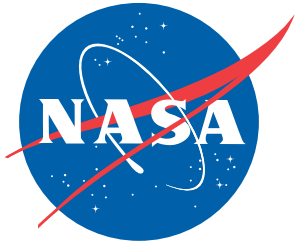# Selected Papers and Presentations Presented at MODSIM World 2010 Conference & Expo

Edited by

Thomas E. Pinelli
Langley Research Center, Hampton, Virginia

March 2011

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at *http://www.sti.nasa.gov*

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at 443-757-5803

- Phone the NASA STI Help Desk at 443-757-5802

- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076-1320

NASA/TM-2011-217069/Part 2
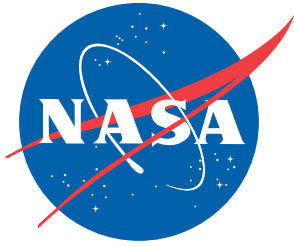
Selected Papers and Presentations Presented at MODSIM World 2010 Conference & Expo

Edited by

Thomas E. Pinelli
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

March 2011

Preface

Modeling and simulation-based (MODSIM) engineering and science is rapidly becoming an essential scientific methodology for nearly all areas of engineering and many branches of science and for research, development, concept generation, product design and manufacturing, and consumer marketing. Continuing advances in computational science and networking technologies have made MODSIM-based engineering and science a "powerful and ubiquitous tool" for engineers and scientists and have made it possible to extend the range, depth, and applications of MODSIM vastly, especially when the phenomena being investigated are not observable or measurements are impractical or too expensive. According to a National Science Foundation Blue-Ribbon Panel, MODSIM-based engineering and science (1) is an equal and indispensable partner, along with theory and experiment, in the quest for enhanced technological innovation; (2) holds great promise for the pervasive advancement of knowledge and understanding through discovery; (3) is indispensable to the nation's continued leadership in innovation and economic global competitiveness; and (4) is "key" to advances in a variety of fields − biomedicine, manufacturing, systems engineering, nanotechnology, health care, atmospheric and climate science, energy and environmental sciences, advanced materials, and product development. MODSIM-based engineering and science is also essential to the success of NASA's research, missions, and projects.

The MODSIM World Conference and Expo began in 2007 when the Hampton Roads Partnership's, Center for Public/Private Partnership (CP3) saw the need to share information about and interests in the vast amount of MODSIM-based research and development occurring in the Hampton Roads region of Virginia. Because of the synergy created by the efforts of Joint Forces Command; Virginia Modeling, Analysis and Simulation Center (VMASC); Eastern Virginia Medical School (EVMS), the NASA Langley Research Center (LaRC), etc., it became obvious to the CP3 membership that there was a need to establish an "interdisciplinary" forum for sharing of MODSIM knowledge and achievement. Their efforts created MODSIM World Conference & Expo. The MODSIM Conference & Expo is now in its fourth cycle.

MODSIM World 2010 was held in Hampton, Virginia, October 13−15, 2010. The theme of the 2010 conference & expo was "21st Century Decision-Making: The Art of Modeling& Simulation". The conference program consisted of seven technical tracks – Defense, Engineering and Science, Health & Medicine, Homeland Security & First Responders, The Human Dimension, K-20 STEM Education, and Serious Games & Virtual Worlds. Selected papers and presentations from MODSIM World 2010 Conference & Expo are contained in this NASA Conference Publication (CP). Section 8.0 of this CP contains papers from MODSIM World 2009 Conference & Expo that were unavailable at the time of publication of NASA/CP-2010-216205 *Selected Papers Presented at MODSIM World 2009 Conference and Expo*, March 2010.

As a condition for inclusion in the conference proceedings, the first author was responsible for securing/obtaining all permissions associated with the general release and public availability of the paper/presentation. Further, the first authors also had to grant NASA the right to include their work in the NASA CP. There are 62 papers and 41 presentations in this NASA CP. There are two appendices in this publication. Appendix A contains the names and affiliations of the conference organizers. Appendix B includes a description of the technical tracks and the names

of the individuals who organized each track. Preparing the proceedings of this conference required the collaborative efforts of many individuals.

The preparation of NASA/CP-2011-217069 would not have been possible without the expertise and contributions of Ms. Leanna "Dee" Bullock (ATK Space Systems, INC.) and Ms. Jennifer McNamara (BreakAway, LTD.).

Table of Contents

4.0 HOMELAND SECURITY & FIRST RESPONDERS TRACK

4.1   Examining Passenger Flow Choke Points at Airports Using Discrete Event Simulation

Examining Passenger Flow Choke Points at Airports Using Discrete Event Simulation

Jeremy R. Brown & Poornima Madhavan
Old Dominion University
jrbrown@odu.edu   pmadhava@odu.edu

Abstract. The movement of passengers through an airport quickly, safely, and efficiently is the main function of the various checkpoints (check-in, security, etc) found in airports. Human error combined with other breakdowns in the complex system of the airport can disrupt passenger flow through the airport leading to lengthy waiting times, missing luggage and missed flights. In this paper we present a model of passenger flow through an airport using discrete event simulation that will provide a closer look into the possible reasons for breakdowns and their implications for passenger flow. The simulation is based on data collected at Norfolk International Airport (ORF). The primary goal of this simulation is to present ways to optimize the work force to keep passenger flow smooth even during peak travel times and for emergency preparedness at ORF in case of adverse events. In this simulation we ran three different scenarios: real world, increased check-in stations, and multiple waiting lines. Increased check-in stations increased waiting time and instantaneous utilization, while the multiple waiting lines decreased both the waiting time and instantaneous utilization. This simulation was able to show how different changes affected the passenger flow through the airport.

1.0 INTRODUCTION

At the turn of the millennium, erroneous information typed into a central database at Hong Kong's $20 billion Chek Lap Kok airport triggered a domino effect that sent the new facility into almost comic confusion: flights taking off without luggage, airport officials tracking flights with plastic pieces on a magnetic board, and airlines calling confused ground staff on cellular phones to say where even more confused passengers could find their planes. Similar scenes were played out at Malaysia's $2.2 billion Kuala Lumpur International Airport, where stranded cargo translated quickly in the tropical heat into rotting refuse. Such examples drive home one of the oldest rules of computer programming, the simple postulate that a machine is only as good as the humans using it.

Clearly airports are very complex environments in which passengers are the consumers and efficiency is the key to organizing the complexity. Airports can be thought of as systems with many parts that need to work together in order to accomplish a task. This task is to get passengers through the airport and onto waiting airplanes. This system can break down when problems occur. Therefore, modeling of processes to optimize traffic flow is where emergency planning can come into play. Some of the integral components of an airport are infrastructure features such as buildings, passenger ground transport systems, runways, taxiways, and vehicles (needed for getting baggage, fuel, and food onto the planes).

Additional features of the system are the computer systems such as baggage check computers and x-ray baggage machines. The final link in the airport system is the human component, i.e., workers that operate the machinery and computers.

1.1 Role of the runway

The runway plays an important part in regulating traffic flow by allowing aircraft to land and take off safely. Taxiways serve the same purpose, although they are primarily used to get the planes from the runway to the terminal. The bigger the aircraft, the longer and tougher the runway and taxiways must be to handle the weight.

1.2 Role of computer systems

The computer systems in an airport are important to the flow of traffic in that they help keep track of all the flights coming and going, as well as the flow of passengers and their baggage. In addition, computer systems play an important role in airport security by screening luggage, and profiling passengers using video cameras.

1.3 The human component

The presence of humans is integral to the running of all the above components. The workers that operate the systems are an important factor to take into account when looking at the airport as a system of systems. It is the humans that make the decisions, and keep the other systems working. A significant proportion of errors in these systems, therefore, are due to incidences of human error. This raises the importance of modeling human behavior to better understand the behavioral

implications on traffic flow in a large system of systems such as an airport.

A few attempts have been made so far to quantify and model passenger flow in various contexts ranging from train station platforms to elevators of tall buildings [1,2]. Nahke created a simulation of Hartsfield Atlanta International Airport's passenger movement system which consisted of nine trains moving passengers from terminal to terminal [3]. Through the use of this simulation, Nahke were able to see what effects increasing the number of trains had to try and increase passenger capacity. They were able to show that through small changes the train system that was designed for a maximum of nine trains could easily handle ten trains, increasing passenger capacity [3]. Ke, Zizheng, and Liling used simulation to optimize bus schedules during peak times [4]. Wusheng and Qian created a simulation using queuing theory to examine passenger flow at the curbside of an airport [5]. Another study examined the flow of traffic in an airport through simulation and modeling in a similar way to what is being proposed [6]. Although this study effectively examined the problem of passenger flow from the standpoint of scheduling, the model ignored the degree of heterogeneity among the passengers themselves that are largely accountable for several system bottlenecks. Specifically, the model did not take into account passenger behaviors that would be related to their degree of flight experience, physical abilities, presence of children, etc, which would certainly impact the overall rate of passenger flow through an airport. Furthermore, the earlier model is dated and does not include data on baggage screening procedures, which are an integral component of airport security in the present day.

Simulation can also be used in the design process. It can be used to look at how people will move through a building, or to see how a change affects the rest of the system being designed. Brown and Garcia [7] used Simulink, in Matlab, to help design a control system for unmanned aerial vehicle helicopters. This allowed them to try different control systems without incurring the cost of building them and testing them in the real world

The goal of our current research therefore is to develop a working model of an airport using discrete event simulation with particular emphasis on homeland security. The simulation can be used for homeland security purposes to understand better where workers are needed to provide optimal security for waiting passengers.

Through this model, we represent traffic flow through an airport as a chronological set of events that is tied in to passenger behavior. Each event (e.g., arriving at check-in, carry-on baggage check, and final ticket check) occurs as an instant in time and marks a change of state in the system. The simulation was designed using ARENA Discrete Event Simulation software as described in the next section. Discrete Event Simulation (DES) software was created to simulate real world events that have random components to them and that are not time driven. How the simulation moves forward is based on arrival and service times drawn from a random number generator, which can be given functions from which to draw these numbers. These random times tell when an entity will arrive, and how long it takes to process the entity. The reason to use DES for the airport simulation is due to its simplicity in creating, the ability to recreate the random arrival and service times, and that the arrival of passengers and the time it takes to process them is not moved forward by the time moving forward.

2.0 THE SIMULATION

2.1 Materials

Laptop computer with Windows XP running ARENA DES Software Version 10.0 build 30.

2.2 Simulation Components

The simulation can be broken into multiple components each of which is combined in different places of the simulation to create the integrated airport simulation.

2.2.1 <u>Creation module</u>

This module is used to populate the simulation with entities, which in this simulation are passengers. The creation module determines how many passengers are going to arrive at the airport, and how often they arrive. With having a creation module, at the end of the simulation a delete module must be used to remove the passengers and have them leave the simulation.

2.2 Assign module

This module allows specific attributes to be assigned to the passengers, such as a function to predict how long it should take the passenger to get through the baggage check-in.

2.2.3 Decision module

The decision module is used to route passengers through a choice. For example one decision module routes the passenger to either the automated self check-in, or the manned check-in counter based on random chance, based on a percentage of passengers or even a formula.

2.2.4 Process module:

This module is used to carry out a specific process, such as the check-in process or the luggage screening process. Each process has specific resources that are assigned to it, such as the security screeners, baggage handlers and check-in agents.

2.3 Data Collection

Data for the simulation was collected from the Norfolk International Airport with consent from the different airlines and also the Transportation Security Administration for the airport. Data was collected between 7:00 am and 3:00 pm Monday through Thursday for two weeks. Arrival times were collected by using a stop watch and measuring the time between each passenger crossing a particular point when arriving into the airport building. These times were then recorded for later use in the simulation.

The processing times for the check-in were measured by observing passengers checking in. When the passenger started talking with the ticket counter agent or when they first touched the computer screen the stop watch was started. When the passenger gathered their luggage and moved away was when the time would stop. This data was recorded for later use in the simulation.

Processing times for the carry-on luggage screening were collected by observing passengers going through the security checkpoint. The stop watch was started once passengers put their luggage on the conveyor belt and stepped away to go through the metal detector. The time was stopped once they picked up their luggage.

These different times were put into the input analyzer of Arena DES so that an equation could be fit to the data and then put into the simulation. See table 1 for the airport data.

2.4 The Airport Simulation

Since this simulation deals primarily with passenger flow through an airport, the only parts of the airport that were simulated were those that directly affect the passengers themselves as they enter and travel through the airport, and finally board their plane.

Three main areas that were used in the simulation:
> (i) the initial check-in,
> (ii) the carry-on luggage screening

Input Equations	
Arrival times	-0.001 + GAMM(0, 0)
Manned check-in Times	54 + EXPO(0)
Self-check-in times	60 + WEIB(0, 0)
Security check point	10 + WEIB(0, 0)

See Figure 1 for a diagram of the ARENA simulation. These two points were chosen because they are the points where passenger flow is controlled by airport authorities, yet have the most impact on passenger behavior. The time when the passenger arrives at the airport cannot be controlled, and is therefore a random variable within the simulation, and is treated as such.

The arrival times of passengers are randomized based on data collected at the Norfolk International Airport. The passengers were categorized based on the main air carriers operating at the Norfolk Airport:

- American/Continental Airlines
- Southwest Airlines
- USAirways

American and Continental Airlines were grouped together due to the extremely low passenger rate observed at the airport. Each passenger category was assigned a different process time based on times collected from each processing area. The check-in area (see Figure 1) is divided into self check-in and manned check-in, for each airline. For the self check-in, the primary resource is the automated check-in machine. For the manned

station the primary resource is personnel manually checking the passengers and their luggage.

The next area the passengers went through was the luggage screening security checkpoint (see Figure 1). Each passenger goes through this section, just as they do in the real world. Random stops were able to be initiated in the simulation at this point. For example, the number of passengers stopped could be set as a predetermined percentage, and that many passengers will be stopped. Alternatively, certain passengers can be assigned a particular attribute tag such as race, gender or physical ability; then those passengers would be stopped more often in the simulation than other types of passengers. The simulation was run for 80 iterations, one iteration being a 24 hour a day.

3.0 SIMULATION RESULTS

After running the simulation, the average number of entities that entered the simulation was 179.14 for American/Continental, 1068.43 for Southwest, and 957.09 for USAirways. See table 2 for the range and average wait times.

The wait time for USAirways in the simulation indicated a significant difference between the manned check-in and self-check-in ($t(78) = 4.33$, $p < .001$), with the manned check in having a lower wait time ($M = 2.64$, $SD = 1.23$) than the self-check-in ($M = 12.07$, $SD = 3.98$). The manned check-in and automated check-in for American/Continental and Southwest airlines were not statistically different ($t(78) = 0.16$, $p = ns$; $t(78) = 0.07$, $p = ns$). In the simulation, Southwest's manned and self-check-in ($t(78) = 2.11$, $p < .05$; $t(78) = 2.72$, $p < .01$) and USAirways's self-check-in ($t(78) = 5.10$, $p < .001$) had significantly longer wait times than did the security checkpoints.
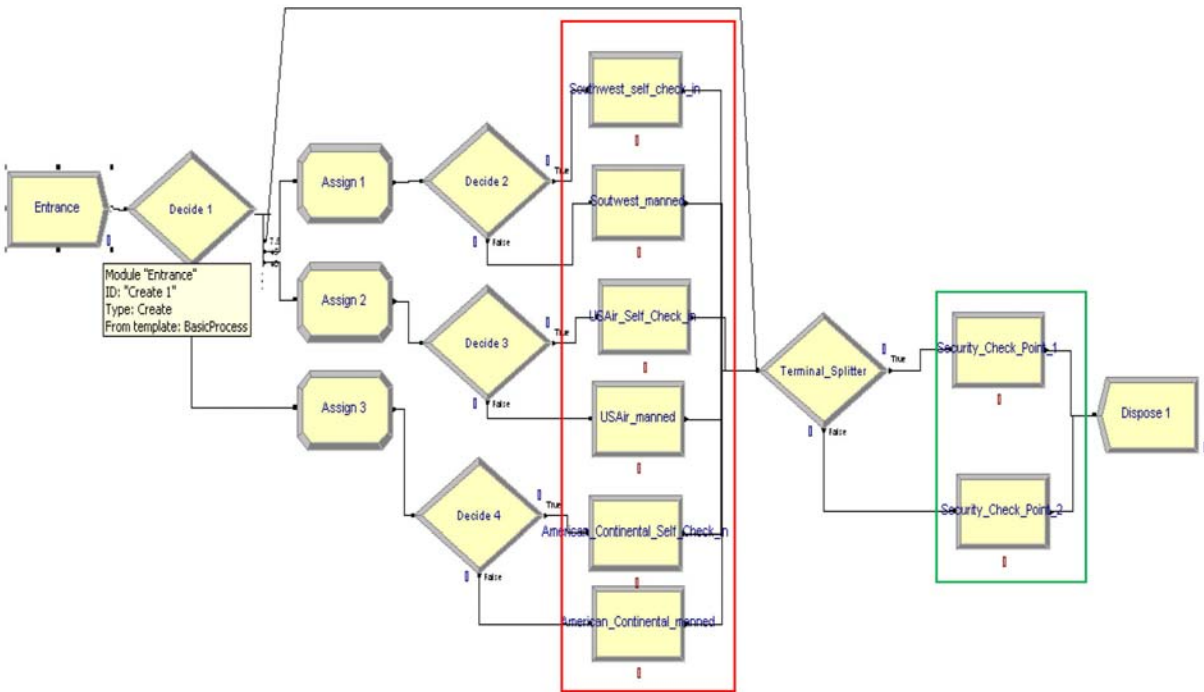

Fig. 1. ARENA diagram of Airport Simulation. The red area represents the check-in area. The green area represents the carry-on luggage check points.
  Table 2

Wait Time

Airline/Checkpoint	Average	Minimum	Max
American_Continental			
manned	2.51	0.00	33.28
self check-in	2.71	0.00	56.57
Southwest			
manned	6.74*	0.00	67.72
self check-in	6.60*	0.00	75.26
USAir			
manned	2.64	0.00	33.92
self check-in	12.07***	0.00	105.48
Security Check point 1	1.46	0.00	17.71
Security Check point 2	1.45	0.00	18.78

Note: time in minutes;
*p<.05, ***p<.001

Instantaneous utilization is another way to look at how resources are being used within the simulation. See table 3. Instantaneous utilization shows the percentage of time that the resource was used. The higher the percentage, the more the resource was used.

Table 3.

Instantaneous Utilization	
Type of Service	Average
Norfolk Airport	
Manned Check-in	78.40%
Self Check-in	83.38%
Security Checkpoint	82.43%

4.0 OPTIMIZING PASSENGER FLOW

As the results indicate, Southwest and USAir have the longest waiting times for both manned check in as well as self check-in. To decrease the wait times, one would assume that increasing the number of stations would decrease this wait time. As table 4 shows, the wait times actually increase significantly with increased stations, in this case 2 additional workers and 10 additional self check-in stations.

Another way to optimize the Southwest and USAir wait times is to divide the check-in

stations into multiple lines, for this case two. This allows people to choose a line that is shorter decreasing their wait time. As Table 5 shows, dividing the check-in lines to two lines for Southwest and USAir, wait times were cut by over half.

These wait times, however, do not tell the whole story. To get the full story, we also need to examine the resource utilization for these two changes. See table 6. By increasing the number of check-in stations, utilization was increased by five to six percent, though the security checkpoint utilization was decreased. So even though the wait times were increased, the resource usage was also increased. By dividing the waiting lines, table 6 shows that resource usage was cut by 37-39 percent for the check-in stations. This means that the workers were only working for 9-10 hours of the 24 when the waiting lines were split in two.

Table 4

Wait Time

Airline/Checkpoint	Average	Min	Max
American_Continental			
manned	2.51	0.00	33.28
self check-in	2.71	0.00	56.57
manned 2 additional	1.38	0.00	33.28
self check-in 10 additional	0.83	0.00	33.28
Southwest			
manned	6.74	0.00	67.72
self check-in	6.60	0.00	75.26
manned 2 additional	25.13***	0.00	33.28
self check-in 10 additional	34.05***	0.00	33.28
USAir			
manned	2.64	0.00	33.92
self check-in	12.07	0.00	105.48
manned 2 additional	24.63***	0.00	33.28
self check-in 10 additional	34.45***	0.00	33.28

***p<.001

Note: time in minutes

Table 5

Wait Time

Airline/Checkpoint	Average
American_Continental	
manned	2.51
self check-in	2.71
Southwest	
manned	6.74
self check-in	6.60
manned 2 lines	0.19***
self check-in 2 lines	0.088***
USAir	
manned	2.64
self check-in	12.07
manned 2 lines	0.06**
self check-in 2 lines	0.14***

Note: time in minutes
**$p<.01$, ***$p<.001$

Table 6

Instantaneous Utilization

Type of Service	Average
Two Lines	
Manned Check-in	40.70%
Self Check-in	43.89%
Security Checkpoint	82.32%
plus 10 Self Check-in	
Manned Check-in	85.39%
Self Check-in	89.11%
Security Checkpoint	79.69%

5.0 DISCUSSION AND IMPLICATIONS FOR OPTIMIZING PASSENGER FLOW

Airplanes move a large percentage of the population - about 580 million passengers just in the year 2008 in the US [8]. When there is a procedural failure in any one section of an airport it can have a drastic effect on the entire air transportation system. Therefore, the primary application of this simulation is to assist in optimizing traffic flow within airports. The results from the simulation runs indicate that the chokepoint at Norfolk International Airport resides with the initial ticketing and baggage checkpoints.

Southwest and USAirways are the primary carriers that can take a number of actions to try and reduce the waiting time associated with check-in. One possible method of redressal is to increase the number of self-check-in stations so that more people can use them at once. Another option is to do a usability analysis on the self-check-in station to make sure the process is smooth, efficient, and easy for inexperienced travelers to use. Finally, more workers could be brought in to help the passengers check-in.

The maximum utility of the airport model is that the effects of these changes can be tested in the simulation before changes in the system can be made. The number of self-check-in stations and manned stations can be repeatedly adjusted and the wait times can be analyzed to see what the optimum number is. The effects of failures and emergencies can also be examined within the model.

For emergency planning and error redressal, the ultimate goal is to try and plan for future events by using past experience [9,10]. As described above, our model allows for the quantification of each contingency situation into a discrete variable. These discrete variables include passenger behaviors that can be quantified to create individual 'agents' that exhibit different behaviors at different points in time. Each variable is then built into the simulation as described to ultimately predict the parameters required for optimal rate of passenger flow inside an airport. All of these ideas will be done in future testing of the simulation model.

6.0 CONCLUSIONS

The simulation model has indicated that there are choke points within the Norfolk International Airport. Those choke points are the check-in stations where passengers check their luggage. We recommend that the airlines in charge of the specific stations should decrease the wait time by increasing the number of staff and/or increasing the number of self-check-in stations. Besides to the obvious economic advantages of regulating passenger flow, minimizing choke

points will also ensure fewer instances of confusion and crowding at airports thereby strengthening the degree of passenger security to a large extent.

7.0 REFERENCES

[1] Li, J.P. (2000) "Train station passenger flow study," In J.A.Joines, R.R.Barton, K.Kang & P.A.Fishwick, Eds., *Proceedings of the 2000 Winter Simulation Conference,* 1173-1176.

[2] Siikonen, M., Susi, T., & Hakonen, H. (2000) "Passenger traffic flow simulation in tall buildings," *Proceedings of the International Conference on Multi-purpose high-rise Towers and Tall Buildings,* 1-16. Netherlands,

[3] Nahke, J., (2001) "Using simulation tools for upgrading existing people mover systems at major international airports," *American Society of Civil Engineers Conference Proceeding, 106, 65,* 1-17.

[4] Ke, C., Zizheng, G., & Liling, Z., (2009) "The optimization method of bus passenger flow peak," *International Conference on Transportation Engineering 2009,* 2207-2212

[5] Wusheng, L., & Qian, T., (2009) "Passenger-flow characteristics and scale of departure curbside in airport landside," *International Conference on Transportation Engineering 2009,* 3236-3241.

[6] Gatersleben, M. R., & Van der Weij, S.W. (1999) "Analysis and simulation of passenger flows in an airport terminal," In P.A. Farrington, H.B.Nembhard, D.T.Sturrock & G.W.Evans, Eds., *Proceedings of the 1999 Winter Simulation Conference,* 1226-1231.

[7] Brown, A., & Garcia, R., (2009) "Concepts and validation of a small-scale rotorcraft proportional integral derivative (PID) controller in a unique simulation environment," *Journal of Intelligent & Robotic Systems, 54,* 511-532.

[8] US Department of Transportation Bureau of Transportation Statistics (2008) "September 2008 airline traffic data: system traffic down 8.4 percent in September from 2007 and down 2.1 percent for January-to-September," Washington, DC: Smallen, D.

[9] Perry, R.W., & Mankin, L.D., (2005) "Preparing for the unthinkable: Managers, terrorism, and the HRM function," *Public Personnel Management, 34,* 175-193.

