variables that contribute to motivation and organizational support. Perhaps even those variables can begin to be explored via simulation.

Like any new process being attempted by human beings, there is potential to improve RimSim implementations through repetition and modification by expanded exposure of the process to a wider range of scenarios. Creating interfaces for human beings is not an exact science and one person's ideal interface might not be able to become another person's ideal interface. By starting from a defined architecture, we can provide the opportunity to plug-in different components to adapt to different environmental conditions and personal preferences. We can distribute the job of improving the whole simulation process across people who each take a piece of the architecture and implement solutions that improve that piece as part of the whole.

Since it was so difficult to schedule professional emergency response personnel, we were able to work with only two teams, with many variables, in tailoring our skill set to work effectively with trained emergency responders. We found our time with a retired fireman to be extremely useful because he had the time to devote to our endeavor without having emergency response responsibilities active in his mind. He had a perspective gained from years of thinking about emergency response and working in that culture that was invaluable to helping us evolve our point of view to better work with emergency responders. Many of today's retirees have little or no computer literacy or exposure to geospatial-temporal interfaces outside of static paper maps. As younger people retire after having lived a career in emergency response during the digital and information ages, we see our work becoming more feasible to more groups of people without having to interrupt emergency responders during their day-to-day responsibilities.

We must continue to evolve our thinking in terms of all subject matters we reviewed in chapter 2 of this dissertation. Each subject matter sheds some light on how we can improve our process to be more effective. Adding those people trained in evaluation to our team will help us quantify our contribution well beyond the humble attempt by minimally competent graduate students making their first attempt at evaluation.

Chapter 11 - *Summary and Conclusions*

11.1 Tools built

To contribute to the research on simulation environments as a useful platform for first responder training and planning, we built two software-based tools: 1) The RimSim Response multi-user role-play simulation platform enables first responders the opportunity to train on a community-emergency response scenario and consider their role when planning their contribution for potential events, and 2) The RimSim Visualization tool provides an interactive visual analytics platform for post-event sense-making of a First Response Effort. As both tools are built from the same underlying Java-based library modules, component software that supports one of the tools can be integrated into the other with the ease of a typical good software engineering process.

As the literature suggested that real-time situation awareness and asynchronous sense-making could be considered two distinct distributed cognitive activities, we developed the two independently to allow first responders to interact with both tools to improve their emergency response distributed cognition.

First responders trained with the RSR tool through role-playing multiple scenarios with different levels of specificity and detail. Medical logistics team members role-played a resource allocation scenario to train their cognition for a medical supplies caching and real-time hospital allocation task that would be typical of many emergency response scenarios that led to human injury. RSR software stress-testing teams played a first responder transport allocation scenario to train their cognition in allocating police, fire, medical, and other first responder staff to varying incident levels across metro-Seattle, Detroit, Vancouver, BC, and Christchurch, NZ communities. Upon iterating the tool for effectiveness, robustness, and reliability, we pursued a specific community-wide hospital evacuation scenario to provide to a team of medical and emergency response staff interested in hospital evacuation scenario training. Two teams from different communities role-played the hospital control, hospital transportation control, fire transportation control and receiving hospital control roles through the use of the RSR tool.

We provided the two teams that performed the King County, Washington hospital evacuation scenario with the RimSim Visualization tool so they could help us interpret their performance and identify insights that contributed to their overall situation awareness. We made the task of identifying situation awareness after the fact a sense-making exercise that could help us iterate the RSV tool while at the same time help us evaluate subject experiments attempting to test our hypotheses in this thesis. The teams used the RSV to negotiate a consensus evaluation of their performance.

11.2 Research findings

Through the design and development process of building tools to support first responder training and planning, we found that the process consistently provided us with insights into the nature of first responder roles. As a result, we were hopeful the tools would provide the first responders with insights that would improve their performance during RSR role-playing sessions. By bringing the King County Hospital Coalition into the design and development process for the RSR hospital evacuation scenario configuration, we found the process to be insightful for those participants as well. The necessity of making decisions associated with interface design brings up discussions about the affordances necessary to perform emergency response tasks. These discussions drive the first response team to consider all aspects of first response tasks in order to help the software team build visual software components that allow them to perform the tasks during role-play sessions.