[10] Anderson, E., (2003) "Be prepared for the unforeseen," *Journal of Contingencies and Crisis Management, 11,* 129-131

4.2   Integrating Advanced Airspace System Components in a NAS-Wide Simulation

Integrating Advanced Airspace System Components in a NAS-Wide Simulation

Patricia Glaab

NASA Langley Research Center

October 14, 2010

Agenda

- Organization and programs supported
- NAS-wide simulation for systems analysis
- ACES simulation quick overview
- Enhancements for new capabilities
- Demonstration videos
- Future research possibilities

2

Organization

- Aeronautics Systems Analysis Branch (ASAB), NASA Langley Research Center
- Aircraft and airspace system concept analysis
  - Both customer supplied and internally defined
  - Identification of promising new technologies
  - Support agency's strategic research planning
  - Support competitive aerospace proposal generation and evaluation
  - Use and advance an integrated suite of tools to conduct this analysis

3

ASAB Support for NextGen

- NextGen time frames
  - Near-term – by 2012
  - Mid-term – by 2018
  - Final capabilities - post 2025
- ASAB supports far-term goals
  - Assumes advanced airspace management tools
  - Highly automated decision making
- Research areas
  - Demand/capacity/constraint analysis
  - Metroplex operations

4

MODSIM WORLD
Conference & Expo

NAS-Wide Simulations

- Systems Analysis for NextGen requires capability to model at National Airspace System (NAS) level
- Focuses on overall benefits, rather than individual components and capabilities of a particular aircraft
- Large number of flights modeled
  - FAA Terminal Area Forecast (TAF) report:
    - 30000 flights/day (current day avg, cont. US, commercial)
    - > 40000 flights/day - projected for 2030

5

MODSIM WORLD
Conference & Expo

NAS-Wide Simulations

- NASPAC (FAA)
  "National Airspace System Performance Capability"
- SIMMOD (FAA)
- PNP (Sensis)
  "Probabilistic NAS Platform"
- RAMS – Eurocontrol Experimental Center
  "Reorganized ATC Mathematical Simulator"
- TAAM (MITRE)
  "Total Airport and Airspace Model"
- ACES (NASA Ames)
  "Airspace Concepts Evaluation System"
 - Open source

6

ACES Simulation Overview

- Developed to assess system-wide impacts of airspace technologies and operational concepts
- Agent-based simulation
  - Event-driven components
  - Time-driven components (event = time step)
- Provides modeling of current day NAS
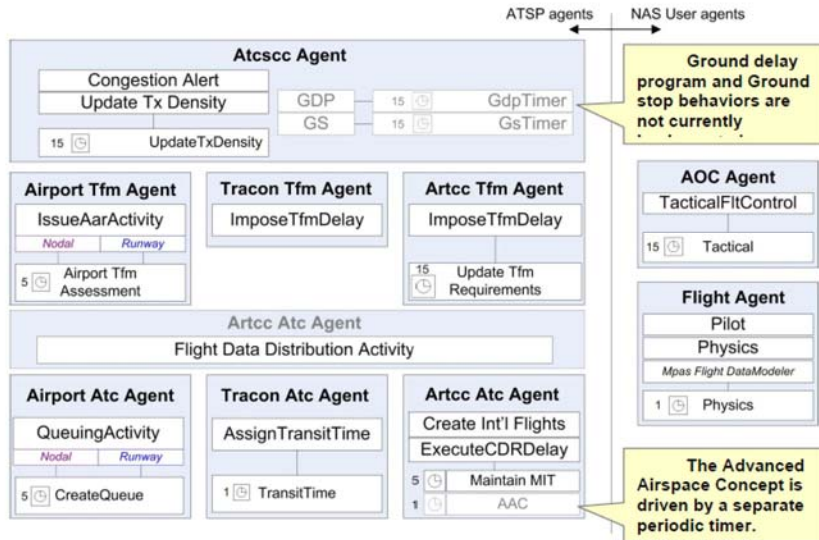- Extensible (via "plugins") framework

7

ACES Capabilities

- Uses Cybele (IAI) as core executive
- Agents in ACES map to real world entities in the National Airspace System (NAS)
  - Flights
  - Airports
  - TRACON ATC
  - En-route ATC
  - Surveillance
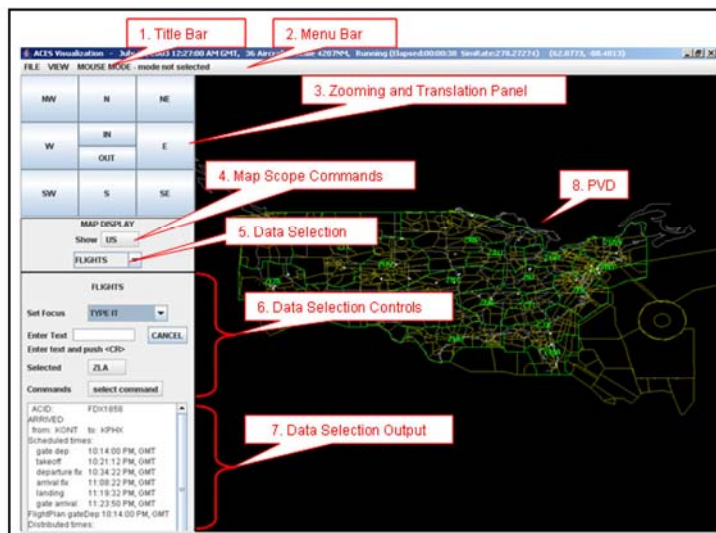 - Physical layout of airspace (sectors, centers)
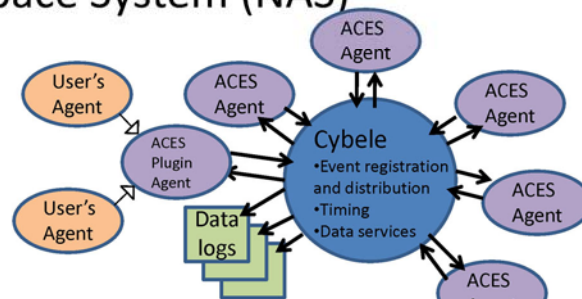
8

ACES Overview

CDRL 18 ACES Programmer's Guide, Rev 2, p 13.

9

ACES Visualization

CDRL 17 Airspace Concept Evaluation System (ACES) User Guide, Version 4, p. 92.

10

ACES Demonstration

(Video of ACES visualization window running a typical simulation scenario with midday traffic volume)

11

ACES Viewer

- ACES support tool for post-run visualization
- Runs using IV4D
  - Built for Air Force Research Labs by Aerospace Computing, Inc (ACI)
- Visualizes anything with lat/long/alt/time points
- Extended to support ACES output style

12

ACES Viewer Demo

Video of previous ACES demo video, now run in ACES Viewer with 3-D view rotated and manipulated

13

ACES Enhancements

- ACES provides a powerful framework, but must be extended for new concept testing
  - Merging and Spacing (M&S) in the airport vicinity
  - Conflict Detection and Resolution (CD&R)
    - Tactical
      - State-based
      - Prevent impending (< 2 minute) loss of separation (LOS)
    - Strategic
      - Intent-based
      - Prevent future (10-20 minutes out) LOS event
- Default ACES cannot support this type of study

14

ACES Capabilities

- CD&R in ACES
  - Tactical only
  - Based on NAS Center boundaries
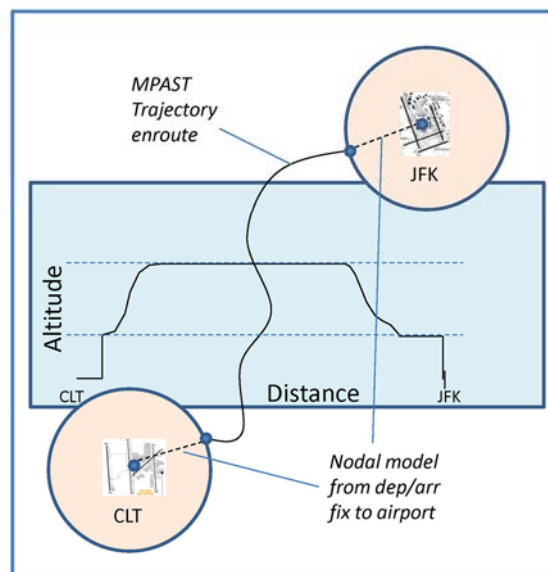 - Very limited capability

Nearby aircraft across Center boundary ignored

15

ACES Capabilities

- No M&S in ACES
  - Default TG is MPAST
  - MPAST does not model trajectory between arrival/ departure fix and airport

  (Node/ Queuing model)

MPAST Trajectory enroute

JFK

Nodal model from dep/arr fix to airport

CLT

16

ACES Enhancements

- NASA Langley contracted software development for prototype
- Intelligent Automation, Inc. (IAI)
  - ACES development team member
  - M&S concept developed in previous initiative
  - CD&R (tactical) developed in previous initiative
- Expanded CD&R
  - ACCoRD (tactical) – NASA LaRC, NIA
  - Stratway (intent) – NASA LaRC
- M&S
  - Refinement of IAI concept design
  - Multi-Point Scheduling Algorithm – NASA ARC

17

Current Status – M&S

- Two airports with detailed databases
  - Atlanta Hartsfield (KATL)
  - Dallas/Fort Worth (KDFW)
- M&S development complete
- Testing mostly complete
- Demonstration of full system in progress

18

Current Status – M&S

Video of ACES simulation run with M&S running traffic to KATL

19

Current Status – CD&R

- Implementation complete for tactical and strategic CD&R
- Work on-going with CD&R Stratway and ACCoRD team to provide feedback for continued tool development
- Integration with M&S completed
- Testing mostly complete
- Demonstration of full system (M&S with CD&R) in development

20

Current Status - CD&R

Video of ACES simulation running with strategic CD&R enhancements

Future Research Possibilities

- Quantification of airport throughput as a function of aircraft spacing (R. Brown, 2010)
- Arrival routing concept development to improve airport throughput
- Effect of CD&R maneuver strategies on system delay and fuel efficiency
- Impact of CD&R on M&S efficiency and robustness

Questions/Discussion

23

Backup Slides

24

(Backup Slide) NAS Flights Estimates

FAA's Terminal Area Forecast, 2010, page 18:

2008 (last historic data available)

Yearly National Total Commercial (takeoffs and landings) (includes Alaska and West Indies)	27951930
Yearly Alaska	-937116
Yearly Western Pacific	-4899428
Yearly Continental US (takeoffs and landings)	22115386
Daily Flights (yearly operations/2 ops per flight/365 days)	30295

2030 (Projected Data)

Daily Flights ((36646248 NT – 1059046 AK – 6113579 WP)/365/2)	40375

25

490

4.3   NextGen Future Safety Assessment Game

NextGen Future Safety Assessment Game

Ersin Ancel and Adrian Gheorghe
Old Dominion University, Norfolk, VA
eance001@odu.edu agheorgh@odu.edu

Sharon Monica Jones
NASA Langley Research Center, Hampton, VA
sharon.m.jones@nasa.gov

Abstract. The successful implementation of the next generation infrastructure systems requires solid understanding of their technical, social, political and economic aspects along with their interactions. The lack of historical data that relate to the long-term planning of complex systems introduces unique challenges for decision makers and involved stakeholders which in turn result in unsustainable systems. Also, the need to understand the infrastructure at the societal level and capture the interaction between multiple stakeholders becomes important. This paper proposes a methodology in order to develop a holistic approach aiming to provide an alternative subject-matter expert (SME) elicitation and data collection method for future sociotechnical systems. The methodology is adapted to Next Generation Air Transportation System (NextGen) decision making environment in order to demonstrate the benefits of this holistic approach.

1.0   INTRODUCTION

The world is heavily dependent on various critical infrastructures in areas like transportation, power, communication, water, energy, etc. Today's critical infrastructures are large-scale socio-technical systems, comprised of multiple components, involving various stakeholders, technologies, policies and social factors [1]. In the recent years, numerous sociotechnical systems started to undergo a series of transitions. The definition of system transition is given as "a long-term fundamental change (irreversible, high-impact and of high-magnitude) in the cultures (mental maps, perceptions), structures (institutions, infrastructures and markets), and practices (use of resources) of a societal system" [1]. In other words, the transition includes "a structural change in both technical and social subsystems" [2].

The planning and implementation phases of such large-scale infrastructure transitions require close monitoring of performance parameters like safety, efficiency, and sustainability. Ensuring that infrastructure transition reveals a safer and more sustainable system became a major challenge for the society [3]. In order to do so, decision makers often need to test various strategies and perform analyses to characterize risk and other parameters. However past strategies and historical data regarding previous infrastructure systems are no longer adequate for next generation infrastructure system design due to (1) previous systems evolved via incremental changes and system improvements which lead them to be unsustainable (i.e. congestion, energy shortage, air transportation delays, etc.) and (2) previous infrastructures were made to last, robust but resistant to change [1, 4]. The lack of empirical data causes decision makers to heavily rely on expert opinions for next generation infrastructure planning.

1.1.1   Lack of Data and Expert Elicitation

The future status of man-made systems like energy, transportation, warfare, agriculture, and other infrastructure cannot be predicted over a prolonged time frame. Large-scale sociotechnical systems are made up of multiple components that involve numerous stakeholders, technologies, policies, and social factors [1]. Decision makers and policy makers often require expert opinion to comprehend the complexity and uncertainties within such systems. Expert elicitation methods typically have been used to obtain the necessary data for reliability and risk studies for these types of

technological, environmental, and socioeconomic issues [5, 6]. Furthermore, NextGen will inherently be different from the current system. However, the ability to predict the future remains limited owing to the long-term implementation phase and the large number of uncertainties [5]. "There are no data about the future on which to rely. We are challenged to imagine many different and possible 'futures' as humankind seeks to exert its mastery and control" [7]. The crucial task is to think innovatively and recognize the creative and imaginative capacities of each stakeholder. The overall goal of the methodology is to reduce complexity and uncertainty while inventing the future and analyzing the respective risk for each alternative scenario [7].

1.1.2 Sociotechnical Complexity
The system-wide upgrade of complex systems is a challenging undertaking [8]. The increased complexity adds to the diversity of decision maker's system interpretations that can directly alter the overall system operation and the decision-making processes. Brewer [7] states that real-world problems do not exist independently of their socio-cultural, political, economic, and physiological content, and for that reason, an approach with multiple perspectives and multiple disciplines is necessary to efficiently clarify the matter at hand (which is quite challenging in practice). The presence of multiple actors with frequently divergent and conflicting interests can turn large-scale infrastructure transitions into wicked problems [9]. Traditional policy and market practices have proven ineffective in dealing with problems with a high degree of uncertainty regarding future scenarios and actor interactions [1]. For this reason, creating a methodology that attempts to predict decision pathways for future systems while accounting for the technical, organizational, and contextual complexity of the system is necessary [4, 7].

2.0 TEST CASE
The goal of this research is to develop a methodology that will serve as an aid for decision makers who are responsible for designing and evaluating scenarios for future technological implementations within next generation infrastructure systems. As previously mentioned, the implementation of large system transitions require understanding of the multi-layer complexity and overcoming the lack of experimental data for designing the future phases of the system. In order to demonstrate the proposed methodology, the planning, development, and implementation of the Next Generation Air Transportation System (NextGen) is used as a test case. The following sections will provide insights on NextGen.

2.1 NextGen Overview
The U.S. National Airspace System (NAS) is made up of a number of multifaceted elements, including over 800 billion passengers and input from more than 15,000 air traffic controllers to assist 590,000 pilots onboard 239,000 aircraft that take off and land at 20,000 U.S. airports. This extremely complex system is closely tied to the national economy, contributing $1.2 trillion annually and over 5 percent of the gross domestic product while generating 11 million jobs and $369 billion in earnings [10].

The delays that currently impact passenger travel are forecasted to be even higher in the future as the demand for air transportation is expected to increase. In addition, future airspace is expected to accommodate unmanned aircraft systems and commercial space vehicles as well. Furthermore, the entire system is expected to operate within acceptable safety levels and environmental impact guidelines [10].

To respond to this forecasted increase in demand, the Joint Planning and Development Office (JPDO) was formed during the Bush Administration in 2003. This organization is a partnership between public

492

and private stakeholders, including the Federal Aviation Administration (FAA), the Department of Defense (DoD), the Department of Homeland Security (DHS), NASA, and others [11]. The JPDO is charged with developing concepts, architectures, roadmaps, and implementation plans for transforming the current national Air Transportation System (ATS) into NextGen.

"During the next two decades, demand will increase, creating a need for a system that (1) can provide two to three times the current air vehicle operations; (2) is agile enough to accommodate a changing fleet that includes very light jets (VLJs), unmanned aircraft systems (UASs), and space vehicles; (3) addresses security and national defense requirements; and (4) can ensure that aviation remains an economically viable industry" [8].

2.2 NextGen Challenges
The complex nature of the NAS, combined with numerous operational and management challenges, threatens the NextGen efforts. Reports from the Office of the Inspector General (OIG) reveal that the Federal Aviation Administration (FAA) is facing difficulties in developing a strategy to engage stakeholders, not to mention managing and integrating multiple NextGen efforts [12]. Uncertainties and a lack of data related to shaping a future aviation system also inhibit the ability to employ formal risk analysis methods. As a result, SME opinion has become the primary source of input for the NextGen scenarios, technologies, and safety.

2.3 Need for a Methodology
In the past, traditional engineering design approaches focused primarily on the technical requirements. Similarly, traditional infrastructure designs were treated like traditional engineering problems, causing them to be brittle and resistant to modernization. However, the next generation infrastructures must be treated differently because of their complex nature

[13]. The need to understand the infrastructure at the societal level and capture the interaction between the technical, political, and economic factors becomes more important [4]. Traditional engineering design methods are used in concert with serious gaming approaches to create a holistic decision-making methodology. The goal of the proposed methodology is to enable decision makers and researchers to gather information in regard to NextGen safety values. Toward that purpose, various tools and techniques are employed collectively here to create a methodology that can be used as an alternative to conventional expert elicitation techniques (i.e. Delphi Method, Nominal Group Technique, brainstorming, etc.) for complex systems with multiple stakeholders.

3.0 PROPOSED METHODOLOGY
As discussed above, the methodology for estimating risks within a future system will combine various approaches. Because the air transportation system includes extensive interactions between multiple stakeholders, which can be difficult to track, and because of the lack of historical data, SMEs from diverse backgrounds are the main source of data for this study [6]. Aviation safety within NextGen is measured by using the probability number method. Conventional numerical and qualitative expert elicitation techniques provide the gaming data that are necessary to construct the scenarios, alternatives, attributes, and so on. Commercial-off-the-shelf software packages (i.e., Logical Decisions for Windows® and Precision Tree® by Palisade Corp.) are also used to rank future technologies in order to support SME opinions before and during the gaming cycle. Fig. 1 provides an overview of the methodology. The various tools and techniques are described in more detail in subsequent sections.
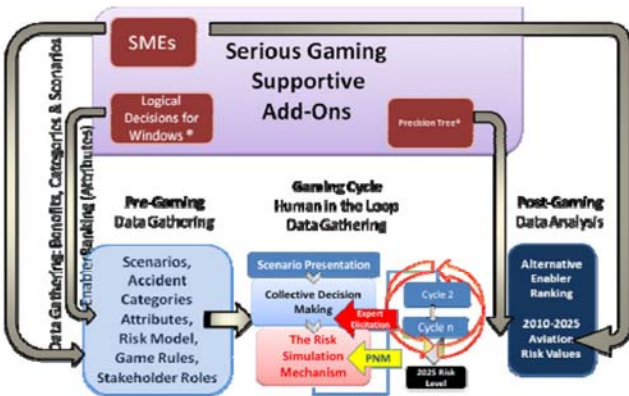
493

Figure 1. Serious gaming methodology: high-level architecture

3.1 Serious Gaming and Infrastructure Design

The use of gaming and simulation techniques as a formal approach to strategy making has gained wide acceptance, as evidenced by the frequency of occurrence within mainstream strategy literature [14]. Gaming methods (or soft systems thinking) have become an alternative to formal complexity modeling techniques like systems dynamics and operations research. These techniques have been successfully applied to well-structured problems; however, when employed on ambiguous and often ill-structured and complex systems, their contribution has been limited because adequate theory and empirical data were absent [14].

Serious gaming methods are able to provide decision makers with an environment in which the totality of the system and its dynamics are present. With a holistic approach that includes the wide-range of perspectives, skills, information, and mental models of the involved parties, the quality of the decision-making environment increases dramatically [13, 14]. Unlike hard-system methods, the gaming and simulation approach is quite flexible and easily adaptable to other quantitative methods, scenarios, and computer models [16]. Policy gaming methods can help both participants and modelers understand the big picture and identify critical elements of

the complex problem at hand. Because of the iterative and experimental nature of these gaming and simulation environments, participants are able to test different approaches within both a safe environment and a condensed timeframe [15].

According to Duke [17], a typical complex real-world situation has the following characteristics: it contains numerous variables in interaction; no realistic basis exists for quantification of these variables or their interactions; and no proven conceptual model or precedent exists on which action decisions can be based. Complex systems are also typified by a sociopolitical context of decision-making, where the actions of the various "players" may be idiosyncratic or irrational; furthermore, the decisions are irreversible and the results are not generally fully understood until well into the future [17]. NextGen fits this model, as a complex real-world air-transportation system that will undergo a full-scale transformation and that contains numerous stakeholders with often conflicting agendas, including those of the general public [18]. The gaming context may help capture the organizational and behavioral dynamics of the decision-making process and ultimately yield a more realistic problem solution.

The following section provides insight on the probability number method (PNM), which is used as the backbone for the NextGen risk calculation method.

3.2 Probability Number Method

The PNM was created through a joint effort between the International Atomic Energy Agency (IAEA) and several United Nations organizations. The method was developed as an affordable solution to quickly determine the risks that are associated with handling, storing, processing, and transporting hazardous materials. The methodology is supported by an extensive database that includes the various factors that impact the risks, including types of substances (i.e., flammable, toxic, or explosive gases or liquids), safety

494

precautions, population densities, environmental factors, and so on [19]. The average accident scenario contains only rough estimates because the purpose of the methodology is to serve as a decision-making aid that enables risk ranking and prioritization for further analysis. Dr. Adrian Gheorghe was a part of the development team for the PNM and brings his expertise to decoding, modifying, and adapting the probability number method to the NextGen system.

3.2.1 Probability Number

Within the PNM, the probability of the occurrence of a certain accident is calculated via a dimensionless probability number N, which is then transformed into an actual probability. The probability number can be adjusted or updated based on various correcting factors. The risk is defined as the product of the consequences and the probabilities of unwanted outcomes (i.e., hazardous events).

3.2.2 Adapted Consequences and Probabilities

The PNM defines risk as the product of the probability of an accident and its respective consequences, calculated separately. Within the NextGen framework, the consequences are defined as fatal aviation accidents (i.e., accidents per 100,000 flight hours). The probability of an accident that involves a passenger fatality is calculated in the following manner. An average probability number that represents the base assumption is determined; then, this number is adjusted by using correcting factors. These factors represent technological improvements and other enablers that are planned within the NextGen framework, namely, runway safety, aircraft reliability, icing, turbulence mitigation, weather, and airborne collision avoidance. The adjusted accident probability N is then converted into a frequency of occurrence.

The probability of occurrence and the consequence factors are inputted into the FAA's Risk Matrix (Fig. 2). The initial conditions for the risk are determined by the averaged accident data (2000-2009) which are obtained from NTSB website[1]. The average severity of aviation accidents is 0.291 fatalities/100,000 flight-hours or severity classification "Minor, 4"; meanwhile the probability of such an accident is 0.208/100,000 flight-hours, indicating a "Remote, C" likelihood category. Departing from the values above, the current aviation risk is determined as "Low Risk, green".
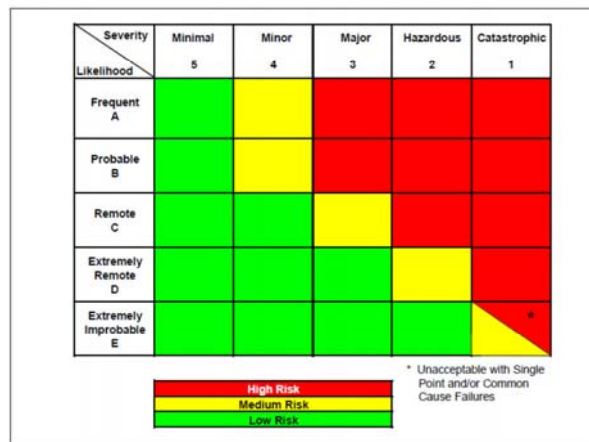

Figure 2. FAA risk matrix.

The PNM was chosen to be the backbone for the risk estimation engine in this application as a result of its intuitive structure and ease of expandability. The main components of the PNM, namely, the consequences, probabilities, and risk outcomes, are incorporated into the NextGen Safety Assessment Methodology and fused with the policy gaming effort.

3.3 Software Add-Ons
The selected gaming platform enables the integration of additional methods and techniques, which allows the methodology to remain flexible and expandable. This in turn ensures a more thorough

---

[1] http://www.ntsb.gov/aviation/Stats.htm

representation of the air transportation infrastructure. Two commercial-off-the-shelf software solutions have been embedded into the methodology to enhance the ranking and prioritization of enablers and other damage indicators. The Logical Decisions for Windows® software helps the prioritization and ranking of NextGen safety-related technological enablers with respect to their benefits, costs, implementation timelines, and other parameters. The Precision Tree® software is used to collect and further analyze the data that are obtained from the gaming exercise.

4.0 GAMING SEQUENCE

The NextGen safety assessment methodology was developed on a serious gaming platform that was adopted from the play sequence of policy gaming developed by Geurts, Duke, and Vermeulen [14]. An adapted version of the play sequence, which accommodates the NextGen safety framework, is given in Fig. 3. The sequence is initiated by the presentation of the game to the stakeholders; this includes providing the game rules, a general overview of NextGen goals, and the available resources. Stakeholder groups that contain participants from various backgrounds are formed, and their respective goals are established (e.g., the FAA is concerned with safety goals, commercial airlines with economic goals, and so on). The groups are asked to evaluate and select from the list of technological advancements that are related to the improvement of safety. However, the implementation of each of these advancements consumes some of the predetermined limited resources. Stakeholders with conflicting agendas must come to a consensus on certain decisions. Following these discussions, the decisions for each time step are entered into the risk simulation mechanism (based on the PNM) and the updated NAS risk values are calculated iteratively for the next three time steps, until year 2025. The gaming exercise is concluded with the debriefing and discussions in order to create the foundation for the data gathering and analysis.
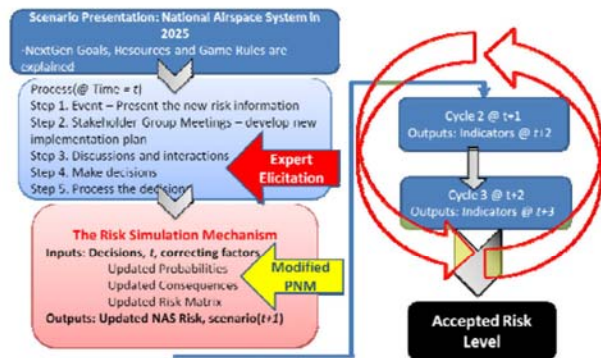

Figure 3. NextGen safety assessment gaming sequence.

4.1 Stakeholders and Game Rules

One of the most productive outcomes of the policy gaming exercise is that the participants are able to interact based on the problem at hand. The "safe" environment allows participants to create and analyze the system complexity while communicating various aspects of the issue among the stakeholders [7]. In order to model such a dynamic environment, a simplified list of involved stakeholders and engagement rules was developed. Stakeholder interactions can be based on rigid rules, free-form rules or combination of the two. Rigid, rule-based gaming is well suited for structured environments, such as military gaming where specific rule sets that can be formalized by mathematical or computational methods are used. However, for social arenas that include both public and intense stakeholder interactions without firm rules, free-form gaming, which relies on game rules, is more suitable. Free-form games enable the participants to challenge, modify, and improve the positions, objects, and rules during the game play. However, the process must be carefully monitored by a control team of experts who act as referees or game directors [16].

Within the scope of this project, the primary goal was to provide insight into the future of NAS safety and data gathering in regard to future systems. Thus, a combination of rigid and free-form gaming rules was employed. The goal of the game was to simulate the

aviation safety values within a given time frame, while taking into consideration technological constraints (i.e., cost versus benefit, feasibility, mixed equipage, and so on) and behavioral concerns (i.e., information overload to pilots, controllers, early technology adopters, and so on).

4.2 Outcomes

Throughout the gaming effort, the discussions and negotiations that occurred between participants with opposing agendas were important and can be used to develop different problem-solving approaches. The gaming exercise serves both as an individual and a collective learning platform for the stakeholders, leading to an overall elevated level of knowledge across the system. The individual learning took place during the decision-making process during which each stakeholder group represents their respective point of view. The acquired awareness that was gained in regard to the overall system complexity will ultimately improve the value of the expert elicitation. Furthermore, the presence of realistic interactions among the players yielded data that can be used in the testing and evaluation of NextGen-related technologies in the future [20]. In addition to the individual learning, the collective learning (or the organizational learning) provided valuable insight that relates to the problems that were discussed (i.e., NextGen aviation safety values).

One of the most tangible outcomes of the gaming exercise was the 2025 NAS safety values with respect to the FAA's Risk Matrix (Fig. 2) acceptability measures. The intermediate risk values that were obtained during the technology implementation phase (i.e., the next 15 years) were elicited under the same assumptions. The cumulative effect of various safety-related technological implementations can be examined, which will enable decision makers to identify technologies or areas that require further analysis and understanding.