We designed formal subject experiments to explore the possibility that the RSR role-play experience would improve distributed situation awareness for those critical success factors that a first responder team asked us to focus on for training purposes. We found the critical success factors they gave us to be reasonable given the critical success factors documented in the various Emergency Operations Centers manuals we explored during emergency response drills we had been invited to during our experiments preparation period. We asked role-play participants to answer questions about the critical success factors through freeze-probe questionnaires that had been used effectively to evaluate team situation awareness in related fields of study. For all five questions on the questionnaire, participants scored higher when using our computer-interface to train their roles than when they did not. Three of the five improvements were strong enough that we were able to state they significantly did better with at least 95% confidence.

We let the role-play participants evaluate their performance asynchronously after they finished interacting with the emergency response crisis simulation. They told us they felt they had done better using the interface but we wanted to observe them as they worked as a team to discuss their performance using a visual analytics tool. We guided their analysis by suggesting that they review all communications and decisions made during the simulation session to identify where they had insights and score those insights as to the value each insight had to the overall performance of the hospital evacuation activity. We asked that the team reach a consensus by level-setting the contribution of each participant to the benefit of the overall performance of the team. We observed the team perform the insight-scoring task intelligently with impressive negotiation skills. We also observed that the team scored higher on the simulation session with our RSR interface than the simulation session without.

11.3 Hypothesis proven

Although we believe our contribution goes beyond the testing of two specific hypotheses, we attempted to frame our research through testing two hypotheses:

1. *A multi-user situational simulation environment can be effectively used as a training tool for generating insight among emergency response personnel.*
2. *A multi-user situational simulation environment can be effectively used as a training tool for improving situation awareness among emergency response personnel.*

We proved the first hypothesis through post-simulation evaluation of two simulation sessions whereby both generated significant insights even after the RSR tool iterative development process had generated insight among the first responders who contributed to tool design. We found the same insight generation process by running the simulation sessions with first responders who did not participate in RSR tool generation.

We proved the second hypothesis through performing an experiment with and without providing a multi-user situational simulation environment for use by simulation participants. The participants demonstrated significantly higher situation awareness when using our RSR tool than when not.

11.4 Contributions made

Our process of defining system architecture to support first responder training through simulation role-play was one of daily iteration, test, discussion, and reflection. Such a process is expected to be rocky at times as sometimes iteration takes us backwards away from our goals temporarily and not always just to get us out of a local maximum. Our process became a nimble one so that we could backtrack to any previous development point where we felt progress could be better attained from that point onwards. Because we wanted to build a software framework that could support simulation for a wide variety of community-wide crisis scenarios, we deemed robustness and effectiveness critical success factors to be the most important yardstick by which we evaluated our software. We believe we have made a contribution through documenting the process by which we designed and development the RSR and RSV tools.

We believe our process is one we can share with a wide variety of researchers from a wide variety of research backgrounds. The RimSim architecture lets researchers pick and choose where to make a contribution. We have used an object-oriented, open source, programming model to allow computer scientists to write code that will enhance any part of the simulation support services that let first responders gain insight into community-wide crisis response. Like our process of helping a linear programming modeler implement a resource optimization model into the heuristics support module of the RimSim system architecture, a community of computer scientists can help others implement and test their hypotheses from within the simulation framework of RimSim. Each time that integration process happens, new light is shed on the effectiveness of the architecture from both a code engineering perspective and from a first responder training tool perspective (among others). Our suggested future work becomes much more evident as we see that process unfold, bringing in one enthusiastic researcher at a time.

In order to evaluate distributed situation awareness and insight generation, we considered a wide variety of questionnaires, probes, and quantitative metrics that could shed light on the usefulness of our tools. We offer our literature search, summary of findings, and implementation of chosen yardsticks to the community for consideration of evaluating team training and planning tools. Just as we benefitted by the sharing of approaches by researchers before us, we wish to benefit others who follow us into this area of research.

REFERENCES

1] Kirschenbaum, A. (1994), Measuring the Effectiveness of Disaster Management Organizations, International Journal of Mass Emergencies and Disasters, Vol. 22, No. 1 pp. 75-102.

[2] Bailey, S. (2005) The Expectation Gap, Presentation to the UW-Microsoft Partnership Studying Emergency Response in Seattle Region, Bellevue, WA.

[3] NVAC (2004), Illuminating the Path: The Research and Development Agenda for Visual Analytics, Dept. of Homeland Security Publications.

[4] Clayton, J. D. (1985). *Volume 3, Triumph and Disaster 1945–1964*. Boston: Houghton Mifflin.

[5] Endsley, M. R. (2000), Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), Situation awareness analysis and measurement. Mahwah, NJ

[6] Magnus, P.D., Distributed Cognition and the Task of Science, Social Studies of Science, 37(2): 297–310.

[7] Ware, C. (2003), Perception for Design, 2ndEdition, Elsevier Publishing.

[8] Mass, C (2008), The Weather of the Pacific Northwest, University of Washington Press, Seattle, WA.

[9] Cohen, S. and Roussel, J. (2004), Strategic Supply Chain Management, Elsevier Publishing.

[10] Rojas, E. & Mukherjee A. (2005) "A General Purpose Situational Simulation Environment for Construction Education," Journal of Construction Engineering and Management, ASCE, 131(3) pp 319-329.

Chapter 2

[11] Heer, J. and Agrawala, M. (2006) Software Design Patterns for Information Visualization, In IEEE Transactions on Visualization and Computer Graphics, September-October 2006 (Vol. 12, No. 5) pp. 853-860

[12] Klein, G.A. (1998) Sources of Power: How People Make Decisions. MIT Press: Cambridge, Massachusetts

[13] Clark, A. (2007) Embodiment and the sciences of mind, Cognitive & Linguistic Sciences 20th Anniversary Public Lecture Series, Brown University, 7:30-9:00pm, February 5, 2007.

[14] Rensink, R. (2002). Change Detection. Annual Review Psychology. 53:245–77
[15] Pringle, H. L., Kramer, A. F., & Irwin, D. E. (2004). Individual differences in the visual representation of scenes. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children*. Westport, CT: Greenwood/Praeger.

[16] Hutchins, E. (1996) Cognition in the Wild. MIT Press: Cambridge, Massachusetts

[17] Rogers, Y. (2006) Distributed Cognition and Communication. In The Encyclopedia of Language and Linguistics 2nd Edition. Edited by Keith Brown Elsevier: Oxford. 181-202

[18] Dawkins, R. (1976) The Selfish Gene. Oxford University Press

[19] Pea, R. (1993), A Heuristic Framework for Raising Empirical Questions on Distributed Intelligence – Artifacts as Cognitive Offloading, In G. Salomon, (Ed.), Distributed Cognition: Psychological and educational considerations. Cambridge University Press.

[20] Cole, M. and Engestrom, Y. (1993) A cultural-historical approach to distributed cognition. In G. Salomon, (Ed.), Distributed Cognition: Psychological and educational considerations. Cambridge University Press.

[21] Wundt, W. (1916). Elements of folk psychology. Outlines of a Psychological History of the Development of Mankind. Authorized Translation by Edward Leroy Schaub. New York: Macmillan

[22] Leontiev, A.N., The Development of Mind, a reproduction of the Progress Publishers 1981 edition, plus "Activity and Consciousness," originally published by Progress Publishers, 1977, published by Erythrospress

[23] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors 37(1), 32-64.

[24] Nullmeyer, R. T., Stella, D., Montijo, G. A., & Harden, S. W. (2005). Human factors in Air Force flight mishaps: Implications for change. Proceedings of the 27th Annual Interservice/Industry Training, Simulation, and Education Conference (paper no. 2260). Arlington, VA: National Training Systems Association.