5.0 CONCLUSIONS

The planning and implementation of next generation infrastructure transitions are challenging due to their complex nature and the lack of historical data. This paper proposes the use of simulation and gaming methods as a platform for evaluating and generating necessary data for designing future infrastructure systems. The Next Generation Air Transportation System (NextGen) decision making environment is used as a test-bed to demonstrate the developed methodology.

Subject-matter-expert opinions were heavily relied upon to develop the gaming components, to decide on the participants, and finally, to evaluate the validity of the framework. Conventional risk calculation methods and commercial-off-the-shelf software capabilities were integrated to provide system-level overview and risk analysis as a decision-making tool. One of the most prominent contributions of the gaming exercise was its ability to aggregate the perspectives of multiple stakeholders with varying agendas, while calculating the effectiveness of future NextGen safety enablers. The gaming environment promotes individual and collective learning across the system, allowing subject-matter experts to express their opinions for a more thorough and accurate modeling of the future infrastructure.

6.0 REFERENCES

[1]   N. Frantzeskaki and D. Loorbach, "Infrastructures in transition: Role and response of infrastructures in societal transitions," in *International Conference on Infrastructure Systems, Next Generations Infrastructures Foundation*, Rotterdam, Netherlands, 2008.

[2]   E. J. L. Chappin and G. P. J. Dijkema, "On the design of system transitions Is Transition Management in the energy domain feasible?," in *IEEE International Engineering Management Conference, IEMC '08*, Estoril, Portugal, 2008, pp. 1-5.

[3]   B. Elzen and A. Wieczorek, "Transitions towards sustainability through system innovation," *Technological Forecasting and Social Change*, vol. 72, pp. 651-661, 2005.

[4] R. J. Hansman, C. Magee, R. de Neufville, R. Robins, and D. Roos, "Research agenda for an integrated approach to infrastructure planning, design and management," *International Journal of Critical Infrastructures,* vol. 2, pp. 146-159, 2006.

[5] T. M. Chytka, "Development of an aggregation methodology for risk analysis in aerospace conceptual vehicle design," Ph.D. Dissertation, Eng. Mgmt and Sys Eng. Dept., Old Dominion University, Norfolk, VA, 2003.

[6] T. M. Chytka, B. A. Conway, and R. Unal, "An Expert Judgment Approach for Addressing Uncertainty in High Technology System Design," in *Tech. Mgmt for the Global Future, PICMET*, Istanbul, Turkey, 2006, pp. 444-449.

[7] G. Brewer, "Inventing the future: scenarios, imagination, mastery and control," *Sustainability Science,* vol. 2, pp. 159-177, 2007.

[8] JPDO, "Concept of Operations for the Next Generation Air Transportation System (NextGen)," Washington, DC, June 13, 2007.

[9] H.W.J. Rittel and M.M. Webber, "Dilemmas in a General Theory of Planning," *Policy Sciences*, vol. 4, no. 2, pp. 155-169, June 1973.

[10] FAA, "FAA's NextGen Implementation Plan 2009," Washington, DC, February 10, 2009.

[11] JPDO, "Next Generation Air Transportation System Integrated Work Plan: A Functional Outline," Washington, DC, Sept. 30, 2008.

[12] U.S. Department of Transportation, "Memorandum: Timely Actions Needed to Advance the Next Generation Air Transportation System," Washington, DC, June 16, 2010.

[13] D. Roos, R. de Neufville, F. Moavenzadeh, and S. Connors, "The design and development of next generation infrastructure systems," in *IEEE Int. Conf. on Systems, Man and Cybernetics*, 2004, pp. 4662-4666.

[14] J. L. A. Geurts, *et al.*, "Policy Gaming for Strategy and Change," *Long Range Planning,* vol. 40, pp. 535-558, 2007.

[15] I. Wenzler, "The role of simulation games in transformational change," in *Planspiele fur die Organisationsentwicklung* W. C. Kritz, Ed., ed Berlin: WVB, 2008.

[16] I. S. Mayer, "The Gaming of Policy and the Politics of Gaming: A Review," *Simulation Gaming,* vol. 40, pp. 825-862, December 1, 2009.

[17] R. D. Duke, "A Paradigm for Game Design," *Simulation Gaming,* vol. 11, pp. 364-377, 1980.

[18] U.S. Department of Transportation, "Budget Estimates Fiscal Year 2011 Federal Aviation Administration," Washington, DC, 2010.

[19] International Atomic Energy Agency, "Manual for the classification and prioritization of risks due to major accidents in process and related industries," IAEA, Vienna, Austria, Nov. 1996.

[20] C. Joldersma and J. L. A. Geurts, "Simulation/Gaming for Policy Development and Organizational Change," *Simulation Gaming,* vol. 29, pp. 391-399, Dec. 1, 1998.

7.0 ACKNOWLEDGMENTS

NextGen Future Safety Assessment Game

Ersin Ancel and Dr. Adrian Gheorghe
Old Dominion University

Dr. Sharon Monica Jones
NASA Langley

MODSIM World Conference & Expo 2010
Hampton, Virginia (October 13-15, 2010)

www.nasa.gov

Outline

- Introduction

- Problem Definition

- Test Case: Next Generation Air Transportation System (NextGen)

- Proposed Methodology

- Gaming Sequence

- Conclusions

Introduction
Large Scale Critical Infrastructures

- Future infrastructures are shaping up to become more complex and more interrelated than ever causing older systems to be unsustainable, faced with congestion, energy shortage, blackouts, etc.

- Transforming today's infrastructures and planning for the future is a challenging task since

  - Infrastructure systems are inherently robust and resistant to change

  - We lack the empirical data for next generation infrastructure planning

  - Increased multi-level complexity level of these systems

Introduction (cont'd)
Lack of Data/Expert Elicitation Challenges

- Large amount of complexity and uncertainty

- Expert elicitation methods have typically been used to obtain the necessary data for reliability and risk studies of technological, environmental and socioeconomic issues

- Necessary to be creative about possible "futures"

Introduction (cont'd)
Sociotechnical Complexity

- Presence of multiple actors with frequently divergent and conflicting interest

- Traditional policy and market practices are proven ineffective in dealing with high degree uncertainty caused by future scenarios and actor interactions

Problem Definition

- A methodology capable of addressing both technical and social aspects of the system, taking into consideration the technical, organizational, and contextual complexity of the system in a holistic manner

- Creating a venue to understand and enhance the communication between multiple stakeholders and decision makers

Test Case
National Airspace System (NAS)

- World's most complex aviation system

- Consists of

  - 15,000 air traffic controllers

  - 20,000 airports

  - 800B passengers

  - About 50,000 flights/day

- Contributing $1.2 trillion to the economy (5% of GDP)

Test Case (cont'd)
NextGen and Its Challenges

- Current NAS capacity is no longer adequate for upcoming demand (i.e. delays)

- Next Generation Air Transportation System (NextGen) is the transformation of NAS to reflect 2025 goals and requirements

  - Up to 3 times the passenger capacity

  - Increased safety levels while reducing the environmental impacts

- FAA is facing difficulties with

  - Managing efforts in an integrated way

  - Engaging private sector in the process

  - Coordinating a multi-agency approach

Proposed Methodology

- Components

  - Conventional risk assessment techniques

  - Serious gaming methods

- Objective is to help decision making process for System of Systems (SoS) level infrastructures by first reaching a common understanding of the system complexity by all stakeholders and then generate knowledge about the system under consideration

Proposed Methodology
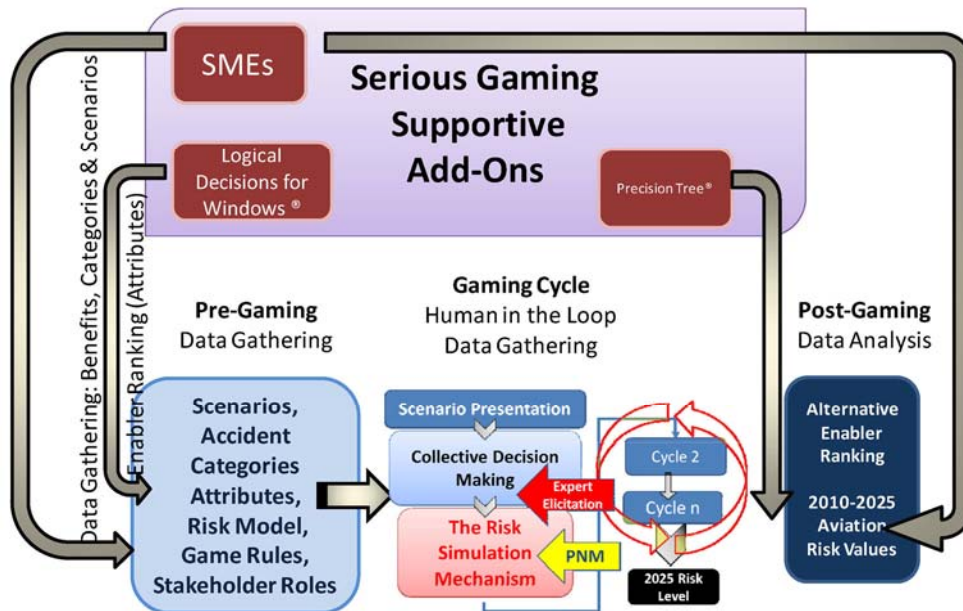
Serious Gaming

- Developed during the 1950 and 1960s

- Initially used within the urban studies, political science and business

- The goal is to engage participants in a safe environment in order to …

    …create and analyze the "futures" they want to explore

    …pre-test strategic initiatives

    …deal with the increasing organizational complexity by enhanced communication

Serious Gaming (cont'd)

- Serious gaming methods became an alternative to formal complexity modeling techniques

- Simulates experimental, rule-based interactive environments for players to learn from their actions

- Flexible and easily coupled with other quantitative methods, scenarios and computer models

- Helps participants and decision makers understand and communicate complexity

Probability Number Method

- Developed with the collaboration of

    - International Atomic Energy Agency (IAEA),

    - United Nations Environment Programme (UNED),

    - United Nations Industrial Development Organization (UNIDO),

    - World Health Organization (WHO) within the United Nations

- Dr. Gheorghe was a part of the Scientific Secretariat

- The method is developed to **determine the risks** associated to major accidents with off-site consequences **in fixed installations handling, storing and processing hazardous materials** (flammable, toxic or explosive gas or liquid) **or the transportation** of such materials by road, rail, pipeline and inland waterway

Probability Number Method (cont'd)

- Instead of calculating the Probability values, the PNM goes with determining a number that will be converted into the probability values (e.g. 3×10^{-4} accidents/year)

- The relationship is given as

$$N = \left| \log_{10} P \right|$$

- The product of probabilities and consequences (calculated separately) of each risk category identifies its respective risk

FAA Risk Matrix

Severity / Likelihood	Minimal 5	Minor 4	Major 3	Hazardous 2	Catastrophic 1
Frequent A	green	yellow	red	red	red
Probable B	green	yellow	red	red	red
Remote C	green	green	yellow	red	red
Extremely Remote D	green	green	green	yellow	red
Extremely Improbable E	green	green	green	green	yellow/red *

* Unacceptable with Single Point and/or Common Cause Failures

High Risk	(red)
Medium Risk	(yellow)
Low Risk	(green)

National Aeronautics and Space Administration

Supportive Add-Ons (COTS Software)

- The serious gaming exercise is supported by two commercial-off-the-shelf (COTS) software packages

Logical Decisions for Windows

- Used to rank and prioritize the technological and other enablers

- Provided to the players throughout the game

- Dynamic ranking helps the decision making process

Precision Tree

- Will be used to collect and further analyze the data extracted from the gaming exercise

Gaming Sequence
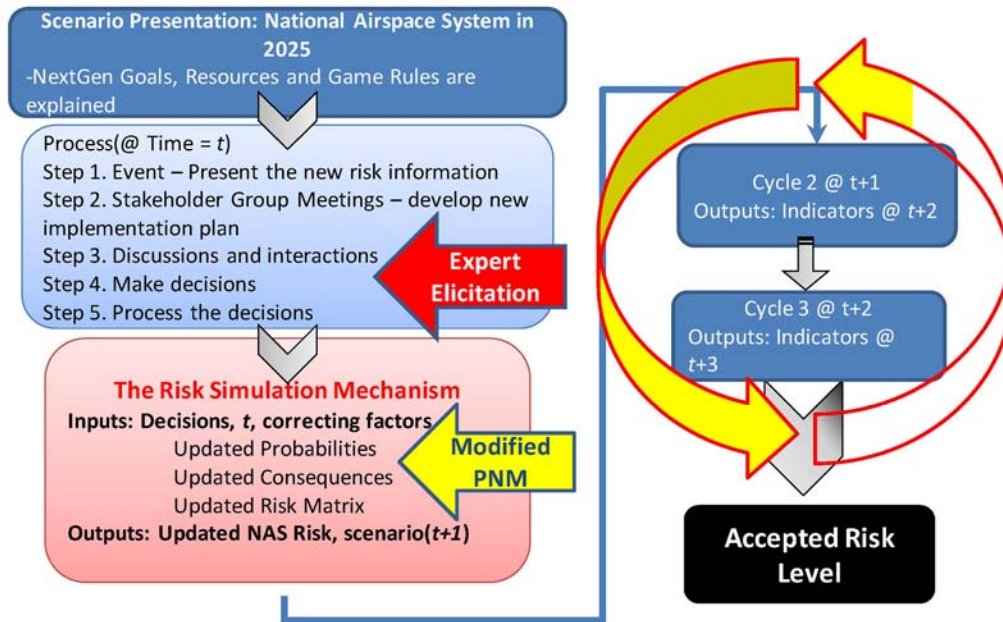
Outcomes

- Holistic approach (i.e. sociotechnical system as a whole)

- Individual learning

  - Learning and communicating complexity among stakeholders

  - Informed participants/decision makers yielding to better expert elicitation

- Collective Learning

  - Future aviation safety values

  - Decision making support for further analysis

Conclusions

- Planning and implementation of next generation infrastructure transitions are challenging due to their complex nature and the lack of historical data

- The ability to aggregate the perspectives of multiple stakeholders with varying agendas, while calculating the effectiveness of future NextGen safety enablers is addressed

4.4   Consequence and Resilience Modeling for Chemical Supply Chains

Consequence and Resilience Modeling for Chemical Supply Chains

Kevin L. Stamber, Eric D. Vugrin, Mark A. Ehlen, Amy C. Sun, Drake E. Warren, and Margaret E. Welk
Sandia National Laboratories
klstamb@sandia.gov edvugri@sandia.gov maehlen@sandia.gov acsun@sandia.gov
dewarre@sandia.gov mewelk@sandia.gov

The U.S. chemical sector produces more than 70,000 chemicals that are essential material inputs to critical infrastructure systems, such as the energy, public health, and food and agriculture sectors. Disruptions to the chemical sector can potentially cascade to other dependent sectors, resulting in serious national consequences. To address this concern, the U.S. Department of Homeland Security (DHS) tasked Sandia National Laboratories to develop a predictive consequence modeling and simulation capability for global chemical supply chains. This paper describes that capability, which includes a dynamic supply chain simulation platform called N-ABLE^{TM}. The paper also presents results from a case study that simulates the consequences of a Gulf Coast hurricane on selected segments of the U.S. chemical sector. The case study identified consequences that include impacted chemical facilities, cascading impacts to other parts of the chemical sector, and estimates of the lengths of chemical shortages and recovery. Overall, these simulation results can DHS prepare for and respond to actual disruptions.

1.0 INTRODUCTION

The U.S. Department of Homeland Security (DHS) [1] has identified that "protecting and ensuring the continuity of the critical infrastructure and key resources (CIKR) of the United States is essential to the Nation's security, public health and safety, economic vitality, and way of life." The chemical sector serves as one of the 18 CIKR sectors identified by DHS.

Analysis of chemical supply chains within this context is an inherently complex task, given the dependence of these supply chains on multiple CIKR systems (e.g., transportation, energy). This effort requires data and information at various levels of resolution, ranging from network-level supply chain systems to individual chemical reactions.

DHS has tasked the National Infrastructure Simulation and Analysis Center (NISAC) with development of a chemical infrastructure analytical capability to assess interdependencies and complexities of the nation's critical infrastructure, including the chemical sector. The Federal Government established NISAC, which includes personnel at Sandia National Laboratories (Sandia) and Los Alamos National Laboratory to support efforts aimed at identification of dependencies within and across sectors, providing consequence assessment to enable National Risk Analysis.

To address this need, DHS's Science and Technology Directorate has funded the Sandia component of NISAC in an ongoing effort to integrate its existing simulation and infrastructure analysis capabilities with various chemical industry datasets. The intent of this effort is to develop and ultimately provide capabilities in consequence and resilience analysis of natural and manmade events that impact the chemical industry and chemical-dependent sectors of the economy.

This document describes key elements of this ongoing development effort, including the modeling and simulation tools utilized in analyzing the chemicals sector from different perspectives. This includes a case study, examining the effects of a Gulf Coast hurricane on segments of the chemicals sector and an examination of consequence and resilience metrics.

2.0 BODY

Consequence and resilience analysis of the chemicals sector requires a wide range of modeling techniques to answer questions of varying scopes, acting on a common data set. To do this, Sandia developed and

populated a common data model, a set of modeling capabilities with different resolutions, and a framework for analyzing resilience.

2.1 Chemical Data Model (CDM)

Central to the development effort aimed at providing consequence and resilience analysis is a common Chemical Data Model (CDM). The CDM draws on infrastructure, population, labor, economic and other data sets from a variety of commercial (e.g., SRI Consulting, PennWell, Minnesota IMPLAN Group) and government (e.g., U.S. Bureau of the Census, National Geospatial-Intelligence Agency, and Surface Transportation Board) data sources, as well as on data developed during the project. CDM data are updated annually, at minimum, or as often as updates become available.

Figure 1 represents a simplified schematic of how the information within CDM is organized, merged, and stored. Each chemical plant in the database has attributes that identify where it is located, what chemicals are produced and stored, the associated capacities for production and storage, and what production technologies are used at the plant.
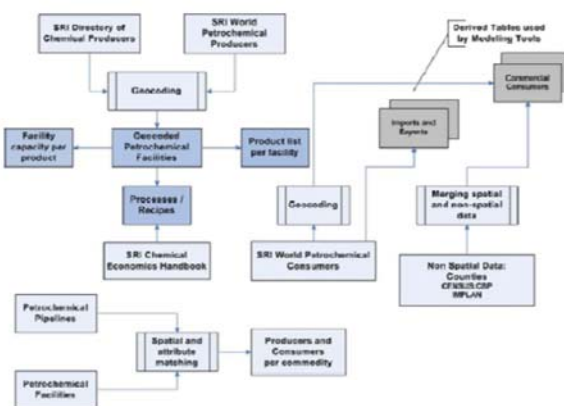

Figure 1. The Chemical Data Model.

2.2 Consequence Analysis Tools

A variety of Sandia-developed tools leverage this common data structure for various aspects of analysis of the chemicals sector (and other sectors as appropriate to the question).

2.2.1 FASTMap

Geospatial analysis is conducted using a tool called FASTMap. FASTMap is a geographic information system (GIS)-based tool that creates common look-and-feel, production-quality maps of the chemicals sector and other CIKR sectors relative to disruption areas, and provides data on CIKR assets (e.g., names, number of facilities) in the disruption area.

2.2.2 Fast Analysis Infrastructure Tool (FAIT)

Infrastructure dependency analysis is conducted within a tool called the Fast Analysis Infrastructure Tool (FAIT). FAIT provides data on the dependencies of specific chemical sector components (e.g., plants and pipelines) on assets in other infrastructure (e.g., electric power, transportation, emergency services).

2.2.3 Loki

Network analysis is conducted using a tool called Loki, which is a network model and analysis tool designed to quickly estimate potential production losses among chemical manufacturing processes.

2.2.4 Railroad-Network Analysis System (R-NAS)

Rail transportation analysis is accomplished through a network tool called the Railroad-Network Analysis System (R-NAS). R-NAS models the U.S. national rail network and estimates the impact to national rail commodity flows given disruptions to the rail system (bridges, rail yards, and so forth).

2.2.5 NISAC Agent-Based Laboratory for Economics (N-ABLE™)

Dynamic supply chain analysis is conducted using the NISAC Agent-Based Laboratory for Economics™ (N-ABLE™), a large-scale microeconomic supply chain model and tool that allows for the analysis of the impacts to individual firms (production, sales,

510

transportation, and inventories) and the broader supply chain over time (output, shipments, and inventories) resulting from disruptions to firms and transportation networks [2], [3]. N-ABLE™ draws on the results of the other analysis tools and subject matter expertise to define disruption parameters and simulate individual firm behaviors within the modeled supply chains. Figure 2 shows a representation of the interaction of a typical N-ABLE™ enterprise firm, containing different types of decision makers with objectives, interaction with each other, with supporting CIKR, and with upstream and downstream 'markets' for input commodities and output products. Entire supply chains are constructed from collections of firms, based on this enterprise design, with each participating firm interacting with others through markets and physical infrastructure.
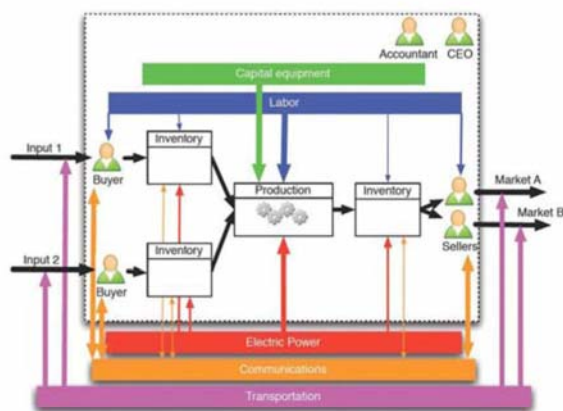
Figure 2. N-ABLE™ Enterprise Model of an Economic Firm.

N-ABLE™ simulation results provide quantitative and qualitative information necessary for consequence analyses. For example, if a hurricane temporarily shuts down a set of chemical production facilities, N-ABLE™ estimates economic impacts resulting from a decreased chemical supply to downstream facilities (e.g., customers of the closed facilities, the customers of the customers, etc.). N-ABLE™ also estimates losses resulting from a decreased demand of input chemicals used by the closed production plants to upstream facilities (e.g.,

suppliers to the closed plants, suppliers of the suppliers, etc.). These economic impact and loss estimates can be used to measure the systemic impacts to the chemical supply chain from a hurricane.

In addition, N-ABLE™ estimates the time necessary for the system to recover from a disruption. In the case of chemical supply chains, supply interruptions can cascade through many other sectors at different rates. Some downstream consumers will feel the impact of interrupted production immediately, some will not feel the impact until days or weeks later, and some will not feel it at all. Inherent in N-ABLE™ is the capability to represent the search for other supply sources when losses occur and any changes in transportation costs associated with the need to use alternate suppliers. The cost estimates associated with the recovery and adaptation processes are crucial to estimating supply chain recovery processes.

2.3   Resilience Analysis Framework

A uniform, methodical approach for assessing resilience of infrastructure systems is required to successfully incorporate resilience into critical infrastructure protection (CIP) policies and business planning practices. This approach needs to be general enough to apply to all types of infrastructure systems to account for dependencies between different infrastructure types and establish standards across all infrastructure types. Furthermore, resilience assessment approaches should explicitly account for the costs of recovery processes in comprehensive disruption cost evaluations.

With these two requirements in mind, Sandia has developed a novel framework for evaluating the resilience of infrastructure and economic systems [4]. The framework includes a new definition of resilience, a mathematical resilience cost measurement approach, and a qualitative analysis methodology that assesses system characteristics that affect resilience. This

511

framework can be applied to studies of natural and manmade disruptions.

The framework as developed presents a mathematical resilience cost measurement approach that can be used to objectively determine the impacts of disruptions on a system and the resilience costs associated with disruptions. The resilience cost measurement approach requires quantification of two key components of the definition of system resilience: systemic impact (*SI*) and total recovery effort (*TRE*). *SI* is the impact that a disruption has on system productivity and is measured by evaluating the difference between a targeted system performance (*TSP*) level and the actual system performance (*SP*) following the disruption. *TRE* refers to the efficiency with which the system recovers from a disruption and is measured by analyzing the amount of resources expended during the recovery process. The measurement of system resilience costs requires the quantification of both *SI* and *TRE*.

Figure 3 graphically represents systemic impact for a hypothetical system that has been disrupted. In this example, system performance decreases immediately following the disruption shock. With the onset of recovery actions, performance levels eventually increase and ultimately attain targeted system performance levels. At this point, recovery is considered complete. *SI* is quantified by calculating the area between the *TSP* and the actual *SP* curves in Fig. 3. This area is calculated using the formula in Eq. (1).
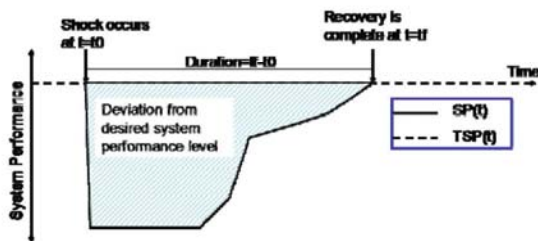

Figure 3. Systemic Impact (SI).

$$SI = \int_{t_0}^{t_f} [TSP(t) - SP(t)] \, dt \qquad \text{Eq. (1).}$$

Figure 4 illustrates the recovery response for the system shown in Fig. 3. After the disruption initiates, the recovery response begins and resources are expended in this effort. The *TRE* is the cumulative amount of resources expended during the recovery period and is represented by the area under the recovery effort (*RE*) curve in Fig. 4. This area is calculated by Eq. (2).
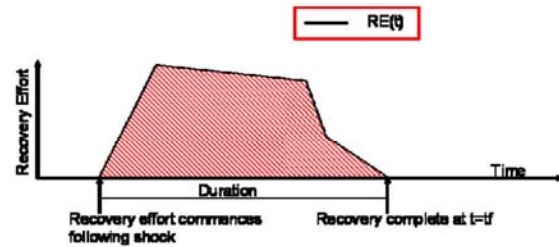

Figure 4. Total Recovery Effort (TRE).

$$TRE = \int_{t_0}^{t_f} [RE(t)] dt \qquad \text{Eq. (2).}$$

System performance is determined by the *RE*. That is, different *RE*s lead to different system performances. For example, if no *RE* is made following the disruption, the loss of system performance may be great. In contrast, if recovery resources are deployed shortly after the system shock, system performance may not be significantly affected, and *SI* may be small. The recognition that *SI* is implicitly determined by the selected recovery strategy leads to the development of recovery-dependent resilience (*RDR*) cost measurements. *RDR* costs are the resilience costs of a system under a particular recovery strategy and are calculated with Eq. (3).

$$RDR(RE) = \frac{SI + (\alpha \times TRE)}{\int_{t_0}^{t_f} |TSP(t)| dt} \qquad \text{Eq. (3).}$$

RDR costs are linear combinations of SI and *TRE*. The denominators in Eq. (3) are normalization factors that permit the comparison of the resilience of systems of different magnitudes. Because resilience represents a balancing of *SI* and *TRE* costs,

the calculation of *RDR* costs includes the parameter α, which is a weighting factor that allows an analyst to assign the relative importance of the systemic impact and total recovery effort terms. Assigning a small positive value to α weighs the systemic impact more heavily; a large positive value for α weighs the cost of recovery more heavily. To equally weigh *SI* and *TRE*, α is set to 1.

In addition to RDR costs, optimal resilience (*OR*) costs can also be considered, but their calculation is beyond the scope of this work at present.

When applied to the CDM, N-ABLE™ simulations can provide quantitative and qualitative information necessary for resilience analyses. For example, if a hurricane temporarily shuts down a set of chemical production facilities, N-ABLE™ can estimate economic impacts resulting from a decreased chemical supply to downstream facilities (e.g., customers of the closed facilities, the customers of the customers, etc.). N-ABLE™ can also predict losses resulting from decreased demand of input chemicals used by the closed production plants to upstream facilities (e.g., suppliers to the closed plants, suppliers of the suppliers, etc.). These economic impact and loss estimates can be used to measure the *SI*s to the chemical supply chain from a hurricane.

In addition, N-ABLE™ can predict how the chemical sector will adapt to and recover from a disruption. The tool has the capability to estimate production curtailments by the customers of the closed plants that cannot find new suppliers, the higher transportation costs associated with new suppliers, the use of chemical substitutes, and the implementation of different production technologies and recipes to adapt to a disruption. The cost estimates associated with the recovery and adaptation processes are crucial to calculating the *TRE* in a resilience analysis.

3.0 DISCUSSION

3.1 Analysis Basis

The methodology, models, data, and other capabilities described above have been applied to a variety of homeland security problems. The following summary of an analysis of a Category 3 hurricane making landfall in the Gulf Coast is an example of this application.

The scenario hurricane was patterned after the actual Hurricane Ike (2008), which developed during the early part of September and made its landfall over coastal Texas on September 13, 2008. The storm moved at a projected forward speed of 12 miles per hour (mph), carrying maximum sustained winds of over 100 mph. The storm size was approximately 230 miles, and it was the third most destructive hurricane in U.S. history. Figure 5 shows the projected path of Hurricane Ike early on September 11, 2008.

Figure 5. Projected Path of Hurricane Ike, NOAA Advisory 41, 0400 CDT September 11, 2008.

Sandia used the scenario storm parameters (trajectory and category) in this analysis to first estimate the damages from wind and surge waters. These damage estimates are translated into areas of probable electric power outage and inland flooding depths. Sandia analysts then assessed potential direct impacts to chemical facilities,

petroleum refineries, and the natural gas network for elements physically affected by this scenario storm. Analysts then assessed the indirect impacts to facilities and infrastructures not in the path of the hurricane but dependent on facilities within the disruption area. Finally, analysts estimated cascading impacts to the chemical industry and petrochemical supply chain at a regional, national, and global level. Figure 6 shows the estimated electric power disruption area for the scenario hurricane. Differences in color reflect the likelihood of power outage (green reflecting a 0- to 25-percent probability of outage, red representing a 75- to 100-percent probability of power outage), while intensity of color reflects the projected duration of disruption (lighter shades representing shorter duration of outage where present, darker shades representing longer duration of outage where present).
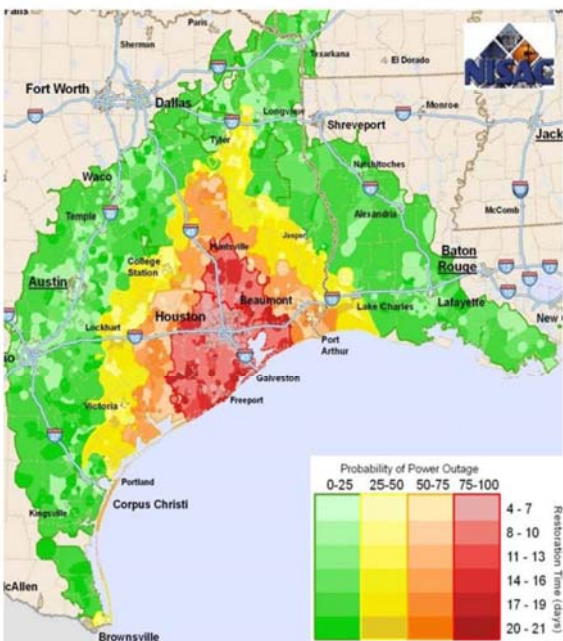
Figure 6. Estimated Disruption Area of the Scenario Hurricane.

It is common practice for Gulf Coast petrochemical production facilities in the projected path of a hurricane to shut down operations 48 hours prior to hurricane landfall. On average, the petrochemical facilities within the electric power outage contours will be without power for a few weeks. Production at these facilities will not likely be restored immediately following restoration of power. Following a plant shutdown, petrochemical facilities often require additional startup time to perform system checks, such as purging pipelines and vessels with inert gases such as nitrogen, to ensure the unit's operability. To simulate the cumulative effects of these, analysts assumed that all petrochemical facilities within the outage contours are shut down for 25 days.

To quantitatively evaluate the resilience of the petrochemical supply chain, we ran two sets of N-ABLE™ simulations. In the baseline scenario, we assumed no disruptions. In the disruption scenario, we assumed that a hurricane is projected to make landfall on day 202 of the simulation and the electric power outage shown in Figure 6 is expected to occur. On day 200, all petrochemical facilities within the contours shut down in anticipation of the storm. Normal production capabilities are assumed to return on day 225 of the simulation.

The market value of production (MVP) is the metric used to measure SI. MVP captures total "street value" of every step of chemical-unit production. It is similar to the sale value of end products, but it counts production at every stage in the production process, whereas the sale value only counts chemicals that are sold on the merchant market. MVP equals sale value of end products if there is absolutely no vertical integration, i.e., outputs of every stage of the production process are sold on the merchant market

For this analysis, two factors are considered in determining TRE: additional aggregate transportation costs (TC) and production plant shutdown/restart costs (RC). When a disruption decreases the supply of available chemicals, consumers of those chemicals will seek new suppliers. These suppliers will likely be farther from the consumers than

514

the original suppliers, so the cost of transporting chemicals from the new suppliers will likely be greater due to the increased transportation distances

Cost engineering estimates RCs as a percentage of the capital costs of the equipment involved. Pre-planned, short-term shutdowns are generally less expensive, based on available data. After literature review, consultation with project subject matter experts and economists at the National Center for Risk and Economic Analysis of Terrorism Events (CREATE) and the American Chemistry Council (ACC), the authors utilized an *RC* of 3 percent of capital costs.

For the sake of simplicity, we only consider the *TC*s and *RC*s when calculating the *TRE* for this example. To calculate *RDR* costs, we set α to 1 in Eq. (3) and approximate the integral with 1-day time-step intervals because N-ABLE™ reports data on a daily basis.

Figure 7 shows *MVP* as a function of time for the base case and the scenario hurricane for the whole Ethylene supply chain. Utilization of inventories (in hand and in transit) helps to buffer some of the effects of the disruption.
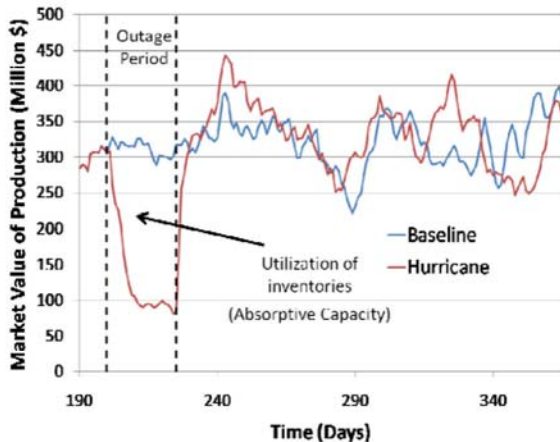

Figure 7. *MVP* as a function of time, Ethylene Supply Chain, Base and Hurricane Scenarios.

Figure 8 shows average shipment distance as a function of time for the base case and the scenario hurricane for the whole Ethylene supply chain. The inventory utilization described in Figure 7 comes at a cost, which reflects through in the calculation of *TC*, and as a result, on *TRE*.
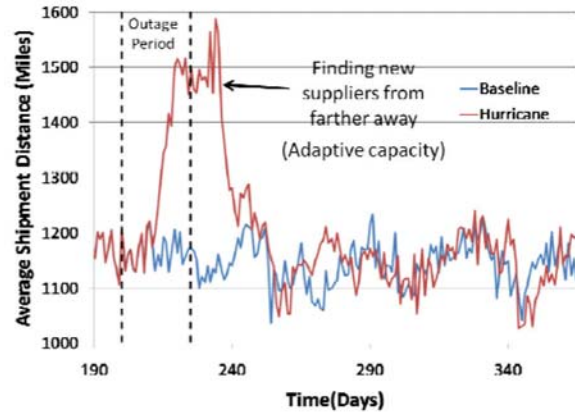

Figure 8. Average Shipment Distance as a Function of Time, Ethylene Supply Chain, Base and Hurricane Scenarios.

Table 1 shows the comparative analysis of the calculation of System Resilience for the whole Ethylene supply chain and for a segment, Vinyl acetate monomer (VAM). Impacts to VAM production are more severe than the aggregate, transportation distances and costs greater, and recovery period longer. As such, the VAM resilience metric is larger by one-third than that of the Ethylene supply chain as a whole (here, a lower value reflects a more resilient system).

Table 1. Comparison of Resilience Values, Hurricane Scenario, VAM and Ethylene Supply Chains

Measure	VAM	Ethylene
Target *MVP* ($M)	856	49,000
SI ($M)	88	4,000
TRE: RC ($M)	11	256
TRE: TC ($M)	1.5	254
Resilience	.12	.09
Resilience = (*SI* + (*RC* + *TRE*))/Target *MVP*		

A more detailed discussion of the consequence and resilience analysis results

for this scenario will be presented at MODSIM World 2010.

4.0 CONCLUSION

Analysis of chemical supply chains is an inherently complex task, given the dependence of these supply chains on multiple infrastructure systems (e.g. transportation and energy). The capability developed at Sandia is intended to provide information to the DHS with respect to the consequences of large-scale disruptions to the chemical sector, including interrupted supply and resulting economic impacts to the nation, which can be utilized to inform response and recovery officials, enabling more effective pre-event planning and more knowledgeable event response. The ongoing development effort includes the development of several tools along with a comprehensive database that feeds the tools. The database is constructed by merging many datasets in combination to provide a high degree of resolution within the data so that individual plants can be uniquely represented.

The hurricane disruption scenario presented herein shows that large-scale disruptions to petrochemical supply chain elements affect many supply chains and, consequently, take considerable time to recover (Figures 7 and 8). Supporting this result in the scenario analysis, information reported in Chemical Week showed that

> Several Texas Gulf Coast chemical plants began to restart operations after shutting down ahead of Hurricane Ike's landfall on September 13, 2008. However, producers claim that the ready availability of utilities, raw materials, and logistics, and the damage at some customer sites negatively affect their effort to restart operations [5].

The disruption to chemical plants cascade both up and down the supply chain, affecting recovery efforts.

5.0 REFERENCES

[1] U.S Department of Homeland Security, "Infrastructure Protection Plan: Partnering to enhance protection and resiliency," 2009, http://www.dhs.gov/xlibrary/assets/NIPP_Plan.pdf.

[2] Eidson, E. D., and M. A. Ehlen, 2005. "NISAC Agent-Based Laboratory for Economics (N-ABLE™): Overview of Agent and Simulation Architectures." Sandia National Laboratories, SAND Report SAND2005-0263, February 2005.

[3] Ehlen, M. A., A. J. Scholand, and K. L. Stamber, 2007, "The Effects of Residential Real-Time Pricing Contracts on Transco Loads, Pricing, and Profitability: Simulations using the N-ABLE™ Agent-Based Model," Energy Economics, 29:2, pp. 211-227, 2007, http://www.sandia.gov/nisac/docs/CERTS.pdf.

[4] Vugrin, E. D., D. E. Warren, and M. A. Ehlen, 2010. "A Resilience Assessment Framework for Infrastructure and Economic Systems: Quantitative and Qualitative Resilience Analysis of Petrochemical Supply Chains to a Hurricane," American Institute of Chemical Engineers, 2010 Spring Meeting, 6th Global Congress on Process Safety, San Antonio, TX, March 22-24, 2010.

[5] Coons, Rebecca, "Hurricane Ike Batters Texas Gulf Coast Production," Chemical Week, September 22, 2008, http://media.web.britannica.com/ebsco/pdf/34/34625778.pdf.

6.0 ACKNOWLEDGMENTS

MODSIM WORLD
Conference & Expo

Consequence and Resilience Modeling for Chemical Supply Chains

Kevin L. Stamber, Eric D. Vugrin, Mark A. Ehlen, Amy C. Sun, Drake E. Warren, and Margaret E. Welk
Sandia National Laboratories
klstamb@sandia.gov

NNSA
National Nuclear Security Administration

Sandia National Laboratories

MODSIM WORLD
Conference & Expo

Acknowledgements

U.S. Department of Homeland Security
Science and Technology Directorate

MODSIM WORLD
Conference & Expo

- Introduction
- Technical Elements
  - Chemical data model (CDM)
  - Consequence Analysis Tools
  - Resilience Analysis Framework
- Measuring Resilience
- Summary
- Questions

Introduction

MODSIM WORLD
Conference & Expo

- Chemical sector is highly connected to multiple infrastructures and commercial sectors
- Consequence analysis capability must consider
  - Disruptions of the chemical sector
  - Disruptions of interdependent infrastructures
 - National, regional, and facility perspectives
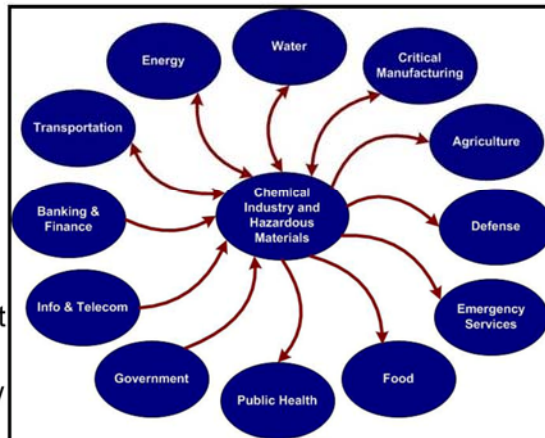
Introduction

## Supply chains face an array of threats

- "Protection, in isolation, is a brittle strategy"
- Effective integration of resilience into critical infrastructure protection policies requires
  - Consistent, broadly applicable definitions and methods
  - Objective methods for measuring progress
  - Comprehensive accounting of resource-constrained recovery strategies
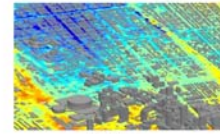
Flooding
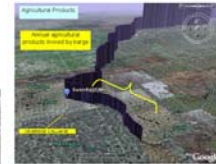
Commodity Flow Impacts
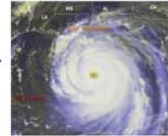
Hurricane

"We are working every day to ensure our country stands ready to respond to any disaster or emergency -- from wildfires and hurricanes, to terrorist attacks and pandemic disease. Our goal is to ensure a more *resilient* Nation."

–President Barack Obama, September 4, 2009

Introduction

- We need to address direct impact questions
  - What is the area of direct impact?
  - What chemical facilities are directly affected?
  - What percentage of capacity does this represent?
- And cascading impact questions
  - How long before we return to 'normal'?
  - What additional facilities will be affected?
- We also need to be able to examine systemic resilience
  - Define
  - Calculate

Introduction

| Organization	Role
DHS Science and Technology (S&T) Directorate, Infrastructure and Geophysical Division	Manage the chemical supply chain and resilience project
Sandia National Laboratories, Interdependencies and Consequence Management Group	Develop analysis and design capabilities
National Infrastructure Simulation and Analysis Center (NISAC) (managed by DHS Office of Infrastructure Protection [IP])	Apply completed capabilities to disruptions of critical infrastructures and key resources (CIKRs)

Technical Elements

- Chemical Data Model (CDM)
- Consequence Analysis Tools
- Resilience Analysis Framework

Actionable Analysis

Software

Models

Chemical Sector Database
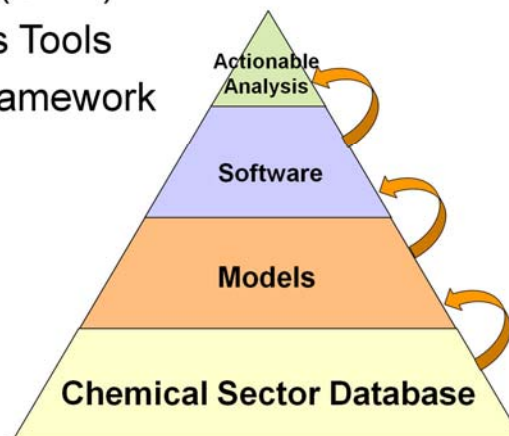
- Data foundation for the project
  - All models are driven from the same set of core input data
  - Differences in model output are due to modeling approach
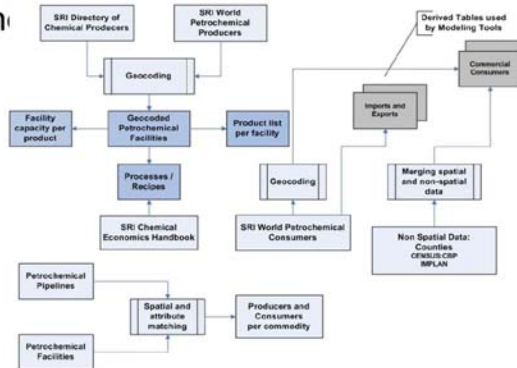
- Project models and analysis tools need the following information

  - Plant facilities
    - Name
    - Location (address and geocodes)
  - Facility productions
    - Chemical types
    - Quantities
    - Processes
  - Infrastructure dependencies
    - Transportation (rail, pipeline, etc.)
    - Energy (electric power, natural gas, petroleum products)
    - Quantities

  - Consumption
    - Categories
    - Locations
    - Quantities
  - Imports/exports
    - Locations
    - Quantities
  - Other factors
    - Economics
    - Population distribution
    - Emergency services

| Dataset Name	Provider
World Petrochemicals Program 2009	SRI Consulting
Chemical Economics Handbook 2009	SRI Consulting
Directory of Chemical Producers 2009	SRI Consulting
Oil & Gas Pipelines	NGA HSIP Gold 2008 (PennWell)
Oil & Gas Facilities	NGA HSIP Gold 2008 (PennWell)
United States Census 2000	U.S. Census Bureau
County Business Patterns 2007	U.S. Census Bureau
County Business Patterns Employees Estimation 2007	U.S. Census Bureau

Dataset Name	Provider
Geographic Names Information System	U.S. Geological Survey
IMPLAN States Summary 2002	Minnesota IMPLAN Group
International Trade Statistics 2007	U.S. Department of Commerce
Refinery Location Data	Argonne National Laboratory*
2005 Commodity Flow Survey, Department of Transportation	2005 Waybill Sample, Surface Transportation Board
2007 Class I Railroad Statistics, Association of American Railroads	2007 Producer Price Index, Department of Labor
E-Plan Emergency Response Information System	U.S. Environmental Protection Agency/U.S. Department of Homeland Security

* Argonne data were updated using 2007 domestic data from the Energy Information Administration (EIA) and foreign data from SRI Consulting.

Technical Elements: CDM

- CDM Building Process
  - Gather
  - Process and integrate
    - Merge datasets into a common, Oracle-based framework
  - Authenticate
  - Document
  - Ensure traceability
  - Test
    - Ensure compatibility with models
  - Iterate

Trade Data | EPLAN | SRI CEH | SRI WP | SRI DCP

Chemical Data Model (CDM)
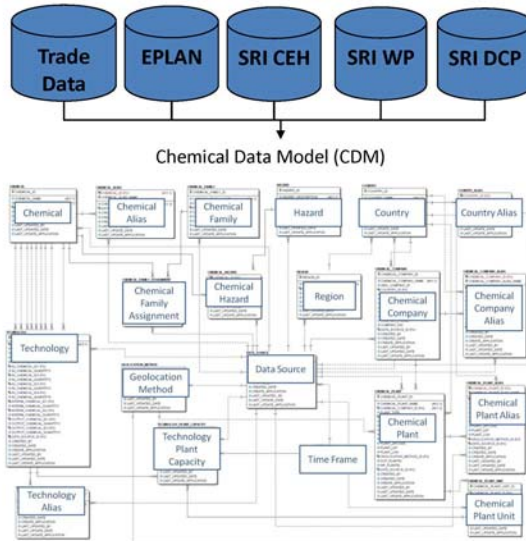
Technical Elements: Consequence Analysis Tools

FAIT

Loki

FASTMap

R-NAS

N-ABLE™

Information Sharing   Network Analysis   Time to respond
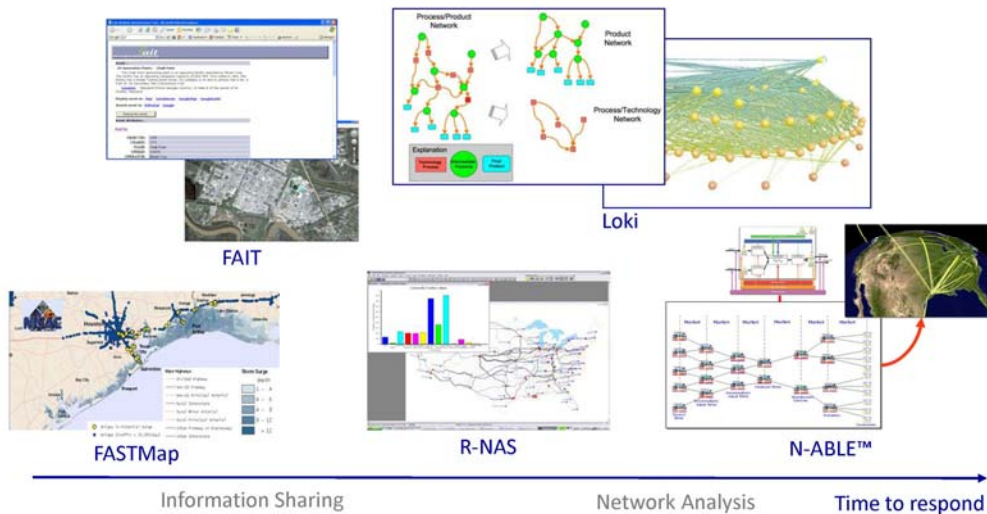
Dynamic Supply Chain Analysis with the NISAC Agent-Based Laboratory for Economics (N-ABLE™)

MODSIM WORLD Conference & Expo

Individual enterprises are combined to create networks of enterprises
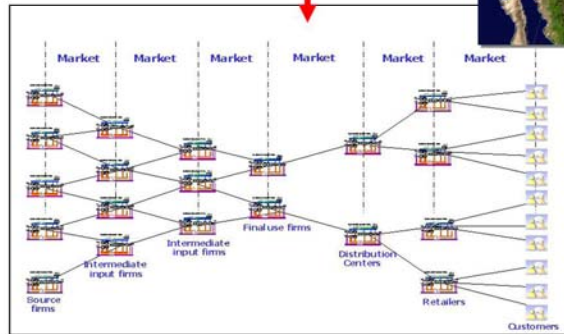
The networks of enterprises comprise national, regional, and local markets

Model Foundation of N-ABLE™: The Enterprise Firm

MODSIM WORLD Conference & Expo

- N-ABLE™
  - Generates data-driven microeconomic "enterprises"
  - Simulates enterprise operations (buyers, production, sellers, inventories, and shipping)
  - Identifies interactions in markets and dependencies on critical infrastructures
  - Estimates how enterprises respond individually and collectively to disruptions

524
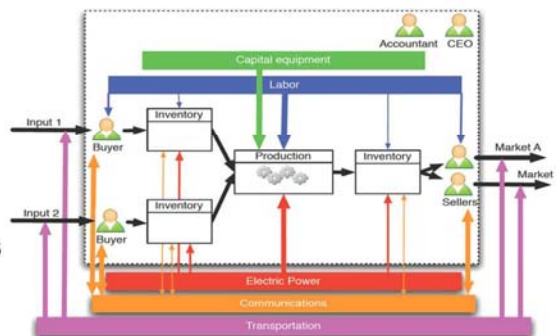
Resilience Framework:
A Definition of Resilience

- Resilience is contextual
  - Relative to disruptive event and performance targets
- System performance is a fundamental factor
  - Structure not as important as performance
  - We consider magnitude and duration
  - We do not assume a system will return to pre-disruption state
- Resource expenditure in recovery processes a fundamental consideration
  – We consider the ability to efficiently reduce system impacts to absorb, adapt, and/or recover

"Given the occurrence of a particular, disruptive event (or set of events), the resilience of a system to that event (or events) is the ability to efficiently reduce both the magnitude and duration of the deviation from targeted system performance levels."