[25] Blandford, A. & Wong, W. (2004). Situation awareness in emergency medical dispatch. International Journal of Human-Computer Studies, volume 61, pp. 421-452.

[26] Smart, P. R., Russell, A., Shadbolt, N. R., Schraefel, M. C. and Carr, L. A. (2007) AKTiveSA: A Technical Demonstrator System for Enhanced Situation Awareness in Military Operations Other Than War. The Computer Journal, 50 (6). pp. 703-716.

[27] Flin, R. & O'Connor, P. (2001). Applying crew resource management in offshore oil platforms. In E. Salas, C.A. Bowers, & E. Edens (Eds.), Improving teamwork in organization: Applications of resource management training (pp. 217-233). Hillsdale, NJ: Erlbaum.

[28] Sarter, N. B. & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. International Journal of Aviation Psychology, 1, 45-57.

[29] Fracker, M. L. (1991). Measures of situation awareness: Review and future directions. Wright-Patterson Air Force Base, OH: Armstrong Laboratories.

[30] Dominguez, C., Vidulich, M., Vogel, E. & McMillan, G. (1994). Situation awareness: Papers and annotated bibliography. Armstrong Laboratory, Human System Center, ref. AL/CF-TR-1994-0085.

[31] Smith, K., & Hancock, P. A., (1995). Situation awareness is adaptive, externally directed consciousness. Human Factors, 37 (1), 137-148.

[32] Moray, N. (2004). Ou sont les neiges d'antan? In D. A. Vincenzi, M. Mouloua & P. A. Hancock (Eds), Human performance, situation awareness and automation: Current research and trends (pp. 1-31). Mahwah: LEA.

[33] Jeannot, E., Kelly, C. & Thompson, D. (2003). The development of situation awareness measures in ATM systems. Brussels: Eurocontrol.

[34] Endsley, M. R. (2004). Situation awareness: Progress and directions. In S. Banbury & S. Tremblay (Eds.), A cognitive approach to situation awareness: Theory, measurement and application (pp. 317-341). Aldershot, UK: Ashgate Publishing.

[35] Endsley, M. R., & Jones, W. M. (2001). A model of inter- and intrateam situation awareness: Implications for design, training and measurement. In M. McNeese, E. Salas & M. Endsley (Eds.), New trends in cooperative activities: Understanding system dynamics in complex environments. Santa Monica, CA: Human Factors and Ergonomics Society.

[36] Endsley, M. R. & Garland, D. J. (Eds.) (2000). Situation awareness analysis and measurement. Mahwah, NJ: Lawrence Erlbaum Associates.

[37] Jones, D. G. & Endsley, M. R. (2000). Examining the validity of real-time probes as a metric of situation awareness. Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society. Santa Monica, CA: Human Factors and Ergonomics Society.

[38] Strater, L. D., Endsley, M. R., Pleban, R. J., & Matthews, M. D. (2001). Measures of platoon leader situation awareness in virtual decision making exercises (No. Research Report 1770). Alexandria, VA: Army Research Institute.

[39] Taylor, R. M. (1989). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations, CP478. Seuilly-sur Seine: NATO AGARD.

[40] Endsley, M. R. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting (pp. 82-86). Santa Monica, CA: The Human Factors and Ergonomics Society.

[41] Matthews, M. D., Pleban, R. J., Endsley, M. R., & Strater, L. G. (2000). Measures of infantry situation awareness for a virtual MOUT environment. Proceedings of the Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium. Savannah, GA: SA Technologies, Inc.

[42] Wilson, G. F. (2000). Strategies for psychophysiological assessment of situation awareness. In M. R. Endsley & D. J. Garland, (Eds.), Situation awareness analysis and measurement (pp. 175-188). Mahwah, NJ: Lawrence Erlbaum Associates.

[43] Garland, D.J., Hopkin, V.D., Wise, J.A. (1999). *Handbook of Aviation Human Factors*, Mahwah, NJ: Lawrence Erlbaum Associates

[44] Salmon, P., Stanton N., Walker, G., & Green, D. (2006) "Situation awareness measurement: A review of applicability for C4i environments", *Applied Ergonomics* 37:225–238

[45] Endsley, M.R., 1995b. Measurement of situation awareness in dynamic systems. *Human Factors* 37:65–84.

[46] Taylor, R.M., (1990). "Situational awareness rating technique (SART): the development of a tool for aircrew systems design". In: *Situational Awareness in Aerospace Operations*. NATO-AGARD, Neuilly Sur Seine, France, pp. 3/1 –3/17.

[47] Endsley, M.R. (1990). Predictive utility of an objective measure of situation awareness. In: *Proceedings of the Human Factors Society 34th Annual Meeting*, Human Factors Society, Santa Monica, CA, pp. 41–45.

[48] Durso, F.T., Hackworth, C.A., Truitt, T., Crutchfield, J., Manning, C.A., (1998). Situation awareness as a predictor of performance in en route air traffic controllers. *Air Traffic Quarterly*, 6:1–20.

[49] Jeannott, E., Kelly, C., Thompson, D., (2003). "The development of situation awareness measures in ATM systems". *EATMP Report*, HRS/HSP-005-REP-01.

[50] McGuiness, B. and Ebbage, L. (2002) "Assessing Human Factors in Command and

Control: Workload and Situational Awareness Metrics". A *BAE Systems Advanced Technology Centre Report*, Bristol, UK.

[51] McGuinness, B., Foy, L., (2000). "A subjective measure of SA: the Crew Awareness Rating Scale (CARS)". Presented at the *Human Performance, Situational Awareness and Automation Conference*, Savannah, Georgia, 16–19 October 2000.

[52] McGuinness, B., (2004). "Quantitative analysis of situational awareness (QUASA): applying signal detection theory to true/false probes and self-ratings". *Command and Control Research and Technology Symposium: The Power of Information Age Concepts and Technologies*, San Diego, California, 15–17 June 2004.

[53] Waag, W.L., Houck, M.R., (1994). "Tools for assessing situational awareness in an operational fighter environment". *Aviation Space Environ. Med.* 65 (5): A13–A19.

[54] Matthews, M.D., Beal, S.A., (2002). "Assessing situation awareness in field training exercises". *Research Report 1795*, U.S. Army Research Institute for the Behavioral and Social Sciences.

[55] Smolensky, M.W., (1993). "Toward the physiological measurement of situation awareness: the case for eye movement measurements". In: *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, Santa Monica, Human Factors and Ergonomics Society.

[56] Endsley, M. R., and Robertson, M., (1996). "Team situation awareness in aviation maintenance," In Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA, pp. 1077-108.

[57] Endsley, M. R., and Jones, W. M. (1997), "Situation awareness, information warfare & information dominance," *Endsley Consulting, Belmont, MA, Technical Report* 97-01.

[58] Prince, C., and Salas, E. (1993), "Training and research for teamwork in the military aircrew," In *Cockpit Resource Management*. E. Wiener, B. Kannki and R. Helmreich, editors, Academic Press, San Diego, CA, pp. 337-365.

[59] Uruguay, A.L.P, and Hirata, C.M. (2009)."Shannon Entropy as a Metric of Situational Awareness in C2 Structures". In the *Proceedings of the International Command and Control Research and Technology Symposium*. June 15-17 Washington, DC.

[60] A. R. Pritchett, R.J. Hansman & E.N. Johnson (1996), Use of Testable Responses For Performance-Based Measurement of Situation Awareness, in the *Proceedings of the International Conference on Experimental Analysis and Measurement of Situation Awareness*, Daytona Beach, FL

[61] Johnson, E.N. & Hansman, R.J. (1995) "Multi-Agent Flight Simulation with Robust Situation Generation" MIT Aeronautical Systems Laboratory Report ASL-95-2

[62] French, H.T., Clark, E., Pomeroy, D. Seymour, M. , & Clarke, C.R. (2007). Psycho-physiological Measures of Situation Awareness. In M. Cook, J. Noyes & Y. Masakowski (eds.), *Decision Making in Complex Environments*. London: Ashgate Publishing.