-Vugrin et al., 2010

Resilience Framework: Calculation of Resilience Costs

Shock occurs at t = t0 Duration = tf - t0 Recovery is complete at t = tf

System Performance

Systemic Impact (SI)

time

SP(t)
TSP(t)

Recovery Effort

Total Recovery Effort (TRE)

RE(t)

Duration

time

Recovery effort commences following shock

Recovery is complete at t = tf
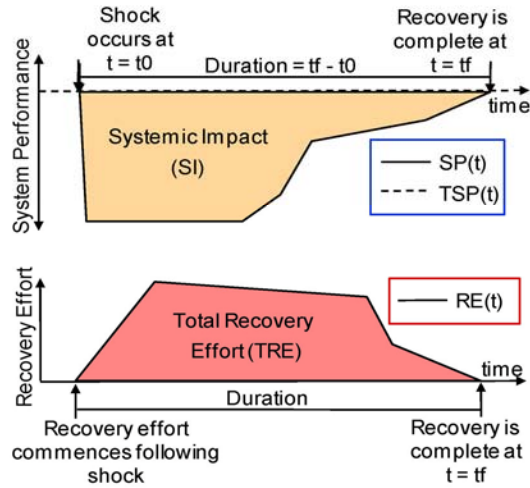
Resilience Costs

=

$$\frac{SI + \alpha \times TRE}{\int_{t0}^{tf} |TSP(t)| \, dt}$$

Resilience Framework: Qualitative Resilience Assessment

Resilience

=

	Systemic Impact		Total Recovery Effort
Component		+	
Determining Features	Absorptive Capacity	Adaptive Capacity	Restorative Capacity
Distinguishing Characteristics of Capacity	Considers aspects that automatically manifest after the disruption	Considers internal aspects that manifest over time after the disruption	Considers ability to affect and repair internal system features
Effort Required	Automatic/ Little Effort	Internal Effort Required	External Effort Often Required
Measurement of Component	Internal Measurement	Exogenous Measurement	

526

Resilience Framework: Resilience Analysis Process

- To apply the conceptual framework to the chemical sector, we must define several key components of the analysis process, such as
  - Chemicals under consideration
  - Performance metric
- We require subject matter expertise for this process
- We will demonstrate the resilience analysis process for a scenario hurricane

1. Define System(s)
⇕
2. Define Scenario(s)
⇕
3. Define Metrics
⇕
4. Obtain Data
⇕
5. Calculate Resilience Costs
⇕
6. Perform Qualitative Assessment
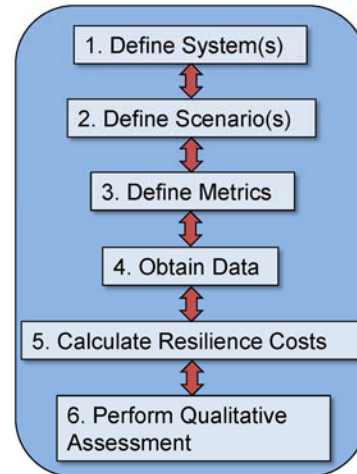
Measuring Resilience: Scenario Assumptions

- Hurricane makes landfall and affects plants in electric power (EP) outage contours
- Facilities within the outage contours shut down 2 days prior to landfall
- These facilities are nonoperational for an additional 23 days
- All facilities that are within outage contours require startup processes
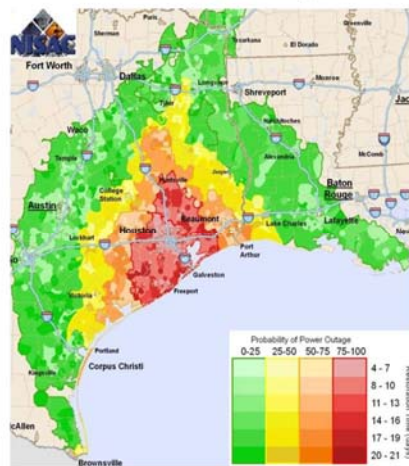
Estimated Power Outages & Restoration Times

527

MODSIM WORLD
Conference & Expo

- Market value of production (MVP)
  - Total "street value" of every step of production for all chemicals and facilities

$$MVP(t) = \sum_i \sum_j Q_{i,j}\left(t\right) \times p_i$$

Chemical price
per unit mass

Mass produced
of chemical

  - Systemic impact metric = MVP for disrupted conditions
  - Targeted system performance = MVP for undisrupted conditions

MODSIM WORLD
Conference & Expo

- Total recovery effort metric 1:
  - Additional transportation costs (TC) due to increased transport distances

Increased average
distance/shipment

$3/car-mile

$$TC(t) = \left[\sum_i MD\left(i,t\right) \right] \times \left[D_{ave}^D\left(t\right) - D_{ave}^B\left(t\right) \right] \times C_{car} \times C_{cost}$$

Met demand for a
chemical (short tons)

(1 car /100 short tons)

Measuring Resilience: Defining Metrics (continued)

- Total recovery effort metric 2:
  - Production plant shutdown/restart costs: cost engineering estimates these cost as a percent of capital costs
  - Pre-planned, short-term shutdown is generally less expensive
  - After consultation with project chemical subject matter expert, literature, and economists at National Center for Risk and Economic Analysis of Terrorism Events (CREATE) and American Chemistry Council (ACC), we use <u>3 percent of capital costs to estimate shutdown/restart costs</u>

$$RC = 0.03 \times \sum_{j} CC_j$$

Capital cost per plant

Restart Costs as a Percent of Capital Costs

Source	Range	Median
Perry (2008)	-	3%
Peters and Timmerhaus (1968)	0.5-2%	1.3%
Peters and Timmerhaus (1980)	8-10%	-
Price (2009)	5-20+%	-

Measuring Resilience: Calculation of Resilience Costs

- Recall that

$$\text{Resilience Costs} = \frac{\text{System Impact} + \alpha \times \text{Total Recovery Effort}}{\text{Targeted System Performance}}$$

- Therefore, for this analysis:

$$RC = \frac{MVP_{Baseline} - MVP_{Disrupted} + TC + RSC}{MVP_{Baseline}}$$

**Measuring Resilience:
Obtaining Data through N-ABLE™
Simulations**

- Two sets of N-ABLE™ 1-year simulations were executed
  - Baseline and disrupted conditions
- In the disrupted simulation
  - Plant shutdown is assumed to occur on day 200, and
  - Affected plants are assumed to be fully operational on day 225
- Simulations provide MVP and TC data
- Restart costs (RC) are estimated external to the simulation

MODSIM WORLD
Conference & Expo

**Measuring Resilience:
Systemic Impact,
Whole Ethylene Supply Chain**
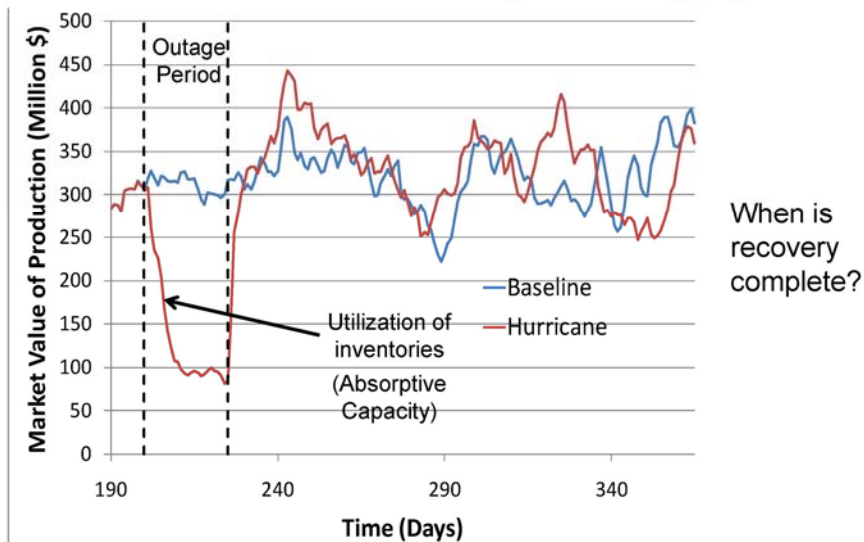
Measuring Resilience: Adaptive Behaviors, Whole Ethylene Supply Chain
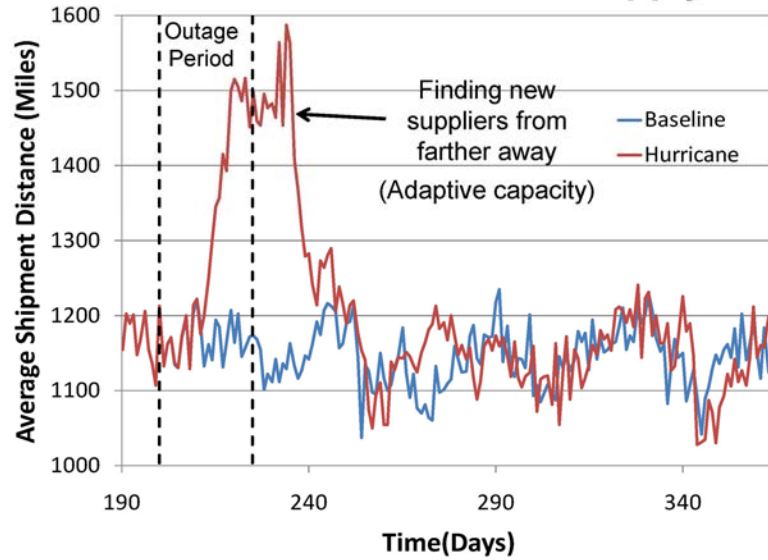
Measuring Resilience: Calculating Resilience Costs

Measure	VAM	PVC Chain	Entire Chain
Recovery "Complete" on Day	264	260	250
Target MVP ($M)	856	14,800	49,000
Systemic Impact ($M)	88	1,100	4,000
Recovery Effort: Restart ($M)	11	23	256
Recovery Effort: Transportation ($M)	1.5	9.6	254
Resilience Cost	.12	.08	.09
Resilience Cost = (Systemic Impact + Total Recovery Effort)/Target MVP			

531

Measuring Resilience:
Calculating Resilience Costs

- Systemic impact dominates recovery costs in all systems
- Restart costs far outweigh increased transportation costs
- Restart and transportation costs for VAM are relatively high
- Though the VAM system is the smallest, it is also the least resilient
  - Simplicity may hinder resilience

Summary

- This project takes a multidisciplinary approach to chemical supply chain modeling and resilience analysis
- We have integrated our consequence analysis capabilities into a resilience framework to enhance analytic capabilities
- We plan to continue capability development efforts this year
- We welcome, encourage, and value feedback

MODSIM WORLD Conference & Expo

Acknowledgements

- The authors would like to thank John Masciantoni for providing N-ABLE™ assistance.
- The authors would like to thank Adam Rose and Tony Barrett from the National Center for Risk and Economic Analysis of Terrorism Events (CREATE) for their thoughtful insights on this work.

MODSIM WORLD Conference & Expo

Questions?

4.5 Integrated Modeling, Mapping, and Simulation (IMMS) Framework for Exercise and Response Planning

Integrated Modeling, Mapping, and Simulation (IMMS) Framework for Exercise and Response Planning

Jalal Mapar
Science and Technology Directorate
Department of Homeland Security
jalal.mapar@dhs.gov
Trisha Hoette: tmhoett@sandia.gov
Karim Mahrous: kmmahro@sandia.gov
Carmen M. Pancerella: carmen@sandia.gov
Todd Plantenga: tplante@sandia.gov
Christine Yang: clyang@sandia.gov
Lynn Yang: liyang@sandia.gov
Sandia National Laboratories
Michael Hopmeier: hopmeier@unconventional-inc.com
Unconventional Concepts, Inc.

Abstract. Emergency management personnel at federal, state, and local levels can benefit from the increased situational awareness and operational efficiency afforded by simulation and modeling for emergency preparedness, including planning, training and exercises. To support this goal, the Department of Homeland Security's Science & Technology Directorate is funding the Integrated Modeling, Mapping, and Simulation (IMMS) program to create an integrating framework that brings together diverse models for use by the emergency response community. SUMMIT, one piece of the IMMS program, is the initial software framework that connects users such as emergency planners and exercise developers with modeling resources, bridging the gap in expertise and technical skills between these two communities. SUMMIT was recently deployed to support exercise planning for National Level Exercise 2010. Threat, casualty, infrastructure, and medical surge models were combined within SUMMIT to estimate health care resource requirements for the exercise ground truth.

1.0 INTRODUCTION

1.1 Exercise Management

Exercises provide a way to assess and validate the speed, effectiveness and efficiency of emergency management capabilities, and test the adequacy of policies, plans, procedures, and protocols of emergency response in a risk-free environment. In 2002, the U.S. Department of Homeland Security (DHS) and Federal Emergency Management Agency (FEMA) developed policies, incorporated into the Homeland Security Exercise and Evaluation Program (HSEEP) [1], to guide the design, development, conduct, and evaluation of exercises. This served as an opportunity to standardize the language and concepts used in the exercise planning and evaluation process for the homeland security community.

Under DHS leadership, the National Exercise Program (NEP) [2,3] provides a framework for prioritizing and coordinating federal, regional and state exercise activities, without replacing any individual department or agency exercises. NEP defines four tiers of exercises:

Tier I: White House directed, U.S. Government-wide Strategy & Policy Focus, Full Participation

Tier II: Federal Strategy & Policy Focus, Significant Simulation

Tier III: Other Federal Exercises, Operational, Tactical or Organizational Focus, Simulation

Tier IV: State, Territorial, Local, Tribal or Private Sector Focus

Typically, Tier I and II are supported by the FEMA's National Exercise Division (NED). Each year one exercise is designated as the National Level Exercise (NLE), a Tier I event requiring senior level participation among the Federal interagency community. NLEs are full-scale exercises that are typically scheduled five years prior, and are

planned for up to two years in advance. A full-scale exercise is a multi-agency, multi-jurisdictional, multi-discipline exercise involving functional and "boots on the ground" response.

The role of modeling and simulation (M&S) in HSEEP exercises is still evolving. Though dozens of federally-funded modeling efforts have been identified that are relevant for emergency response to hazards and threats and that could greatly enhance exercise planning and conduct, there is no formal mechanism for the emergency management community to discover, access and use these M&S capabilities. Furthermore, the wide range and disparity of M&S capabilities for emergency management necessitates a way to integrate and run models.

1.2 IMMS Program
Realizing the opportunities to enhance exercises and planning through modeling and simulation, the Department of Homeland Security Science and Technology Directorate (DHS/S&T) spearheaded the Integrated Mapping, Modeling and Simulation (IMMS) program.[1] IMMS is a research and development effort to develop a common framework for integrating incident-related M&S tools to enhance situational awareness and operational efficiency of emergency managers and exercise planners. Among the many applications are support to exercises, including Tiers I-IV. Recognizing that the modeling and simulation community is large and highly diverse, DHS/S&T is providing, through IMMS, the M&S and exercise communities a way to discover models, integrate them quickly and economically, and apply them in analyses to improve exercise, planning, and response

---

[1] "High-Priority Technology Needs", Department of Homeland Security Science and Technology Directorate, Version 2.0, June 2008. The Incident Management Integrated Product Team (IPT), led by FEMA and the Office of Emergency Communications, identified an integrated modeling, mapping and simulation capability as a high-priority technology.

efforts. The central technological component of IMMS is the Standard Unified Modeling and Mapping Integration Toolkit (SUMMIT), which connects users such as emergency planners and exercise developers with modeling resources in an easy-to-use format. This paper provides a high-level overview of the SUMMIT architecture, and a description of IMMS support to National Level Exercise 2010 (NLE10). NLE10 lessons learned from M&S support of exercises are being incorporated as requirements and concepts of operation (ConOps) for SUMMIT, and are being used to develop and refine the IMMS vision.

2.0 BODY

2.1 SUMMIT
To create a capability for linking together M&S tools, SUMMIT is being iteratively designed and prototyped. SUMMIT provides a platform-neutral framework that brings together distributed M&S codes and a wide range of users. The framework makes it easier to discover and integrate models, provision them for a specific scenario, execute models on available resources, and deliver results to a collaborating set of users.

The SUMMIT architecture allows for considerable flexibility, placing few restrictions on federated models, but still providing necessary capabilities for integration. Model owners decide who has permission to use their code, and where it is hosted. Exercise planners and other emergency response users link models as needed to address specific scenarios.

For example, suppose the exercise scenario is the release of chlorine gas from a railcar in an urban setting, and the emergency responder wants to know if there are sufficient medical supplies for first response. A computational approach (Fig. 1) might incorporate a finite element model that computes the chemical gas dispersion plume for given weather conditions, another model that quantifies casualties in the

population at risk, and a third model that tallies available medical resources. SUMMIT makes it possible for the planner to link these three models, execute them, and view results, all from a single client interface.

Figure 1 depicts a SUMMIT simulation template, an abstraction that shows how models are connected to address a specific scenario. Although the template displays only a high level view, "under the hood" there is sufficient detail to define a software federation of models that can execute automatically.
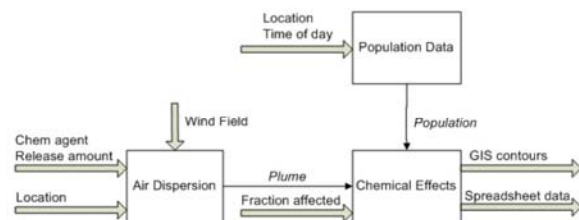

Figure 1. Example simulation template linking three models (boxes) via data flows (arrows).

Software components in SUMMIT are divided between a central server, user clients, and executable models that are often (but not always) hosted on the machines of model owners. Figure 2 shows the main components of the architecture. The "data" and "model" icons in the figure are components owned by model contributors. All other icons are part of the SUMMIT framework. A SUMMIT Client component allows for interaction with the user, and a SUMMIT Software Development Kit (SDK) provides tools for model owners to integrate their models and verify that they are SUMMIT-compliant. Components in the SUMMIT core server provide system functionality through a set of distributed services.
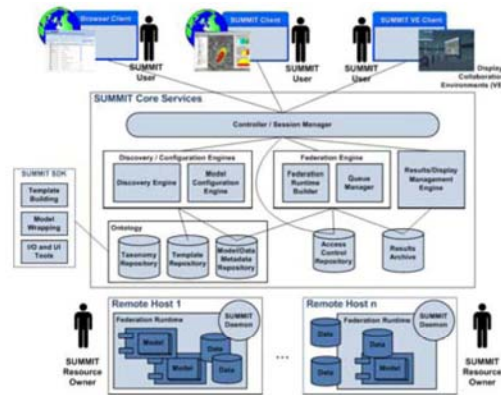

Figure 2. SUMMIT architecture.

SUMMIT provides for three types of users: emergency responders and exercise planners (the primary end user), model owners, and scenario planners. *End users* access content through a SUMMIT Client. They log into the system, discover an appropriate simulation template (such as Fig. 1), configure inputs, and then view results after SUMMIT automatically executes the models that compose the simulation template. *Model owners* use SUMMIT SDK tools to create a software wrapper that enables execution of their model as part of a SUMMIT-mediated federation of models. (Note that in this paper the term "federation of models" does not refer to a High Level Architecture [4] type of federation, but to a collection of models run consecutively with interconnected data.) For example, the three models in the chlorine gas scenario described earlier might be contributed by three different model owners and hosted at three different remote sites. *Scenario planners* use SUMMIT SDK tools to create simulation templates that specify inputs and needed outputs and bring together models for a specific incident scenario. In the chlorine gas example a scenario planner created the simulation template by linking three models at a conceptual level to produce the desired output.

Further details on the coordination and execution of models in the SUMMIT

framework are discussed in references [5] and [6].

SUMMIT has created a flexible external interface that allows for multiple client environments. These include a native rich client platform, a browser-based client, and interfaces to advanced commercial and government GIS technologies. Utilization of these environments allows users to grasp complex multivariate data quickly and intuitively. The IMMS team is also investigating the capabilities of virtual environment technology to bring telepresence and a heightened sense of situational awareness into exercises.

SUMMIT is also providing external interfaces so that its M&S integration capabilities can be leveraged by external tools. For example, exercise management tools will be able to access SUMMIT-archived data, allowing for even greater intra-exercise coordination.

2.2 National Level Exercise 2010

To better understand the potential utilization and role of IMMS in exercise, the IMMS team participated in NLE10 in May 2010 [7]. The objective of this first pilot was to apply the existing reference implementation of SUMMIT in a live exercise so that architecture requirements and concepts of operation (ConOps) could be evaluated and improved. The specific scenario for NLE10 was response and recovery from incidents involving an improvised nuclear device (IND) detonation in a U.S. city. The scenario was derived from National Planning Scenario 1: IND Detonation [8]. The major objectives of NLE10 were to exercise:

- Intelligence and information sharing and dissemination
- Incident Management
- Critical Infrastructure protection
- Medical Surge
- Public information
- Continuity of Operations (COOP)
- Economic and Community Recovery

NLE10 was conducted in the National Capital Region. Exercise players in NLE10 represented over 60 federal agencies, including U.S. Department of Defense, Central Intelligence Agency, Department of Energy, Department of Health and Human Services, DHS, Department of Justice, Department of State, Department of Transportation, and Environmental Protection Agency. Due to a late venue change, this NLE consisted of federal play only, and no local or state play; however, representatives from local and state governments and FEMA Regional Office V contributed to the planning process, and participated in the Simulation Cell (SimCell), providing "boots on the ground", realistic scenario injects that drove the operations-based exercise play.

Exercise conduct consisted of a Master Control Cell (MCC) releasing injects from the Master Scenario Event List (MSEL) to exercise players. The MCC included:

- A control room acting as the key node of communication, hosting both exercise controllers (who plan and manage exercise play) and evaluators (who track action relative to evaluation criteria and analyze exercise results without disturbing exercise flow) from 62 federal departments and agencies.

- A SimCell hosting representatives from the region, state, local, international, private sector, law enforcement, etc., which provided injects to and answered requests for information from the exercise players.

MSEL injects were released via phone, email, fax and the DHS Lessons Learned Information Sharing (LLIS) portal [9].

2.3 How SUMMIT Supported NLE10

SUMMIT was one of several M&S providers for NLE10, supporting both exercise planning and execution. The IMMS team used SUMMIT to integrate multiple M&S tools contributed from different agencies. Threat, casualty, and infrastructure models

and data were provided by the DOE National Atmospheric Release Advisory Center (NARAC), and DHS Homeland Infrastructure Threat and Risk Analysis Center (HITRAC) National Infrastructure Simulation and Analysis Center (NISAC). A medical surge model from the Department of Health and Human Services (HHS) Agency for Healthcare Research and Quality (AHRQ) provided health care resource surge requirements. SUMMIT was used for the pre-planned exercise ground truth, calculated prior to the exercise, and the real-time scenario injects, computed during the exercise execution.

A new SUMMIT simulation template was created specifically for the NLE10 scenario (Fig. 3). The template integrated M&S tools for nuclear effects, infrastructure effects and medical surge needs. The NLE10 template defines how models connect and the data flows between models. In the future, this template may be reused to support similar exercises at the federal, state or local level.
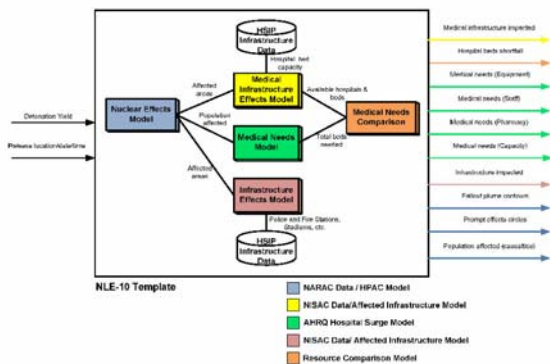

Figure 3. A representation of the NLE10 template depicting input parameters, models, and the data flow between models. Outputs are colored to match the model from which they are produced.

For exercise planning, SUMMIT was used to provide ground truth injects on the amount of surge equipment and staff that would be required in the medical response (Fig. 4). For exercise conduct, SUMMIT was used to provide real-time scenario injects on the equipment and staff needs estimated by on-scene responders.

Figure 4. Example output data from models and data federated in SUMMIT for NLE10.

3.0 DISCUSSION

The application and deployment of SUMMIT in NLE10 provided valuable lessons learned on M&S support to exercises. These lessons learned are being incorporated into SUMMIT requirements and ConOps, and will enhance SUMMIT's support of NLE11, an earthquake scenario in the New Madrid Seismic Zone based on National Planning Scenario 9: Natural Disaster – Major Earthquake [8].

NLE10 lessons learned include:

SUMMIT can facilitate the use of M&S in exercise planning.

By having an integrated framework for M&S, exercise planners were able to more easily run various scenarios in order to generate the ground truth data. Exercise planners did not have to expend time to locate the individual models, execute a series of distributed M&S tools, and gather outputs. SUMMIT enabled multiple executions to be made easily so that exercise controllers could carefully plan and scope their exercise. Furthermore, the template may be reused by exercise planners who are using this same National Planning Scenario.

Exercise controllers require a common and consistent picture of the exercise.

Maintaining situational awareness of exercise events and a common exercise picture across the dozens of exercise controllers in the Master Control Cell and SimCell is vital for sustaining a realistic scenario. During NLE10, there were several instances in which a common exercise picture and improved situational awareness would have been valuable. For example, at one point the Master Control Cell released a scenario inject that provided incorrect ground truth data.

SUMMIT provided some enhanced situational awareness in NLE10 by integrating several M&S tools into a single conceptual simulation template. A more comprehensive common exercise picture can be provided through exercise management tools that link with SUMMIT. Exercise management tools provide timelines of exercise injects, expected player actions, and actual player actions. They show how player actions affect the ground truth scenario. The exercise management tools used in NLE10 provided a global view; however, according to feedback received from the exercise controllers, these tools were neither easy nor intuitive to navigate and query. Additionally, controllers reported that there was little to no feedback on player receipt of injects and player responses to injects. The SUMMIT architecture is being designed so exercise management tools can be integrated with M&S tools that generate ground truth. This will help ensure a common exercise picture, enabling exercise planners and managers to record and access exercise objectives, scenario ground truth data, expected and actual player actions, exercise management team actions, consequences of player actions, and scenario outcomes in one location.

Collaboration environments and visualization tools greatly enhance exercise planning and conduct, but must be user-friendly, intuitive and robust.

Visualization tools, such as GIS-based and virtual world technologies, can be used to display a common exercise picture of the exercise scenario data, release of injects, player actions and consequences of player actions. For exercise planners and controllers who are not techno-savvy (which several people stated about themselves during the hot wash feedback session immediately following NLE10), a virtual world or other visualization should make it much easier for them to interact and make changes to the scenario and common exercise picture. The SUMMIT architecture is providing a means for virtual world technologies to be federated with M&S and exercise management tools. Exercise data can be displayed in an immersive environment and accessed by distributed exercise planners, managers and players.

Scenario data coordination and consistency is imperative.

During the exercise planning phase, M&S was used to develop ground truth data and exercise injects. One of the benefits of using M&S for this purpose is to help ensure that the underlying scenario is consistent and realistic. It is much less likely for exercise planners or controllers to create conflicting scenario data when the ground truth data are calculated or derived from a physics-based model, objective data and a consistent set of assumptions. In NLE10 some inconsistencies in the ground truth data did appear because of the use of several models with different assumptions. For example, two of the data providers calculated casualty numbers which differed significantly due to the fact that different population databases were used. Through the use of a unifying M&S framework, discrepancies between models can be managed by following these guidelines:

1) Models with the same inputs and outputs should be managed in a single simulation template, making it easy to set up comparative runs. The same

template should be used to compute all ground truth data.

2) All assumptions and input parameters are documented, openly shared and used among the data providers.
3) All M&S tools (including databases) that are used to calculate scenario data are integrated.

To minimize inconsistencies and enhance coordination in scenario development, SUMMIT is providing an integrating framework through which multiple models and datasets can be used together to generate consistent data.

4.0 CONCLUSIONS

NLE10 provided important lessons learned on architecture requirements and ConOps for SUMMIT. These are being implemented in SUMMIT and will help enhance exercise planning and conduct in NLE11. Research on SUMMIT support to Tier II-IV exercises is underway and will build upon the lessons learned from support to NLEs.

The SUMMIT architecture has proven to be flexible enough to create simulations via model federation that allow for complex scenario construction in an intuitive manner. The flexibility and extensibility of the architecture also allows for evolutionary growth with participation of the M&S and exercise communities. Integration of data visualization tools and virtual environments allows M&S data and results to be readily accessed by the exercise community.

A SUMMIT *early adopter program* has been established to evaluate the integration process with the participation of model contributors in the M&S community. Information about SUMMIT and this program can be found at the SUMMIT web site (**http://dhs-summit.com**).

The current focus for SUMMIT is support for emergency response exercises; subsequent research will focus on emergency planning and response operations. Bringing modeling and simulation tools to emergency planning

and operations will allow for improved accuracy in exercise parameters, creating more realistic training exercises and better prepared emergency responders.

5.0 REFERENCES

[1] Homeland Security Exercise and Evaluation Program (HSEEP), http://hseep.dhs.gov/.

[2] Department of Homeland Security, "National Exercise Program," presentation at the DOD Worldwide Joint Training and Scheduling Conference 2007-1 (WJTSC 07-1), Mar 8, 2007.

[3] Homeland Security Council, *National Exercise Program Implementation Plan*, June 20, 2008.

[4] F. Kuhl, R. Weatherly, and J. Dahman (1999). *Creating Computer Simulation Systems: An Introduction to the High Level Architecture*, Upper Saddle River: Prentice-Hall.

[5] E. Friedman-Hill, T. Plantenga, and H. Ammerlahn, (2010). "Simulation Templates in the SUMMIT System", in *2010 SISO Spring Interoperability Workshop*, Orlando, FL, Apr 2010.

[6] T. Plantenga, E. Friedman-Hill, (2010). "Integrated Modeling, Mapping, and Simulation (IMMS) Framework for Planning Exercises", submitted to *Interservice/Industry Training, Simulation, and Education Conferences (I/ITSEC)*, 2010.

[7] FEMA National Exercise Division, "National Level Exercise 2010: Scenario Ground Truth", May 2010.

[8] Homeland Security Council, "NATIONAL PLANNING SCENARIOS: Executive Summaries", April 2005.