[63] Wilson, G. F. (2000). Strategies for psychophysiological assessment of situation awareness. In M. R. Endsley & D. J. Garland, (Eds.), Situation awareness analysis and measurement (pp. 175-188). Mahwah, NJ: Lawrence Erlbaum Associates.

[64] Barfield, W. & Weghorst, S. (1993). The Sense of Presence Within virtual environments: a conceptual framework. In *Human Computer Interaction: Software and Hardware Interfaces*. UK:Elsevier Science Publishers.

[65] Durso, F. T., Truitt, T. R., Hackworth, C. A., Crutchfield, J. M., Nikolic, D., Moertl, P. M., Ohrt, D., & Manning, C. A. (1995). Expertise and chess: A pilot study comparing situation awareness methodologies. In D.J. Garland & M.R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 295-303). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

[66] Harwood, K., Barnett, B., & Wickens, C.D. (1988). Situational awareness: A conceptual and methodological framework. In F.E. McIntire (Ed.), *Proceedings of the 11th Biennial Psychology in the Department of Defense Symposium* (pp. 23-27). Colorado Springs, CO: U.S. Air Force Academy.

[67] Campbell, B. (1990). A Fire Insurance Underwriting System for The Travelers Insurance Company, *Report and Software to the Travelers Insurance Company*, Hartford, CT.

[68] Jackson, P. (1999). Introduction to Expert Systems, 3rd Edition, Harlow, England: Addison Wesley Longman.

[69] Billinghurst, M., Savage-Carmona, J., Oppenheimer, P. and Edmond, C. (1995). The Expert Surgical Assistant: An Intelligent Virtual Environment with Multimodal Input. In Weghorst, S., Sieberg, H.B. and Morgan, K.S. *Proceedings of Medicine Meets Virtual Reality IV*, pp. 590-607.

[70] U. Cortes, M. Sanchez-Marre, L. Ceccaroni, I. R-Roda and M. Poch, Artificial intelligence and environmental decision support systems, Applied Intelligence, Vol. 13, 2000, pp. 225-239.

[71] Amparo Alonso-Betanzos, Oscar Fontenla-Romeroa, Bertha Guijarro-Berdiñasa, Elena Hernández-Pereiraa, María Inmaculada Paz Andradeb, Eulogio Jiménezc, Jose Luis Legido Sotod and Tarsy Carballase, An intelligent system for forest fire risk prediction and fire fighting management in Galicia, Expert Systems with Applications

Volume 25, Issue 4, November 2003, Pages 545-554.

[72] Chi, S., Lim, Y., Lee, J., Lee, J., Hwang, S., Song, B. (2003), Proceedings of the Italian Association for Artificial Intelligence Congress No. 8, Pisa, Italy (23/09/2003), vol. 2829, pp. 487-498

[73] Su, K., Hwang, S., and Wu, C. (2007), Developing a Usable Mobile Expert Support System for Emergency Response Center, IAENG International Journal of Computer Science, 32:4, IJCS_32_4_18

[74] Humpfrey, S. (1991), An expert system for improving nuclear emergency response, doctoral dissertation, Rensselaer Polytechnic Institute, Troy, NY.

[75] Moore, D. (1998), Expert systems in emergency response (1998), Proceedings of the Process Safety Symposium, Texas A&M University, College Station, TX. March 30, 1998

[76] T. Guimaraes, Y. Yoon, and A. Clevenson, Factors important to expert systems success: A field test, Information & Management, Vol. 30, 1996, pp. 119-130.

[77] M. Wojtek, R. Steven, S. Roman and W. Szymon, Mobile clinical support system for pediatric emergencies, Decision Support Systems, Vol. 36, 2003, pp. 161-176.

[78] Chan, C. C. H. (2007), Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer, Expert Systems with Applications, Volume 34, Issue 4, May 2008, Pages 2754-2762

[79] D. Ruiz, J. Canton and J.M. Nougues (2001), On-line fault diagnosis system support for reactive scheduling in multipurpose batch chemical plants, Computers and Chemical Engineering, Vol. 25, 2001, pp. 829-837.

[80] Tomasello, M. (2000), The Cultural Origins of Human Cognition, Cambridge, MA: Harvard University Press, 272 pages.

[81] Gobet, F., Lane, P.C.R., Croker, S., Cheng, P.C.H., Jones, G., Oliver, I., & Pine, J.M. (2001). *Chunking mechanisms in human learning*. Trends in Cognitive Sciences, 5, 236-243.

[82] Varela, Francisco J., Thompson, Evan T., and Rosch, Eleanor. (1992). The Embodied Mind: Cognitive Science and Human Experience. Cambridge, MA: The MIT Press. ISBN

[83] Gibson, J.J. (1986) The Ecological Approach to Visual Perception. Lawrence Erlbaum Associates: Hillsdale, NJ.

[84] Russell, D.M., Stefik, M.J., Pirolli, P., & Card, S.K. (1993), The cost structure of sensemaking, Proceedings of INTERCHI, April 1993, Amsterdam, Netherlands (ACM Press).

[85] Klein, G., Moon, B., and Hoffman, R. (2006), Making Sense of Sensemaking 1: Alternative Perspectives, Intelligent Systems,Vol. 21, No. 4, July/August 2006.

[86] Wright, P., Fields, R. & Harrison, M. (2000). Analyzing Human-Computer Interaction As Distributed
Cognition: The Resources Model. Human Computer Interaction, 51(1), 1-41.

[87] Bertin, J. (1981). Graphics and Graphic Information Processing. Walter De Gruyer, Inc.: Berlin

[88] Cleveland, W., and McGill, M. (1984). Graphical Perception: Theory, Experimentation and Application to the Development of Graphical Methods. Journal of the American Statistical Association, 79(387): 531-554.

[89] Wilkinson, L. (1999), The Grammar of Graphics, Springer–Verlag

[90] MacEachren, A.M., (2004), How Maps Work: Representation, Visualization, and Design, The Guilford Press

[91] Shneiderman, B. (1994), Dynamic queries for visual information seeking, IEEE Software, 11(6):70–77.

[92] Tufte, E. (1983), The Visual Display of Quantitative Information, Graphics Press, Cheshire, CT.

[93] Tufte, E. (1990), Envisioning Information, Graphics Press, Cheshire, CT.

[94] Tufte, E. (1997), Visual Explanations: Images and Quantities, Evidence and Narrative, Graphics Press, Cheshire, CT.

[95] Treisman, A., and Davies, A. (1973), Divided Attention to Ear and Eye, In Attention and Performance IV, S Kornblum (Ed.). New York and London: Academic Press, Pp. 101-117

[96] Gibson, JJ. (1966), The Senses Considered as Perceptual Systems, Greenwood Publishing Group.

[97] VanRulllen, R., Koch, C. (2003), Visual Selective Behavior Can Be Triggered by a Feed-Forward Process, Journal of Cognitive Neuroscience, 15:209-217, MIT Press, Cambridge, MA.

[98] Weaver, C. (2004), Building Highly-Coordinated Visualizations In Improvise, In Proceedings of the IEEE Symposium on Information Visualization, Austin, TX.

[99] Heer, J., Card, S.K., Landay, A. (2005), Prefuse: A Toolkit for Interactive Information Visualization, In ACM Human Factors in Computing Systems (CHI), 421-430

[100] Schroeder, W., Martin, K., Lorensen, B. (2006), The Visualization Toolkit: An Object-Oriented Approach To 3D Graphics (4th Edition), Prentice Hall, Upper Saddle River, N.J.