[9] U.S. Department of Homeland Security/FEMA, Lessons Learned Information Sharing Portal, http://www.llis.dhs.gov.

6.0 ACKNOWLEDGMENTS

4.6   Virtual Worlds and Homeland Security

October 13–15, 2010
Hampton, Virginia

Virtual Worlds and Homeland Security
Michael Macedonia, Ph.D.
SAIC

October 15, 2010

MODSIM WORLD
Conference & Expo

Agenda

- Overview
- Requirements
- Applications

Avatars

- Avatars mimic natural human movement
  - Controlled via simple keyboard or controller input
  - Avatars use realistic animations and advanced blending techniques
  - Emotion and expression framework combines user input and scripted behavior to mimic culturally specific movement patterns
  - Integrated physiology model
  - FaceGen integration provides photo-specific avatars

3

Physics

- Programmable physics engine adapts to network latency
  - The physics engine can be programmed to simulate real-world dynamics
  - Simulation is accurate, validated on the server

4

Networking

- Supports distributed operations
  - Users login to the virtual environment from remote locations across the globe and participate just as they would if they were co-located
- Networking engine minimizes bandwidth requirements
  - Efficient communication protocol minimizes necessary bandwidth, allowing simulation to run over LANs, WANs, and the Internet (such as long-haul networks)

LANs = large area networks WANs = wide area networks

OLIVE is a trademark of Science Applications International Corporation in the U.S. and/or other countries.

5

In-World Communication

- Multiple forms of communication between users
  - Spatially accurate voice-over-IP (VoIP)
  - Highly integrated voice communications with lip sync, automated gesticulations and speaker attention
  - Instant messaging (broadcast or person to person)
  - Built-in radio communication
  - Manual hand signals and gestures
  - Culturally specific library integration
  - Telephony for external access


6

Collaboration Features

- Supports in-world presentation screens that support a variety of rich media
  - PowerPoint®
  - Streaming video
  - Live streaming video
  - Application sharing
- Multiple screens can be placed throughout the world
- Prompter, zoom support
- Laser pointers
- Presence indicators

7

Geospecific Terrain

- Supports large-area, geospecific terrain
  - CDT SE CORE databases
  - WGS-84 Datum
 - OpenFlight interoperability
  - Double precision processing
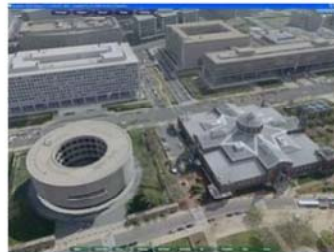
CDT SE CORE = Common Driver Trainer Synthetic Environment Core

8

Session Replay

- Built-in distributed replay
  - Collects all voice, keyboard/mouse and controller inputs across the system
  - Plays results back through system, allowing free-cam
  - Full data mart for external analysis
  - VCR playback features
  - Distributed camera control
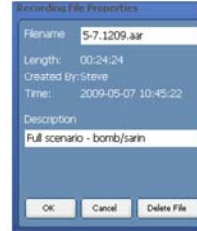
9

Non-Player Characters

- General, open API for integrating external artificial intelligence
  - Ability for external application to instantiate and control entities
  - API provides information on in-world activity to external application
  - Support for low-level-of-detail avatars for crowd scenes
 - Can also be used to support real-time telemetry

API = application programming interface

10

545

Special Effects

- Supports a variety of special effects to add realism to the scene
  - Particle-based effects for natural phenomenon
  - Hold tools to build items with which avatars can interact
  - Time of day and weather support
  - Full suite of weapons, including small arms and rocket-propelled grenades (RPGs)
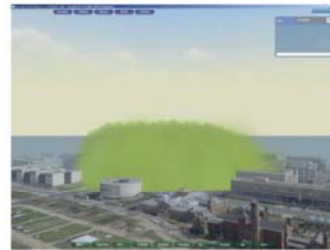
11

Enterprise IT

- Working to support deployment challenges
  - Full support for behind the firewall operation
  - Port multiplexing to support single port communication through firewalls
  - Lightweight Directory Access Protocol (LDAP) integration
  - Integration with eAuthentication to support Level 2 authentication
  - Secure Socket Layer (SSL) encryption available between server and client

12

Group Meetings

- Branded rooms and accessories
- Identity
  - Personalized avatars
  - Profiles
- Media sharing
 - PowerPoint®
  - Video
  - Desktop applications

PowerPoint is a registered trademark of Microsoft Corporation in the U.S. and/or other countries. Lotus and Sametime are registered trademarks of International Business Machines Corporation in the U.S. and/or other countries.

13

Events

- 3D models
- Event roles
  - Organizer
  - Moderator
  - Presenter
- Optimized attendee experience

Team Project Management

MODSIM WORLD
Conference & Expo
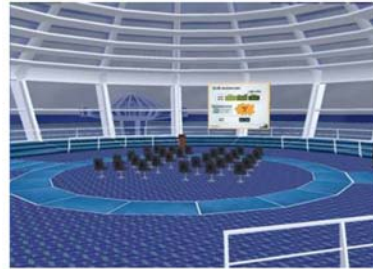
- Persistent room
- Screen placement optimized for team use
- Team documents

Training

MODSIM WORLD
Conference & Expo

- Realistic and hypothetical scenarios
  - Scenes
  - Simulations
  - Role players
  - Non-player characters (NPCs)
- Scenario and Scene Editor
- CBT or SCORM® integration
  - Instructor-led
  - Self-paced
- Record and replay
- Data mart

CBT = computer-based training
SCORM is a registered trademark of the Department of Defense in the U.S. and/or other countries.

Operational Solutions

- Virtual emergency operation centers
- Common operating picture
- Context-specific operation centers
- Connection to real world – GPS, RFID and other sensors
- Embedded rehearsal environments

GPS = Geographic Information System

RFID = radio frequency identification

5.0 THE HUMAN DIMENSION TRACK

5.1 The Army's Human Dimension

HD Operational Problem Statement
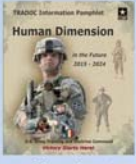
Research & Concept Development Communities

Study

Joint & Army Senior Leaders Congressional Staffers

Concept

General Audience

Handout

Given the requirement for Full Spectrum Operations in an Era of Persistent Conflict with the demands of the Army Force Generation cycle (Reset – Train/Ready – Available) and that Soldiers are the centerpiece of our formations, human capabilities are the key to winning our current and future wars …

The Operational Problem: The Army must focus our Human Dimension efforts to ensure: *sustained quality* of the All-Volunteer Force, trained Soldiers, Civilians, Leaders and units prepared for Full Spectrum Operations (FSO), a resilient force Reset and Trained/Ready for *deployment and prepared for complex and demanding Joint, Interagency, Intergovernmental, and Multi-national (JIIM) environments* now and in the future.

Requires an adaptive institution

3

http://www.tradoc.army.mil/tpubs/pamndx.htm

The Human Dimension Concept

Previous Army concepts acknowledge the Soldier as the centerpiece of our formations, but none, individually or collectively, adequately addresses the human dimension capabilities for current or future operations.

Using available and emerging tools that …

Cognitive
– optimize cognitive flexibility, mental intellect and information processing through enhanced screening, recurring assessment and tracking of individual's potential and attributes; *dynamic, scalable, adaptive, immersive, sensory enabled, multi-layered tailored training*; Adaptive material systems maximizing individual attributes.

Physical
– develop lifelong total fitness habits through comprehensive wellness programs that build aerobic/mental capacity, strength, endurance, agility, focused nutrition, stress and sleep deprivation management, behavioral health. Build resilience thru the Physical, Mental, Family, Social and Spiritual domains of strength.

Social
-- strengthen character and intercultural adaptability that reflects confidence in tough moral, culturally sensitive situations grounded in law, Warrior Ethos and Army Values; develop improved understanding of social / family dynamics, respect, interpersonal relationships, spirit and faith; strengthen team building, foster cohesion

The future domestic / global operating environment requires agile policies to support a comprehensive Human Dimension approach.

4

552

A recruited Future Force managed and retained based, in part, on continuous cognitive assessments (e.g. Attention, Learning, Leadership potential, Adaptability, Decision Making, Vision). Advanced technologies and tools assist in the selection of individuals for assignments and advanced accelerated / measurable training. Enhanced tailorable multi-layered training (to facilitate and accelerate task learning), leader development and Mission Command systems that adapt to cognitive styles to maximize readiness. Leaders provided PDA-like devices that access training and decision-making tools to track soldier readiness and assist matching talents/skills to mission requirements.

5
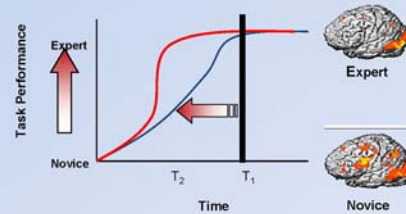
Facts:

- 100 bil neurons, .6 cft, uses 2w/hr vs. Super Computer @ 1600 sft, 5000w just for cooling
- Everyone "wired" differently – More synaptic connections than all known bodies in the universe i.e., billions
- 2% body mass, yet consumes 20% of the energy - alert or asleep
- 3 seasons of the brain – Maturing, Adult, Aging
- Spatial Navigation differences – Men and Women *ARE* different
- Cognitive Peaks differ individually for D-making

Possibilities:

- Predict leadership potential and decision-making capabilities
- Identify cognitive styles, special skills and attributes
- Cognitive gym -- Cognitive PT Test that builds capability & experience
- Cognitive UCOFT – exercise full spectrum skills
- Accelerate learning -- tailored to individual potential and preferences
- Train the untrainable?
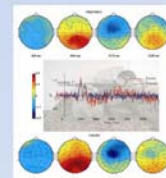
UCOFT - Unit Conduct Of Fire Trainer

Maximize a Soldier's inherent cognitive potential and learning.

6

553

A Future Force that adheres to a continuum of holistic fitness tailored to the individual and subsequent mission requirements (measurable physiological, neurological, psychological, nutritional, and developmental fitness training). Programs that identify, mitigate, treat and rapidly restore soldiers who become holistically "unfit" due to combat operational and stress-related injuries. Retention of qualified physically disabled soldiers is the norm.
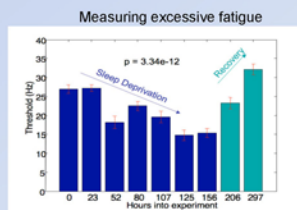
Physical Component

Office of Comprehensive Soldier Fitness, HQDA G-3/5/7

Global Assessment Tool (GAT)

Master Resiliency Program
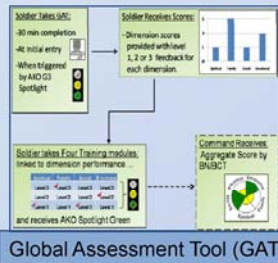
Finding strength to work thru adversity...

Finding humor in something that has gone wrong ...

Possibilities:
- Effective resilience building and stress mitigation
- Identify combat stress / PTSD vulnerability –
- Soldiers physically, mentally, emotionally, socially fit

Maximize a Soldier's inherent physical potential and holistic health.

9

Social Component Outcome

A Future Force that functions and behaves in accordance with: law; Army Values; and national/international expectations and standards. Grounded by a continuum of adaptable, scalable and measurable training programs that include operational challenges in tough ethical/moral situations. Leaders achieving intercultural adaptability, language skills and respect of the potential strategic impact how ethical behavior affects ones' self, the Army and Family values.

WELCOME HOME DADDY!

"... Soldiers and leaders must feel confident ... to interact day-to-day with people of different cultural backgrounds and perspectives... GEN Casey

10

555

ARMY FORCE MANAGEMENT

Army Deliberate Processes

Oh No!
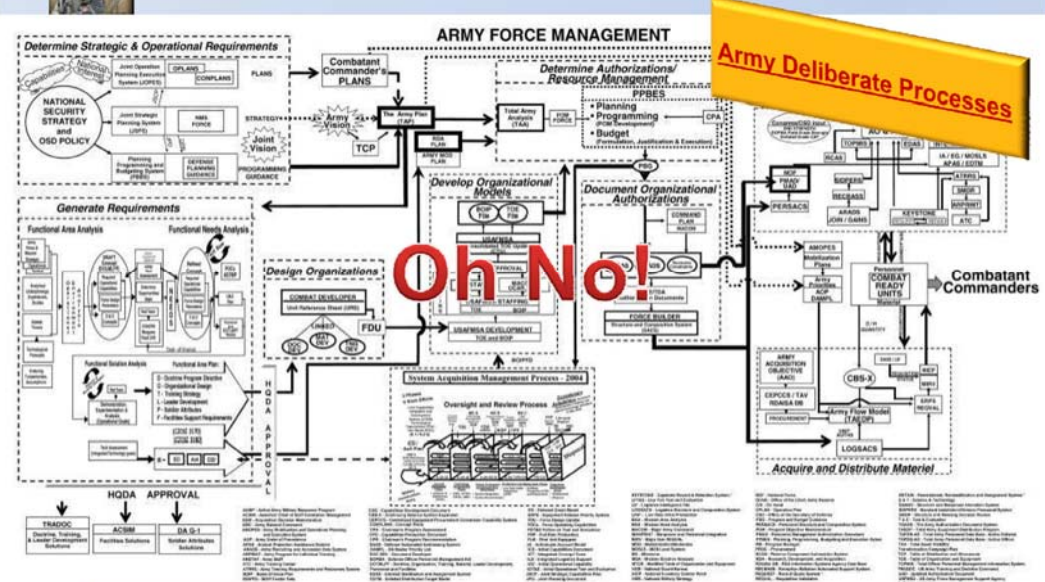
Not responsive to current fight or synchronized with ARFORGEN

13

Traditional Tools and Methods

EMERGING Tools and Methods

-- Psychological assessments

-- Physiological assessments
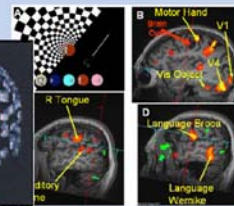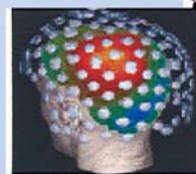
EEG & fMRI Functional Mapping
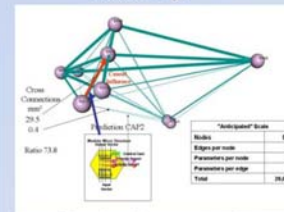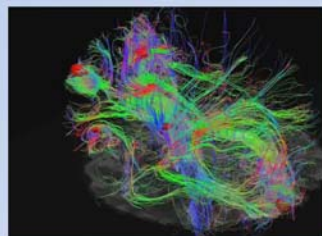
Brain Connective Topology

HD impact on Training

- *Live, virtual, constructive and mixed venues*

- *Enable the Future Force to impart more skills, faster, at lower cost and with greater retention than currently achievable*

- *Use non-traditional home station training techniques and technology, train prior to employment*

- *Enhance and account for individual proficiencies and learning rates (outcome based)*

- *Leader development must be completely adaptable, scalable, multi-layered in complexity*

Guidelines we need to follow

1. The measurement and assessment of human performance is a centerpiece, *not an ancillary benefit*.

2. LVC training must contain immersive, decision-making stimuli with increasing variables that do not replicate what was done before. *Nor permit unintended bad habits.*

3. Tactical, moral, ethical decision-making must be stressed *and pervasive in all training*. Such contexts lead to individual and small unit self-confidence. It should also have open-ended objectives and challenge the task-condition standard construct.

4. HD brings a NEW way of accessing, selecting, training, developing and transitioning Soldiers requiring evidence-based, measurable results that can be correlated to return on investment.

16

We are about

- ✓ Empowering Soldiers / units to dominate the Land Domain
- ✓ Improving, optimizing and restoring cognitive, physical, and social abilities
- ✓ Enabling Soldiers to function efficiently as an integral component of a network and society

17

Back up

18

Taking the Capability Approach

Outcome: An Army poised to achieve its maximum cognitive, physical, and social potential for Soldiers, Civilians, and Families by optimizing their abilities, experience, education and fitness.

Integrated Capability Development Team (ICDT)

DCR – DOTMLPF-P Change Request; DICR - DOTMLPF-P Integrated Change Request; ICD – Initial Capabilities Document; CNA – Capability Needs Analysis

19

HD Quad Chart

HD ICD Outcomes nested with The Army Capstone Concept:

Operational Adaptability — CPS
- Quality Army Leaders and forces ☑
- Applied critical thinking ☑
- Comfort w/ ambiguity & decentralization ☑
- Accepting prudent risks ☑
- Continuous assessments & rapid adjustments ☑
- Educated Leaders culturally astute ☑
- Proficient & Cohesive Teams ☑
- Resilient Soldiers successfully enduring psychological ☑
  & moral challenges

Context:

HD reaches across Army forces, warfighting functions and various Force Modernization proponents.

No organization or process to develop holistic resourced-informed; outcomes-based; integration-focused HD CPS capabilities.

AR 5-22, The Army Force Modernization Proponent System does not designate a Force Modernization Proponent for HD.

Many disparate organizations are developing HD-related capabilities in parallel and stove-piped manner challenging the effort to collect gains made

No HD Center of Excellence (CoE) or Capabilities Development and Integration Directorate (CDID).

Assertions:

Transformational endeavor requiring proponency to integrate CPS attributes across the Army

Continued investment in R&D of CPS measures is required to determine and predict Soldier potential, performance and resiliency

Policy changes are required for acquisition, selection, development, retention, career management and transition

Recommendation:

Establish HD as a Program of Record

Establish a HD management structure having Force Modernization proponency for HD resourced with the TCM and CDID

Charter a Senior Advisory Group to facilitate the effective and efficient enabling of research, development and experimentation

Redefine Soldier readiness in CPS terms

Add squad/small unit readiness as a pacing item

Utilize the Human Capital Enterprise (HCE) to support HD equities

Soldier Lifecycle Model

5.2   Human Performance Modeling Tools for Better System Design

Human Performance Modeling Tools for Better System Design

October 13–15, 2010
Hampton, Virginia

MODSIM WORLD
Conference & Expo

Charneta Samms
U.S. Army Research Laboratory

US ARMY RDECOM   ARL

October 13-15, 2010

Agenda

MODSIM WORLD
Conference & Expo

- Why Human Performance Modeling (HPM)?
- Importance of HPM to System Design
- HPM Tools and Applications
  - CogTool
  - C3TRACE
  - IMPRINT
  - MIDS Plug-in
- Expansion of tools
- Summary

Why Human Performance Modeling (HPM)?

MODSIM WORLD
Conference & Expo

Concept System

Many Variables

Field Study Not Feasible

Too Dangerous

Model – Test – Model

System Performance $\cong$ f(human performance)

3

Importance of HPM to System Design
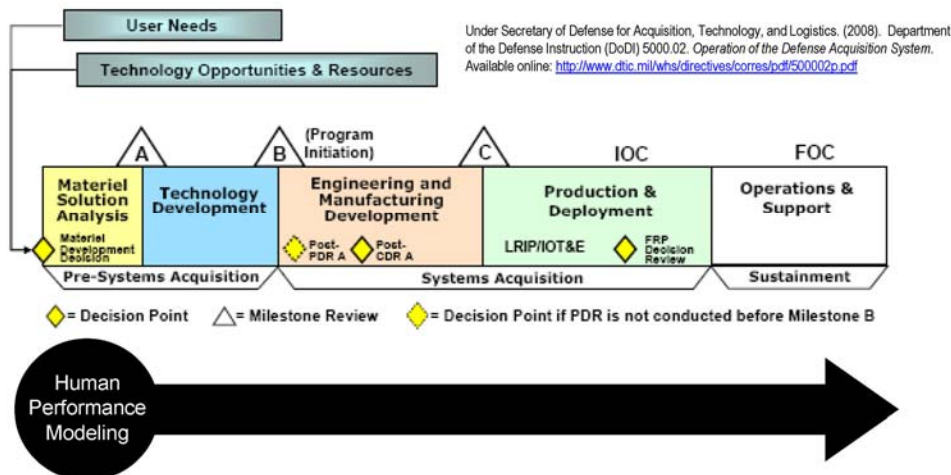
MODSIM WORLD
Conference & Expo

Under Secretary of Defense for Acquisition, Technology, and Logistics. (2008). Department of the Defense Instruction (DoDI) 5000.02. *Operation of the Defense Acquisition System.* Available online: http://www.dtic.mil/whs/directives/corres/pdf/500002p.pdf

Provide quantitative data to inform trade off decisions early in design process

CogTool

- General purpose UI prototyping tool
- Automatically evaluates design with predictive human performance model
- "cognitive crash dummy"

1. Set up a project to compare design alternatives on a suite of tasks

2. Lay out a storyboard of frames (what the user will see) and transitions between them (what the user will do)

3. Detail each frame with the interactive widgets available to the user

4. Demonstrate the tasks, CogTool creates a valid cognitive model of a skilled user.

5. Press "Compute" and CogTool creates ACT-R code, runs it and produces a prediction of skilled execution time

6. Examine what the ACT-R model did to produce the prediction in an interactive timeline visualization.

Bonnie E. John, Principle Investigator
Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
http://cogtool.org

John, B. E., (2010) CogTool: Predictive Human Performance Modeling by Demonstration. Proceedings of the 19th Annual Conference on Behavior Representation in Modeling and Simulation (BRIMS) (Charleston, SC, March 21-24, 2010).

CogTool Application

- Compared time to complete programming tasks within two environments
  - 2002 – Unix command line and Vim editor
  - 2010 - Eclipse Parallel Tools Platform (PTP)

The help system opens

Operations to find the 1st term

Operations to find the 2nd term

Results

Eclipse PTP interface will improve performance of skilled programmers over 2002 command line interface

* Project: PERCS_CLI_PTP_Comparison_20091216_1513

Tasks	Command Line	PTP-Eclipse
▼ HelloWorld_mpi	Min: 114.405 s	Min: 40.090 s
HelloWorld_mpi with keyboard	161.111 s	
HelloWorld_mpi with mouse	114.405 s	40.090 s
▼ F1Help	Min: 30.726 s	Min: 10.462 s
F1 Help with keyboard	30.726 s	
F1 Help with mouse	31.247 s	10.462 s
▼ Code Folding	Min: 5.780 s	Min: 3.563 s
Code folding with keyboard	10.490 s	
Code folding with mouse	5.780 s	3.563 s
Barrier Analysis	40.149 s	11.833 s

Richards, J., Bellamy, R., John, B., Swart, C. & Thomas, J. (2010) Using CogTool to Model Programming Tasks. Psychology of Programming Interest Group WIP (PPIGWIP) (Dundee 2010). www.ppig.org.

C3TRACE

Command, Control, and Communications - Techniques for the Reliable Assessment of Concept Execution

Goal: *To conduct "what-if" analyses based on information flow and quality, to discover alternative organizational, personnel, and system configurations that increase performance*

- Evaluate effects of different personnel architectures and information technology on system and human performance

- Investigate efficiency and effectiveness of message processing in Command & Control environments
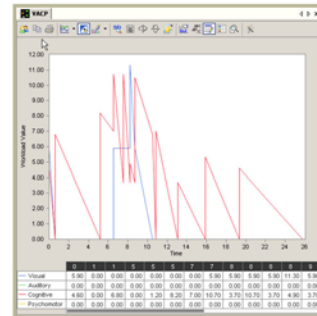
C3TRACE Application

Future Command and Control Cell Analysis

System Issue

Requirements and Force Design are in conflict

physics

Desired vehicle design → Restricted vehicle design

1. Modify deployment concept
   Require different aircraft

2. Accept degradation in 16 soldier C2 cell
   Reduced capability

3. Accept more C2Vs
   More money, maintenance, lifts
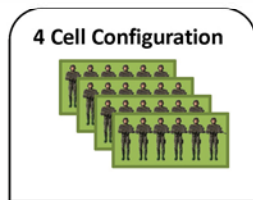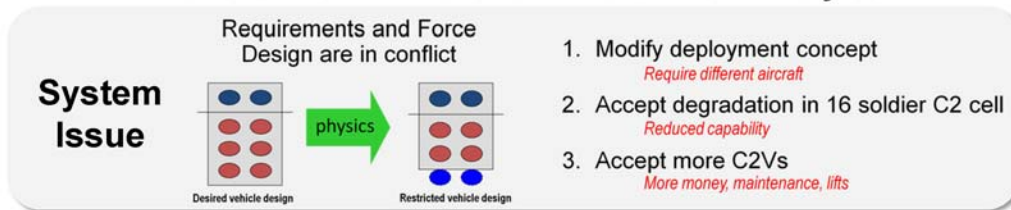
Results

Performance Measures
Utilization
Probability of "good decision"
Messages handling

4 Cell Configuration

6 Cell Configuration
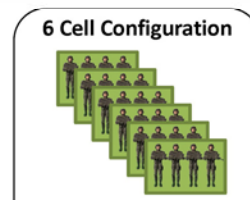
- 19 of 24 - 100% utilization
- 6 of 24 – 25+% poor decision quality
- 18 of 24 - dropped 50+% of messages

- 13 of 24 - 100% utilization
- 5 of 24 – 25+% poor decision quality
- 8 of 24 – dropped 50+% of messages

Mitchell, D. K., Samms, C., Kozycki, R., Kilduff, P., Swoboda, J., & Animashaun, A. (2006) *Soldier Mental Workload, Space Claims, and Information Flow Analysis of the Combined Arms Battalion Headquarters Command and Control (C2) Cells* (ARL-TR-3861). Army Research Laboratory, APG, MD.

Improved Performance Research Integration Tool

MODSIM WORLD
Conference & Expo

IMPRINT
Improved Performance Research Integration Tool

334 users supporting Army, Navy, Air Force, Marines, NASA, Department of Homeland Security (DHS), Department of Transportation (DoT), Joint and other organizations across the country

http://www.arl.army.mil/IMPRINT

https://km3.alionscience.com/sites/imprint

IMPRINT can be used to

- Set realistic system requirements

- Identify future manpower & personnel constraints

- Evaluate operator & crew workload

- Test alternate system-crew function allocations

- Assess required maintenance man-hours

- Assess performance during extreme conditions

- Examine performance as a function of personnel characteristics and training frequency & recency

- Identify areas to focus test and evaluation resources

- Quantify human system integration risks in mission performance terms to support milestone review

- Represent humans in federated simulations

IMPRINT is a trade-off analysis tool

IMPRINT Application

MODSIM WORLD Conference & Expo

System Issue

- Rely on heavy armor and artillery to protect the forces
- Heavy, large systems are difficult to deploy rapidly

Current Tank
Commander, Driver, Gunner, Loader

Survivability Onion:
- Don't be There
- Don't be Seen
- Don't be Acquired
- Don't be Hit
- Don't be Penetrated
- Don't be Killed

Future Tank
Commander, Driver

- Rely on situation awareness to protect the forces
- Lighter, smaller systems are easier to deploy rapidly

Methodology

- Identified functions and tasks via knowledge elicitation
- Set up experimental conditions to model based on varying function allocations
- Built models
- Validated models by walking-through with Soldiers
- Completed runs and prepared results

Results

Commander - Driver and Gunner
Highest workload of all conditions

↓

Gunner - Driver and Commander
No shooting on the move

↓

Commander - Gunner and Driver
Best two crewmember function allocation; single vehicle commander

↓

Commander, Driver and Gunner
Two crewmembers scanning; allows hunter-killer philosophy

Mitchell, D. K., Samms, C. L., Henthorn, T., & Wojciechowski, J. Q. (2003). Trade study: A Two-versus three-Soldier crew for the Mounted Combat System (MCS) and other future combat system platforms (Technical report ARL-TR-3026). Aberdeen Proving Ground, MD, U.S. Army Research Laboratory.