[101] The Numerical Algorithms Group (2006), IRIS Explorer User's Guide, Leeds, UK.

[102] Hill, R., Levenhagen, M. (1995), Metaphors and mental models: sensemaking and sensegiving in innovative and entrepreneurial activities, Journal of Management, Nov-Dec, 1995.

[103] MacEachren, A.M. and Kraak, M.J. (1997), Exploratory cartographic visualization: advancing the agenda. *Computers & Geosciences*, 23(4), pp. 335-343.

[104] MacEachren, A.M. (2004), Geovisualization for knowledge construction and decision support. *IEEE computer graphics and applications*, 24(1), pp.13-17.

[105] Jiang, B., Huang, B., and Vasek, V. (2003), Geovisualisation for Planning Support Systems. *In Planning Support Systems in Practice, Geertman, S., and Stillwell*, J. (Eds.). Berlin: Springer.

[106] Watters, M. 2005. Geovisualization: an Example from the Catholme Ceremonial Complex. *Archaeological Prospection*, 13, pp. 282-290.

[107] Andrienko, G., Andrienko, N., Jankowski, P, Keim, D., Kraak, M.-J., MacEachren, A.M., and Wrobel, S. (2007), Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, 21(8), pp. 839-857.

[108] Danada, J., Dias, E., Romao, T., Correia, N., Trabuco, A., Santos, C., Serpa, J., Costa, M., Camara, A. (2005), Mobile Environmental Visualization. *The Cartographic Journal*, 42(1), pp. 61-68.

[109] Google, Inc. (2008), *Google Earth: Explore, Search, and Discover*, http://earth.google.com/ (accessed November 15, 2008).

[110] Microsoft, Inc (2008), *Virtual Earth*, http://www.microsoft.com/virtualearth/ (accessed November 15, 2008).

[111] National Aeronautics and Space Administration (2008), *Learning Technologies: World Wind*, http://worldwind.arc.nasa.gov/ (accessed November 15, 2008).

[112] Card, S. (2007). *Lifetime achievement award acceptance presentation*. Visualization Conference 2007 (Sacramento, CA).

Chapter 3

[113] Winn, W.D. (2002). Learning in Artificial Environments: Embodiment, Embeddedness and Dynamic Adaptation, *Cognition and Learning* Vol 1, 2002

[114] Novak, J. (1993), Human Constructivism: A unification of psychological and epistemological phenomena in meaning making, *International Journal of Personal Construct Psychology*, 6, pp167-193

[115] Rojas, E. & Mukherjee, A. (2006), A Multi-Agent Framework for General Purpose Situational Simulations in the Construction Management Domain, *Journal of Computing in Civil Engineering*, ASCE, 20(6) pp 1-12

[116] Mukherjee, A. (2005) A Multi-Agent Framework for General Purpose Situational Simulations in Construction Management, *Doctoral Dissertation*, University of Washington, Seattle, WA

[117] Ritchey, J. (2010) *PRISM: Puget Sound Regional Synthesis Model*, http://www.prism.washington.edu/ (accessed September 17, 2010)

Chapter 4

[118] Campbell, B., Mete, O., Furness, T., Weghorst, S., Zabinsky, Z. (2008), Emergency Response Planning and Training through Interactive Simulation and Visualization with Decision Support, In *Proceedings for the 2008 IEEE International Conference on Technologies for Homeland Security*.

[119] MacEachren, A.M., (2010) *GeoVista Center: A Center for Innovation Across the GIScience Research Landscape*, http://www.geovista.psu.edu/ (accessed June 22, 2010).

Chapter 8

[120] Chang, R., Ziemkiewicz, C., Green, T.M., and Ribarsky, W. (2009).“Defining Insight for Visual Analytics”. In *IEEE Computer Graphics and Applications*. March-April 2009.

[121] Johnson, C.W. (2006) Using Computer Simulations to Support A Risk-Based Approach For Hospital Evacuation, *A Department of Computing Science Briefing*, University of Glasgow, Glasgow, United Kingdom