A Decade of Impact on Soldier-System Integration

MODSIM WORLD Conference & Expo

Combined Arms Testbed
First identification of workload issues associated with a 2 Soldier common crew

Command and Control Cell
Supported requirement for 24 personnel allocated to the battalion in the Unit Reference Sheet

Lightweight Howitzer
Supported the possibility of reducing crew size

Future Howitzer
Workload issues associated with rearming resulted in an automated rearming concept to be included in system design

Autonomous Navigation System
Provided support for the ANS technology to increase crew performance

1999 — 2001 — 2003 — 2005 — 2007 — 2009

Situational Understanding STO
Identified critical information requirements for system and display development

Future Tank Platoon Leader Variant
High workload analysis predictions matched experimental results

All Future Concept Vehicle analyses
Soldier workload identified as #1 issue during preliminary design review

Future Tank
Identified workload issues which resulted in system design change from 2 to 3 Soldier crew

Future Reconnaissance and Surveillance Vehicle
Served as basis of manning assessment and justified need for all operators to have displays

567

Multimodal Information Design Support (MIDS) Tool Plug-in

- Develop potential mitigation strategies from multimodal design guidelines matched to areas of high workload as identified in IMPRINT
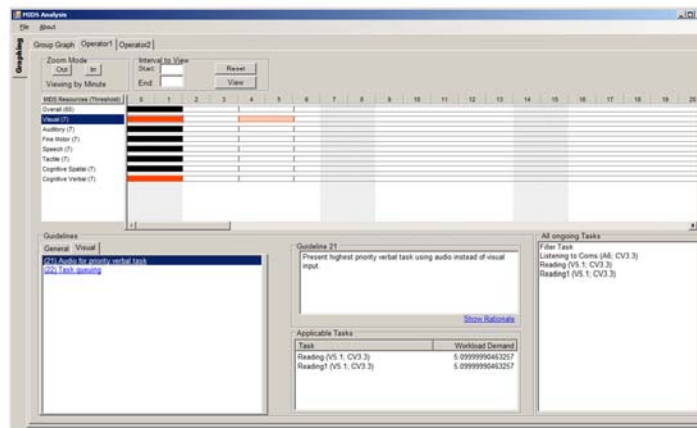
Expansion of Tools

- Develop smart "links" between tools
- Keep up with evolving analysis demands

- Specific Enhancements
  - CogTool
    - Additional measures
  - C3TRACE
    - Visualization of impacts to decision quality
  - IMPRINT
    - Connect to system engineering
  - MIDS Plug-in
    - Predict effect of incorporating mitigation strategies

Summary

- Use of human performance modeling tools can
  - Provide quantitative data to inform trade off decisions early in design process
    - Cost savings
    - Better design
  - Focus test and evaluation resources
    - Model – Test – Model approach
- Expand tools to answer new analytic

5.3 ACT-UP: A Toolkit for Hampton, Cognitive Modeling Composition, Reuse and Integration
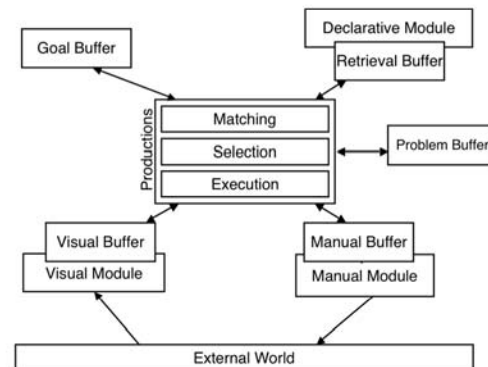
Motivations and Applications

- **Philosophy**: Unified understanding of the mind.
- **Psychology**: Account for experimental data.
- **Education**: Provide cognitive models for intelligent tutoring systems and other learning environments.
- **Human Computer Interaction**: Evaluate artifacts and help in their design.
- **Computer Generated Forces**: Provide cognitive agents to inhabit training environments & games.
- **Neuroscience**: Provide a framework for interpreting data from brain imaging.

Goals

- Enable the implementation of more complex ACT-R models
- Scale up cognitive models to simulate learning / adaptation in communities (e.g., about 1,000 models in parallel)
- Treat models as hard claims
  - Evaluate each specified component against data
  - Underspecify the rest and fit free parameters

The Argument

- **Constraints:** Architectural advances require further constraints
- **Scaling it up:** Complex tasks, broad coverage of behavior (e.g., linguistic), use of microstrategies and predictive modeling may serve to motivate further architectural constraints
- **Difficulties:** ACT-R is heavily constrained already, and models are difficult to develop, reuse and exchange

Control Structure

Flow Chart
(Finite State
Automaton)
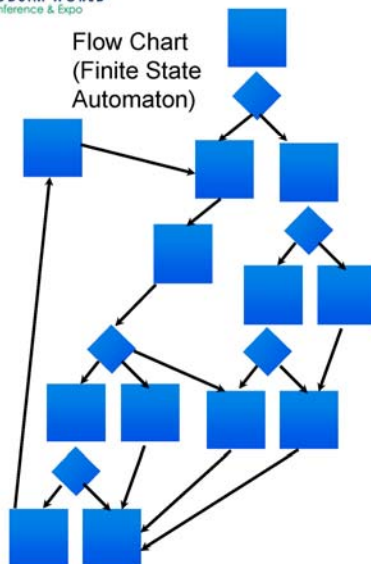
A flow-chart describes an algorithm (or a cognitive strategy)

Decision-making points and states

Not easy to reuse: it fails to capture generalizations

Computer Science: pre-Object Orientation, pre-Functional Programming

Decomposition
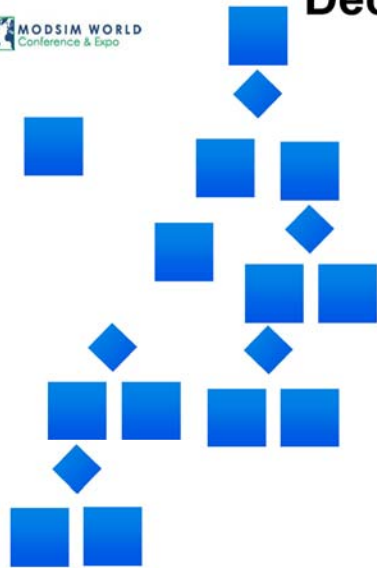
Production Rule System
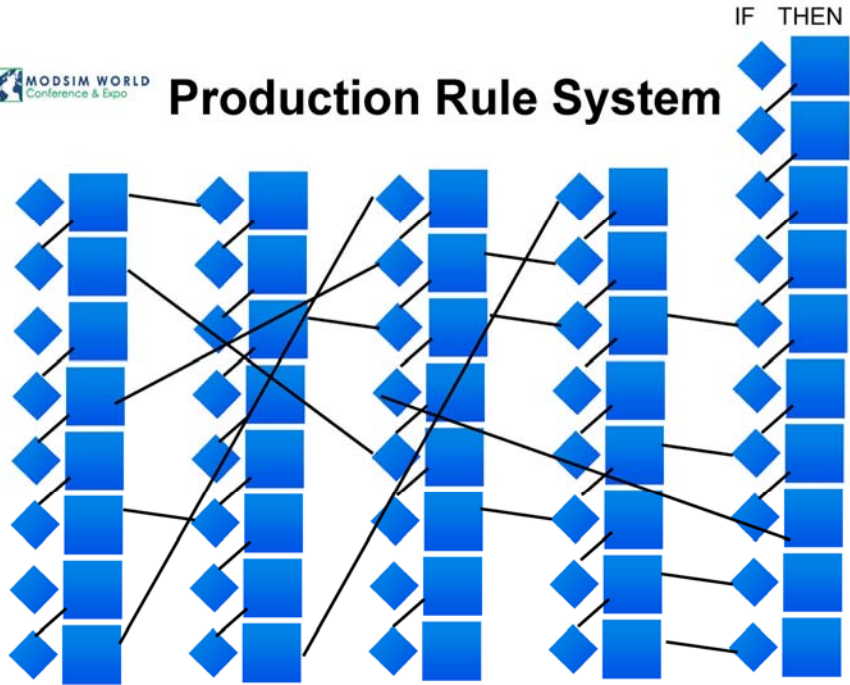
The Argument

- **Constraints:** Architectural advances require further constraints
- **Scaling it up:** Complex tasks, broad coverage of behavior (e.g., linguistic), use of microstrategies and predictive modeling may serve to motivate further architectural constraints
- **Difficulties:** ACT-R is heavily constrained already, and models are difficult to develop, reuse and exchange
- **We need to produce models at a higher abstraction level**
  - However, we'd like to leverage successful cognitive modules, describing memory retention, cue-based retrieval, routinization, reinforcement learning

Cognitive Strategy
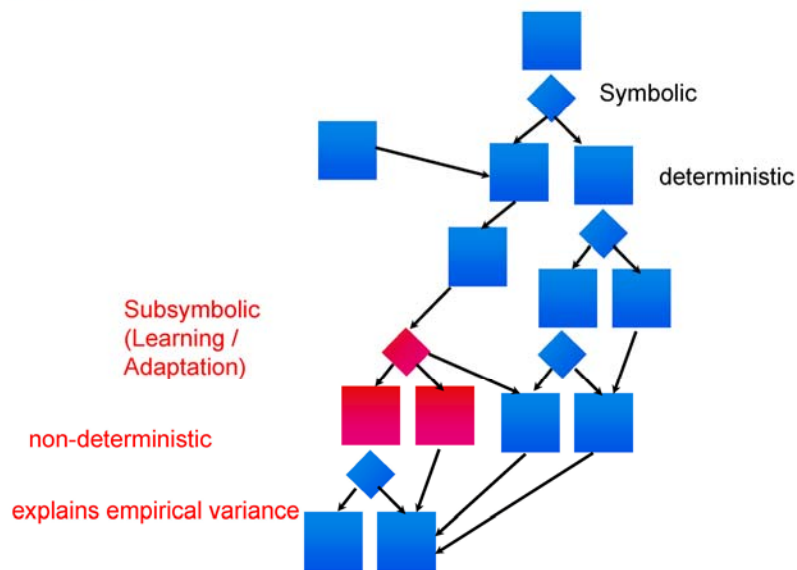
Priming Model

Crucial request of a chunk from declarative memory

- Only a small portion of the model explains the behavioral data at hand
- The rest explains that the task can be accomplished in principle with a parallel architecture and with specific cognitive representations (chunk types)

Production Systems vs. assembly language

```
evensum:     cir.1    D1          ;Zero-out
                                  ;Accumulator
sumloop:     add.1    D0,D1       ;Add current
                                  ;counter value to
accumulator
             subq     #1,D0       ;Decrement
                                  ;counter by one
             bne      sumloop     ;until it
                                  ;reaches zero
             muls     #2,D1       ;Double sum to account
                                  ;for even numbers
             rts                  ;Return
                                  ;to caller
```

~1990

575

The Argument

- **Constraints:** Architectural advances require further constraints
- **Scaling it up:** Complex tasks, broad coverage of behavior (e.g., linguistic), use of microstrategies and predictive modeling may serve to motivate further architectural constraints
- **Difficulties:** ACT-R is heavily constrained already, and models are difficult to develop, reuse and exchange
- **Abstraction:** To implement those, we need to produce models at a higher abstraction level
- **Underspecification is the key to focus on verifiable claims, and to avoid overfitting by fitting free parameters to data**

Underspecified Models

underspecify:

deterministic

specify:

non-deterministic

explains empirical variance

577
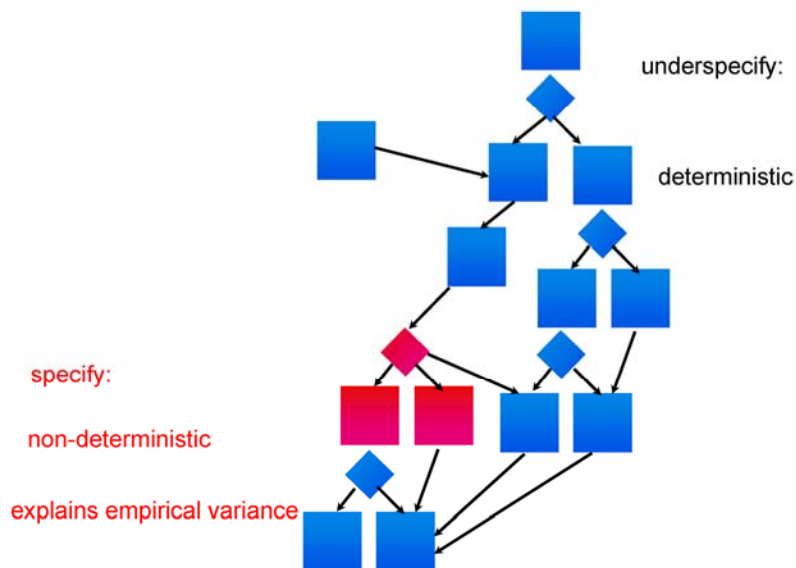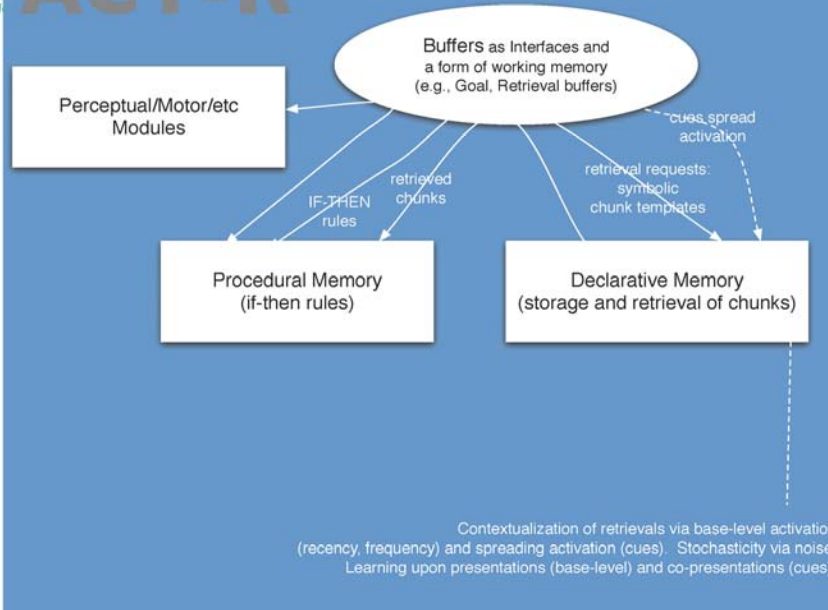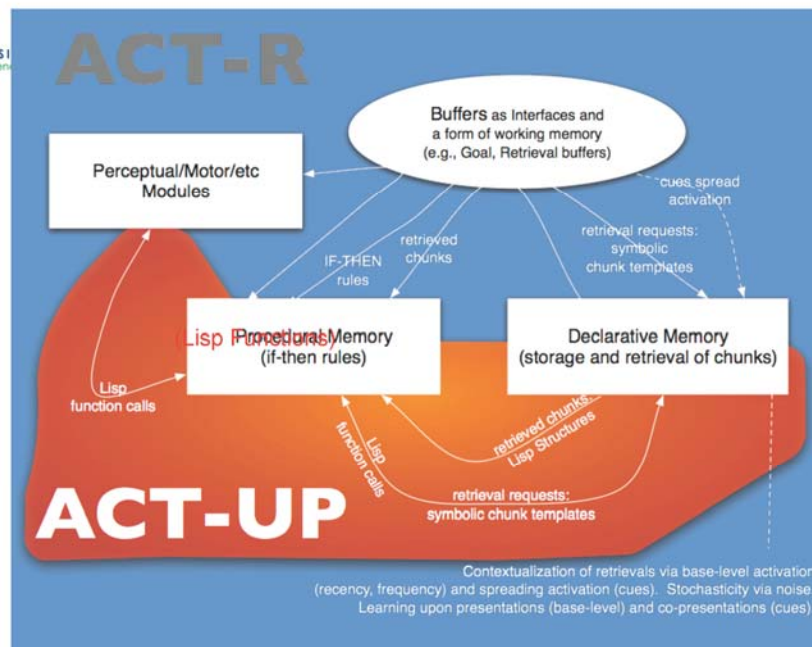
ACT-UP

- A stand-alone system on the basis of Common Lisp
- targets an audience that can write simple Lisp programs (unlike, e.g., CogTool)
- Toolbox approach to ACT-R
  - light-weight: it's a Lisp library
  - does not produce production rules (ACT-R/Lisa, ACT-Simple, CogTool)
- Not aimed at implementing all constraints of ACT-R 6 (unlike Java ACT-R, Python ACT-R)

Declarative Memory

- `define-chunk-type'
  - types are optional
- `make-count-order'
- `learn-chunk'
- `defrule'
- `retrieve-chunk'
- `count-order-second'

```
;; ACT-R parameters                    ACT-UP Code
(setq *lf* .05)
(setq *rt* -1)

;;;; Defining chunk type

(define-chunk-type count-order first second)
```

578

ACT-UP is not ACT-R 6...

- ACT-UP Interface is synchronous
  - Serial execution
  - Deterministic strategies defined as programs
- Parallelism (e.g., perceptual/motor modules) possible [not implemented]
- Non-deterministic rule choice is possible
  - Reinforcement-learning as in ACT-R 6

PM / Utility learning

- `choose-coin'
- calls either `decide-heads or `decide-tails'
- `assign-reward' reinforces the decision
- Exact production rules are underspecified,
  - but decision-making point is explicit
- Choice model replicates ACT-R and empirical results

```
;; Experimental environment
(defun toss-coin ()
  (if (< (random 1.0) .9) 'heads 'tails))

;; The Model
;;;; Rules that return the choice as symbol heads or tails

(defrule decide-tails ()
  :group choose-coin
  'tails)

(defrule decide-heads ()
  :group choose-coin
  'heads)
```

Debugging

```
(defrule count-model (arg1 arg2)
  "Count from ARG1 to ARG1.
ARG1 is the starting point and ARG2 is the ending point.
Each increment is 1 unit."
  (speak arg1)
  (if (not (eq arg1 arg2))
      (let ((p (debug-detail (retrieve-chunk (list :chunk-type 'count-order
                                                    :first arg1)))))
        (if p
            (count-model (count-order-second p) arg2)))
      ;; else return end point
      arg2))
```

Debugging

```
CL-USER> (debug-detail (do-it 1))

make-match-chunk (make-TYPE*): No such chunk in DM.  Returning new chunk (not in DM) of name LOSE
Presentation of chunk LOSE (MP: NIL t=72761.26.  M: MODEL521436, t=0.
Implicitly creating chunk of name LOST.
Presentation of chunk LOST (MP: NIL t=72761.26.  M: MODEL521436, t=0.
Implicitly creating chunk of name BLANK.
Presentation of chunk BLANK (MP: NIL t=72761.305.  M: MODEL521436, t=72761.305.
make-match-chunk (make-TYPE*): No such chunk in DM.  Returning new chunk (not in DM) of name HAVE
Presentation of chunk HAVE (MP: NIL t=72761.445.  M: MODEL521436, t=72761.445.
Implicitly creating chunk of name HAD.
Presentation of chunk HAD (MP: NIL t=72761.445.  M: MODEL521436, t=72761.445.
Group PAST-TENSE-MODEL with 1+0 matching rules, choosing rule PTMODEL (Utility 5.0709996)
Group FORM-PAST-TENSE with 3+0 matching rules, choosing rule STRATEGY-WITHOUT-ANALOGY (Utility 5.225957)
retrieve-chunk:
  spec: (CHUNK-TYPE PASTTENSE VERB GET)
  cues: NIL
  pmat: NIL
filtered 0 matching chunks.
retrieved none out of 0 matching chunks.
NIL
Assigning reward 3.9
Assigning reward 3.853125 to STRATEGY-WITHOUT-ANALOGY. STRATEGY-WITH-ANALOGY remains best regular rule in group FORM-PAST-TENSE.
Assigning reward 0.0 to PTMODEL. Best regular rule among alternatives in group PAST-TENSE-MODEL!
NTI
CL-USER>
```

Implemented Models

- 10 Classic models implemented:
  - count, addition, siegler, zbrodoff, paired, fan, sticks, semantic, choice, past-tense

* past-tense not yet complete

Efficiency

- Sentence production (syntactic priming) model
  - 30 productions in ACT-R, 720 lines of code
  - 82 lines of code in ACT-UP (3 work-days)
  - ACT-R 6: 14 sentences/second
  - ACT-UP: 380 sentences/second

Scalability

- Language evolution model
  - Simulates domain vocabulary emergence (ICCM 2009, JCSR 1010)
  - 40 production rules in ACT-R (could not prototype)
  - 8 participants interacting in communities
- In larger community networks: 1000 agents, 84M interactions (about 1 minute sim. time each), 37 CPU hours

Rapid prototyping/Reuse

- Dynamic Stocks&Flows model (JAGI 2010)
  - Competition entry, model written in < 1 person-month
  - Instance-based learning (IBL, Gonzales&Lebiere 2003)
  - Blending (Wallach&Lebiere 2003)
  - free parameters (timing) estimated from example data
  - Model generalized to novel conditions
    - (.... NOT. but it did so better than others.)
- Same IBL/blending micro-strategy was re-used directly in a *Lemonade Stand Game* entry to a 2009 competition (BRIMS 2010)

Drawbacks

- Less established code-base than ACT-R 6
- Lisp
- Lack of architectural timing predictions from rule matching
- Lack of parallelism (planned: fall 2010)
- lack of perception/motor modules
  - Will be available in ACT/Simple-style interface (Salvucci&Lee 2003)

Beta-Test

- **Limited Release** of ACT-UP test version
  - comes with 10 example models
  - 4 tutorials (paralleling the ACT-R 6 ones)
  - Full API documentation plus *How-do-I...* document
- Testing period: Fall 2010
- Task: implement 1-2 models of your own
- Review letter requested (journal-review style)

5.4   Examining the Relationships Between Education, Social Networks and Democratic Support With ABM

Examining the Relationships Between Education, Social Networks and Democratic Support With ABM

Nick Drucker
MYMIC, LLC.
nick.drucker@mymic.net

Kenyth Campbell
Old Dominion University
kcamp018@odu.edu

Abstract: This paper introduces an agent-based model that explores the relationships between education, social networks, and support for democratic ideals. This study examines two factors that affect democratic support, education, and social networks. Current theory concerning these two variables suggests that positive relationships exist between education and democratic support and between social networks and the spread of ideas. The model contains multiple variables of democratic support, two of which are evaluated through experimentation. The model allows individual entities within the system to make "decisions" about their democratic support independent of one another. The agent-based approach also allows entities to utilize their social networks to spread ideas. Current theory supports experimentation results. In addition, these results show the model is capable of reproducing real world outcomes. This paper addresses the model creation process and the experimentation procedure, as well as future research avenues and potential shortcomings of the model.

1. INTRODUCTION

How do Democracies arise? It is not possible to answer this question in a simple or quick manner. Democracies tend to take decades to fully form, so studying the variables that lead to their rise is not a task one can achieve in a week. The process requires extensive examination of literature and history. However, the study of the relationships between variables that affect democracy can occur in a much shorter period using Agent Based Modeling (ABM). In order to conduct initial experimentation with this new method, this study only examines the effects of a few variables on the rise of democracy; the main variable we examined being democracy, with social networks serving as a medium for ideas to spread, allowing us to examine the effects of social networks on idea transference. Based on a study of the relevant literature, the hypothesis for this experiment is: increases in the education transfer variable will lead to an increased number of democratic supporters, over a 100 step (year) cycle.

2. LITERATURE REVIEW

A review of the current literature on democracy revealed that democratic ideals influenced by education positively affect support of democracy [13], [15], [8], [5], [4]. Democratic states and states transitioning to democracy often have strong liberal education systems. These systems help pass on the basis of democratic ideals to every generation, resulting in a population that approves of and supports democracy [7], [4]. Where these strong liberal education systems are lacking, states often experience lower levels of support for democracy [6].

Much of the literature focuses on the fact that creating a culture of democracy is important to improving democratic support. This tie is into the concept of democratic ideals, or a system of beliefs, which match with a democratic form of government [5], [8]. Consequently, where these ideals are less or completely absent, one would expect democracy to be non-existent, or the

system severely flawed. Therefore, the current theory concludes that support for democracy, on a national level, is contingent on a system of democratic ideals, which the populace receives during the education process.

In addition to education, the literature highlights the relationships between democracy and other cultural and economic influences; with economic variables being a common theme in most of the literature. While this variable appears to influence democratic ideals, in conjuncture with education, there is no clear connection for how the two relate to each other. While it would seem safe to assume that economically secure individuals would receive better educations, none of the literature clearly states this. As a result, for the purpose of this study, economic factors will remain locked, and consist only of a random distribution of wealth among agents.

An additional variable, which the literature identifies as important to the spread of ideas, is social networks. Current literature highlights several ways that an individual's social network can influence them. Many individuals find themselves in situations surrounded by others that share the same democratic ideals as they do; but they also find themselves surrounded by individuals that have different democratic ideals [18]. The likelihood that an individual will accept another's democratic beliefs is based on how strong their current beliefs are, as well as the amount of effort the other person expends trying to instill their democratic ideals. For these reasons, the literature clams there is no more than a fifty percent chance that an individual will accept influence from either side [18].

As opposed to the immediate influence that individuals receive from others within a particular propinquity, current literature also discusses how individuals accept influence from their friends and family. Unlike influences applied by individuals in a person's proximity, a person can choose the friends and family from whom they are willing to accept influence. Recent literature

argues that during the current era people are not limited to only accepting influence from friends and family that hold the same beliefs. In fact, since democratic ideals can change at a rapid rate, individuals are willing to accept influence from those with the same ideals, as well as those with differing views [16], [17].

Because of the above-mentioned literature, this study not only looked at individuals within a person's immediate proximity, but also the individuals that are involved within a person's far-reaching network. Additionally, this study did not limit individuals to accepting influence from those who share the same democratic ideals, but allowed for individuals with differing democratic ideals to influence a person.

3. METHODOLOGY

To create the model for this study we followed a two-step process: create a metacode, and then input the true code into NetLogo. NetLogo is an agent based modeling environment developed on the Java platform. The software allows users to create and program agents, giving them sets of instructions for interactions and behaviors. The user can then create interactions amongst agents, and experiment with the interactions to determine how changes in individual behavior affect the overall behavior of the model. Metacode is a rough outline of the intended process for a program, in this case an agent-based model, and represents a high-level view of how the model will function.

After we created the metacode, we began to write the program in NetLogo. While transferring the metacode into true code we often found problems that required us to add modules to the main program and in some cases change some of the basic processes. Figure 1 shows the final model format, in NetLogo, with the added variables and the final variable control formats.

Agents within the model follow a set of procedures to perform the following actions during each "step": they decide whether to

educate or not; they receive education (if they are at a location); they interact with their social network and local community; they decide whether to become a supporter, detractor or remain neutral; they perform actions to possibly give birth or die; and they move. To examine the main variable, education, agents within the system perform an initial check to determine two things: are they close enough to an education location to attend and are they the right age to attend. The radius in which agents must be to attend a location is determined using a slider (education-influence-radius), which we did not adjust for this experiment due to computer processing constraints. Along with the number of locations available for agents to receive education, we felt that increasing or decreasing these variables would result in predictable outcomes (agents would be more likely to support democracy where radius and education locations were high and vice versa). The important variable we did allow to change in order to examine the effects of education was the level of education agents received at the education locations. Agents who attended a location receive X amount of "education" each year until they reach the age of 18. Agents in the model do not have to go to an education center unless they are within the variable range determined by the education-influence-radius. Therefore, agents who "live" away from education centers (those further away from a center than the value of the education-influence-range variable) would not receive education, while those close by would. In addition, agents could receive education anytime after the age of five, until they were 18. Therefore, agents not encountering a location when their age reached the minimum could still enter a location later.

At the beginning of each run of the model, agents look within a certain radius, as well as looking to a certain number of other agents in their extended social network, to receive influence (support or detract). The model contains a multitude of options for adjusting agent's social networks. The model allows for the selection of the immediate radius from which each agent will look to for support influence. As the range of the social network increases, the agents will have more companions from which they can draw either positive or negative democratic support. Within this process, we built in a measure of randomness by ensuring the distribution of agents would result in different numbers of neighbors in each individual's range. After each step in the model, the agents move a couple of spaces in different directions; this allows agents to move in and out of the influence radius of others.

As for the extended social networks of the agents, or more simply a network that is not limited to a certain radius, there are also options that allow the user to manipulate the model. First, the user has the option to choose in which type of extended network the agents will participate. The three options are normal, uniform, or constant distributions. The normal distributions assign each individual a number of agents to participate in their extended network using the normal distribution to determine the exact number. The uniform option uses a simple random procedure to determine the number of agents within a certain individuals extended network. The uniform distribution does not follow the bell curve but allows every number in the random number range to have an equal opportunity of being selected, resulting in random numbers of agents in each network. Lastly, the constant distribution gives all agents the same number of individuals within their extended network.

During every time step of the model, agents look within their social network, which includes their immediate radius and extended social network, and determine agents from which they will accept either democratic support or non-democratic support. In order to do this, the model is designed to follow the assumptions described in the literature, and a coin-flip procedure determines whether the agents accept influence (i.e. agents have a 50/50 chance of accepting or rejecting influence). This works the same for both democratic

influence and non-democratic influence. If an agent accepts influence the amount they accept, which remains constant, is either added or subtracted from the democratic support variable. Agents receive greater influence from their non-immediate network (representative of their family and friends) than from their neighborhood. We made this decision because influences an individual receives from family and friends tend to be stronger than influences they accept from strangers. If agents accept democratic influence, the support variable increases, but if they accept non-democratic influence, the support variable decreases.

In addition to the two main variables we examined, agents also performed checks to gain or lose wealth and to decide whether to support democracy or not. Because economics was not a focus of our experiment or our hypothesis, but is an important variable in democratic support, we included a procedure to allow agents to gain or lose wealth. For simplicity sake we used very basic procedures to allow agents to gain or lose wealth; if an agent is in the upper 15% of the population in wealth they have a greater chance to gain more wealth, while those below the 50% median have an equal chance to loose or gain wealth. We felt this procedure was necessary to represent the fact that individuals with large amounts of wealth are better able to protect their wealth and may be able to continue to gain it, while those with less wealth have a harder time protecting and gaining wealth.

Agents follow a procedure of checking their wealth, education and support levels to determine if they will became a supporter or detractor. We set thresholds for these variables (for support and wealth they did not change) and agents check all three, deciding to be a detractor if they fell below all three-threshold levels, and deciding to be a supporter if they were above. As explained previously, agents accept support or non-support from their neighbors and social network. We included this variable and allowed it to shift in order to provide a way to examine the effects of social networks, and to allow agents to decide

whether to support democracy based on variables other than just wealth or education. Because it is not realistic to assume that all educated and wealthy people will automatically support democracy, we included the democratic support variable to allow agents to decide not to become supporters, even if they were wealthy and educated.

Fig. 1
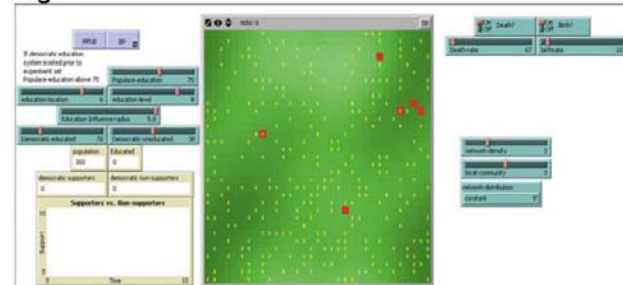
4. RESULTS

In order to experiment with this model we ran 12,961 trials using a variety of variable settings. Utilizing the behavior space feature within NetLogo, we were able to sample six variables across multiple settings. For several variables (Populace-education, Democratic-educated, and Democratic-uneducated) we did not sample the entire variable range due to time constraints. In addition, we did not include the remaining sliders and switches (education-location, death rate, birth rate, and network distribution) in this experiment because we did not wish to test their direct effects on democracy.

Results of our experiment showed that overwhelmingly democratic supporters outnumbered democratic detractors (91% of the time).

	Supporters outnumber Detractors	Detractors outnumber Supporters	Total number of runs
Runs	11,795	1,166	12,961

Table 1: Total times each group outnumbered the other

587

In addition, the average percent of agents in the system that were supporters was 33%, while only 5% of the agents were detractors.

Average number of Supporters	Average number of Detractors	Total number of agents
90	15	300

Table 2: Average results of a run

Across multiple runs, this demonstrates that in almost all instances democratic support arose within the model, across multiple variable settings. However, the 9% of runs, which resulted in democratic detractors being the majority, demonstrated that variable settings did affect the outcome.

In order to gain a better understanding of what our results showed, we constructed a linear regression model of the results, finding that all variables except network density had a significant effect on democratic supporters (results appear in appendix 1). Our regression model included all variables from the model that changed (Populace-education, Democratic-educated, and Democratic-uneducated, Network-density, and local-community). The resultant adjusted R-square value of 0.544 shows that this model was robust and captured a large portion of the variability within the model. In addition, the high F value (2581.219) shows that the model was significant.

The regression coefficients from this model showed that almost all variables had the expected relationships with our dependent variable (based on the literature review). One variable which did not demonstrate an expected relationship with democratic support was the detractor threshold variable. The regression model showed that as the threshold to become a detractor rose, democratic supporters in the model fell. While this result appears counterintuitive, the regression analysis does not take into account the overall number of decorators and supporters in the model (i.e. even as decorators within the system fall due to a higher threshold,

democratic supporters in the system do not necessarily rise). This result demonstrates that the two groups, detractors and supporters, do not vary based on each other's numbers. This finding supports our belief that the model adequately captures real world behaviors. Had the regression analysis shown that these variables had opposite relationships with the dependent variable, it would suggest that they might have an effect on each other as well. For this model to be accurate, the number of democratic supporters or detractors should not influence the other beyond moderately affecting the size of the influential population pool. This result is in no way conclusive that the two variables are not connected, but it does indicate their limited connectivity, which implies the number of supporters and detractors within the model is mostly able to vary independent of one another.

The other variables within our model-demonstrated relationships that the literature suggests should exist. All three remaining variables that were significant had positive relationships with democratic supporter numbers. While one of the variables we focused on (education) had a positive significant relationship with democratic supporters, the variables relating to social networks were not both significant. The variable representing the agent's social network external to their location (i.e. those they agents not in direct or near direct contact with) was not significant. However, because agents were able to move within the system, this is likely the cause for the local community (agents in direct or near direct contact with each other) having an effect on the outcome. The limit of social networks in this model is that they do not expand as agents encounter each other; since the social network is not able to expand throughout the agents "life", it has a fixed effect on the outcomes, which appears to be insignificant. In order to verify this finding we would need to re-run this experiment and allow the agents network density to vary across several distributions to determine if the effect is fixed or not.

5. CONCLUSION

The results of this model, across multiple variable settings, indicate that the model agrees with current theory. The fact that democratic supporters did not always outnumber detractors also indicates that the outcomes are not hard-coded into the model. While the sensitivity does appear somewhat low (as demonstrated by the fact that democratic supporters outnumbered detractors 91% of the time), the model did not produce an overwhelming majority of one outcome or the other. If we had added other variables or added additional variable settings in the experiment, it is likely that the results would likewise have varied. This is especially true for the number of education locations. In order to further test the effect of education on democratic support we would allow the number of locations to vary, and examine how this affects the model's results.

The model also captured the relationship between the spread of democratic ideals and social networks. In addition to the effects education had on democratic support, the relationship between individual agents and their immediate community reveal that external influences also play a large role in determining an individual's views of democracy. However, the model does seem to represent a very small majority of real world situations. The inclusion of social networks, which reach across distances greater than an immediate "neighborhood", is more representative of a country with advanced telecommunication networks. Because countries do not all possess advanced communication networks allowing all their citizens to communicate with friends and family over vast distances instantaneously, the model is not representative of all possible states. However, we could replicate countries without these advanced communications networks by removing the network density variable from experimentation.

While this model is not representative of all countries, which will take further experimentation and testing to correct, it does match well with the current state of the world. In countries with advanced communication networks and good education systems, the predominant form of government is democracy. Our model adequately reflects this, demonstrating that the model is relatively accurate, in terms of recreating real world situations. We expect that removing social networks and running the experiment again would likewise affect the model and produce results more representative of countries without advanced telecommunication networks.

Another possibly inaccurate aspect of the model is the number of education locations we allowed to exist. For the purpose of this experiment, we decided to vary the level of education agents received and not the number of sources where they could receive their education. We expect that lowering or raising the number of locations will have the same impact on the number of democratic supporters. Based on the construction of the model, a high number of locations will inherently affect more agents and introduce more education into the model. We therefore decided to remove this variable from this experiment as we expected its impact would be too great on the outcomes of our experiment. In future tests, we would include this variable and examine how it affects the models results. Should it produce results differing from what we expect, it would provide interesting insight into how the number of education locations available to individuals may positively or adversely affect their education.

In terms of validation, this model appears to be a valid representation of the real world. However, we could not identify a real world case to compare our results too, in furtherance of this conclusion. In future validation procedures we plan to empirically test this model against real world cases of countries where democratic support is the majority opinion of the people, and somewhere it is not. If through further validation our model proves to be an accurate representation of real world situations then our results would further reinforce current theory concerning

589

education and democracy. The results would also support the notion that social influences are an important factor in determining an individual's support of democracy. Our results from this experiment support this notion and suggest that individuals are heavily influenced both by the people they encounter in contact with, and by the education they receive.

6. APPENDIX
6.1 Regression Tables
Democratic Supporters Regression Model Tables

Democratic Support Regression Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.738[a]	.545	.544	83.483

a. Predictors: (Constant), local-community, network-density, education-level, Popeducbegin, detracthreshold, demthreshold

Democratic Support Regression ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.079E8	6	1.799E7	2581219	.000[a]
	Residual	9.028E7	12953	6969.465		
	Total	1.982E8	12959			

a. Predictors: (Constant), local-community, network-density, education-level, Popeducbegin, detracthreshold, demthreshold

b. Dependent Variable: supporters

Democratic Support Regression Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	172.613	4.549		37.945	.000
	demthreshold	-2.651	.045	-.350	-59.028	.000
	detracthreshold	-.314	.045	-.042	-7.002	.000
	Popeducbegin	1.625	.033	.294	49.537	.000
	education-level	16.280	.216	.446	75.284	.000
	network-density	-.052	.357	.000	-.144	.885
	local-community	27.800	.449	.367	61.906	.000

a. Dependent Variable: supporters

7. REFERENCES
[1]. Anderson, C. J., and Guillory, C. A. (1997). "Political Institutions and Satisfaction with Democracy: A Cross-National Analysis of Consensus and Majoritarian Systems." *The American Political Science Review*. Vol. 91, No. 1: pp. 66-81.

[2]. Angle, S. C. (2005). "Decent Democratic Centralism." *Political Theory*. Vol. 33, No. 4: pp. 518-546.

[3]. Canetti-Nisim, D. (2004). "The Effect of Religiosity on Endorsement of Democratic Values: The Mediating Influence of Authoritarianism." *Political Behavior*. Vol. 26, No. 4: pp. 377-398.

[4]. Evans, G., and Whitefield, S. (1995). "The Politics and Economics of Democratic Commitment: Support for Democracy in Transition Societies." *British Journal of Political Science*. Vol. 25, No. 4: pp. 485-514.

[5]. Finifter, A. W., Mickiewicz, E. (1992). "Redefining the Political System of the USSR: Mass Support for Political Change." *The American Political Science Review*. Vol. 86, No. 4: pp. 857-874.

[6]. Finkel, S. E., and Ernst, H. R. (2005). "Civic Education in Post-Apartheid South Africa: Alternative Paths to the Development of Political Knowledge and Democratic Values." *Political Psychology*. Vol. 26, No. 3: pp. 333-364.

[7]. Finkel, S. E., Humphries, S., and Opp, K.D. (2001). "Socialist Values and the Development of Democratic Support in the Former East Germany." *International Political Science Review*. Vol. 22, No. 4: pp. 339-361.

[8]. Gibson, J. T., Duch, R. M., and Tedin, K. T. (1992). "Democratic Values and the Transformation of the Soviet Union." *The Journal of Politics*. Vol. 54, No. 2: pp. 329-371.

[9]. Korzeniewicz, R. P., and Awbery, K. (1992). "Democratic Transitions and the Semiperiphery of the World-Economy." *Sociological Forum*. Vol. 7, No. 4: pp. 609-640.

[10]. Muller, E. N., Jukam, T. O. (1977). "On the Meaning of Political Support." *The American Political Science Review*. Vol. 71, No. 4: pp. 1561-1595.

[11]. Peffley, M., and Rohrschneider, R. (2003). "Democratization and Political Tolerance in Seventeen Countries: A Multi-Level Model of Democratic Learning." *Political Research Quarterly*. Vol. 56, No. 3: pp. 243-257.

[12]. Reisinger, W. M., Miller, A. H., Hesli, V. L., and Maher, K. H. (1994). "Political Values in

Russia, Ukraine and Lithuania: Sources and Implications for Democracy." *British Journal of Political Science*. Vol. 24, No. 2: pp. 183-223.

[13]. Rorschneider, R. (1994). "Report from the Laboratory: Influence of Institutions on Political Elites' Democratic Values in Germany." *The American Political Science Review*. Vol. 88, No. 4: pp. 927-941.

[14]. Rosenblum, N. I. (2003). "Religious Parties, Religious Political Identity, and the Cold Shoulder of Liberal Democratic Thought." *Ethical Theory and Moral Practice*. No. 6: pp. 23-53.

[15]. Seligson, M. A., Booth, J. A. (1993). "Political Culture and Regime Type: Evidence from Nicaragua and Costa Rica." *The Journal of Politics*. Vol. 55, No. 3: pp. 777-792.

[16]. Philpot, Tasha S. (2004). "A Party of a Different Color? Race, Campaign Communication, and Party Politics." *Political Behavior*, Vol. 26, No. 3.

[17]. Best, Samuel J., and Brian S. Krueger. (2005). "Analyzing the Representativeness of Internet Political Participation." *Political Behavior*, Vol. 27, No. 2.

[18]. Burden, Barry C., and Casey A. Klofstad. (2005). "Affect and Cognition in Party Identification." *Political Psychology*, Vol. 26, No. 6.

8. ACKNOWLEDGMENT

591

Education, Social Networks and Democratic Support

Nick Drucker

Kenyth Campbell

MODSIM WORLD 2010

Purpose

- Examining how we can leverage modeling and simulation techniques in the International Relations field
- We were interested in the emergence of democracy supporters and detractors from a neutral population

Research Question

What are the effects of education on the emergence of democratic supporters in a socially connected country?

Agent Based Modeling

- Agents make decisions based on a set of rules
- In the Democracy Emergence model, agents were:
  - Supporters
  - Detractors
  - Neutral
- Supporters and Detractors influence the neutral population, who are eligible to receive education under certain circumstances

Model Logic

- Agents make decisions based on three variables
  - Education is determined by the independent variables in the model
  - Social network provides direction for idea transference
  - Wealth is uniformly distributed and changes in a fixed manner
- Education was the influential variable and was determined by:
  - Radius of education centers influence
  - Amount of education transferred from a center to an agent
  - Agents initial education

Model Logic

- Thresholds for all three variables must be met to support or detract
- Supporters and detractors influence their social networks
- 100 step process with agents dying and being born continuously

Social Network

- Consists of two parts
  - Local radius variable (local-community)
  - Extended network variable (network-density)
- Each of these variables has agents look to other agents for potential influence.
  - Local radius variable has agents look to neighbors within a certain radius.
  - Extended network variable has agents look to friends/family within a social network.

Social Network

- Potential influence is accepted via a coin-flip scenario.
- Influence received from other agents can either be positive or negative.
- Influence is transformed into a support variable that is part of determining whether agents become supporters or dissidents.

Assumptions and Expectations

- Agents exist in a electronically connected environment (e.g. They are able to communicate instantaneously with their distributed network)

- Once an agent is a supporter or detractor they can no longer be influenced

- There is a positive relationship between education and democracy; this assumption is based on current theory

Democracy Emergence Model
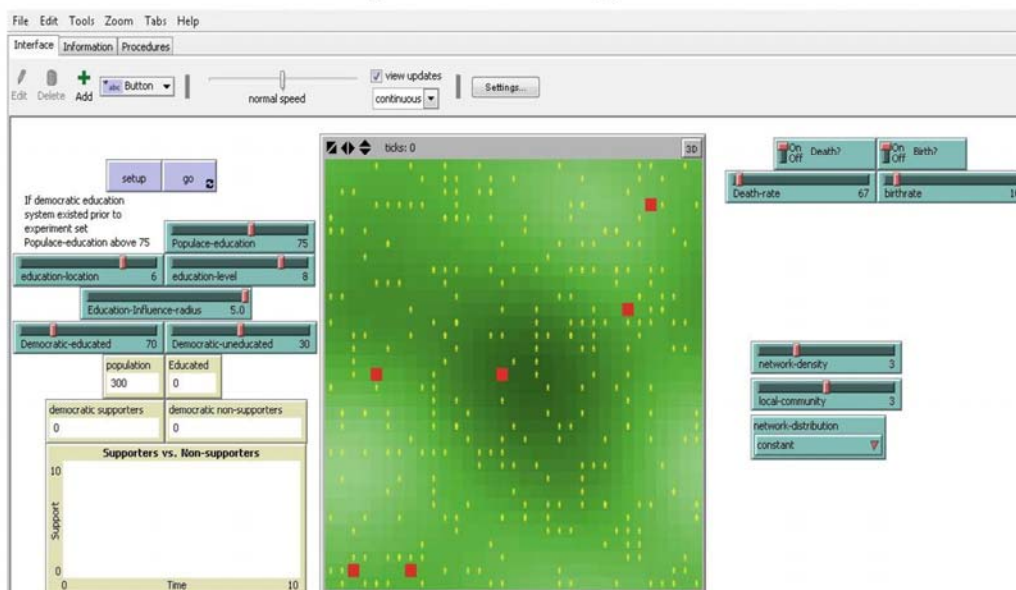
Experimentation

- Set parameters for independent variables (upper and lower limits)
  - Ex. Education influence radius (0-10)
- Established points within the parameters to sample
  - Ex. Education transference variable (2, 4, 6, 8, 10)
- Ran 12,961 trials with different combinations

Results

- 91% of runs resulted in a higher level of support for democracy than initial conditions
- On average 35% of agents in the system were democratic supporters; 5% were detractors
- Linear regression analysis revealed relationships that support the literature
  - Higher education transfer, lower education support threshold, and higher initial education resulted in more supporters
  - Lower threshold to become a detractor did not result in less democratic support (unexpected)

Conclusions

- We were successfully able to apply ABM principles to an International Relations topic
- The model aligns closely with literature
  - Appears to capture real world situations
- Democratic support was not the only end state outcome, non-democratic support was also observed

Future Research
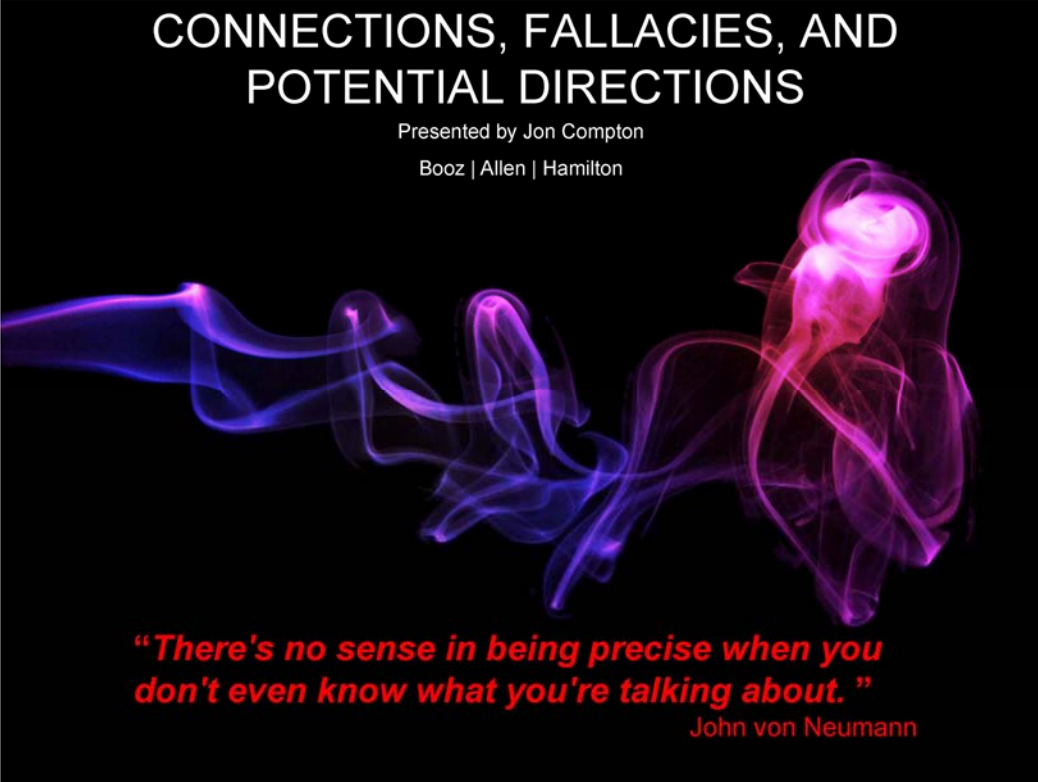
- Apply to other Socio-Cultural issues
- Applying it to additional variables that impact Democracy and its emergence

Questions?

5.5   Connections, Fallacies, and Potential Directions

Lies, damned lies, and statistics

A study published in the American Journal of Clinical Nutrition shows that breastfed infants tested 5.2 IQ points higher than formula fed infants, for a comprehensive study involving 11 different studies and over 7000 children.

"Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital." ~Aaron Levenstein

Forest but for the trees, trees but for the leaves

One of the major problems faced by the early pyramid builders was the need to move huge quantities of rock. While 80 men can drag a 2.5-ton block of stone on a sled, as depicted in carvings in some later Egyptian tombs, this brute-force method was not very efficient.

"Two quite opposite qualities equally bias our minds - habits and novelty" ~Jean de la Bruyere

"When knowledge is well guarded, it's easily lost." ~Anon.

What econometricians know….

Without doubt the most widely cited review is the classic Meta-analysis of research on the relationship of class size and achievement (Glass & Smith, 1978). The two primary conclusions drawn from this material are:

  * reduced class size can be expected to produce increased academic achievement

  * the major benefits from reduced class size are obtained as the size is reduced below 20 pupils

"I never let schooling interfere with my education." ~Mark Twain.

The Terms of the Discourse….

The Climate change debate:

Do you believe in:

      Climate Change?

      Global Warming?

      Anthropogenic Global Warming?

      Catastrophic Anthropogenic Global Warming?

And just who are the "flat Earthers" and just what don't they believe in?

"The most perfidious way of harming a cause consists of defending it deliberately with faulty arguments." ~ Friedrich Nietzsche

Where theory and culture diverge

Prior to the development of consciousness, Julian Jaynes argues humans operated under a previous mentality he called the bicameral ('two-chambered') mind. In the place of an internal dialogue, bicameral people experienced auditory hallucinations directing their actions, similar to the command hallucinations experienced by people with schizophrenia today. These hallucinations were interpreted as the voices of chiefs, rulers, or the gods.

"The weight of original thought in it is so great that it makes me uneasy for the author's well-being…." ~D.C. Stove on Jayne's *The Origin of Consciousness*

The Terms of the Discourse

What is the definition of consciousness?

• Abstract thought?

• Memory?

• Language?

"*The visionary lies to himself, the liar only to others.*" ~ *Friedrich Nietzsche*

Even supermen must amuse themselves (sometimes at the expense of the lives of others)
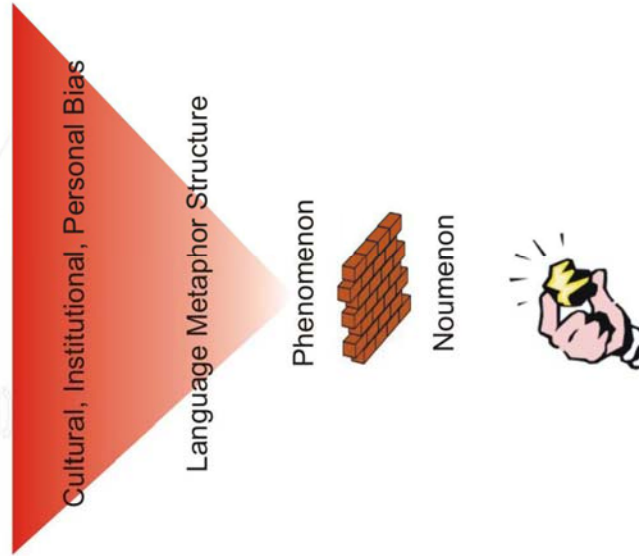
CHUMP
(Compton's Heuristic Ultracrepidarian Macromantic Pantomancer)

The beauty of the rationality assumption (a tired tale indeed)

But what is the real problem?
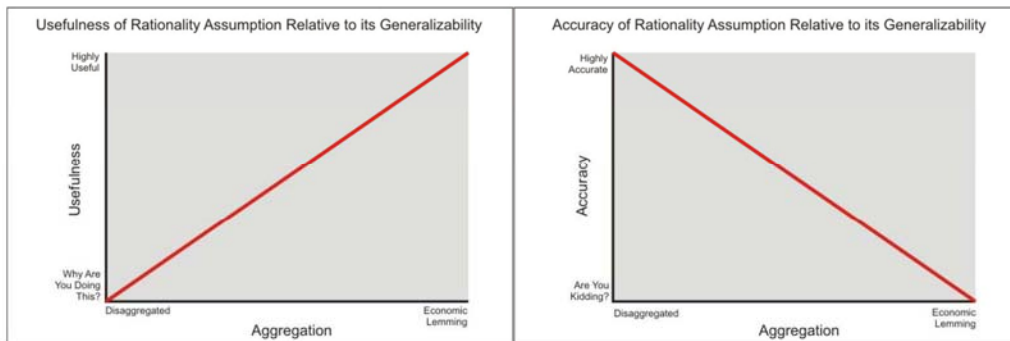
Do you play chess?

"The irrationality of a thing is no argument against its existence, rather a condition of it." ~ Friedrich Nietzsche

CONNECTIONS

No, it's the theory, stupid

ASSUMPTIONS

NEVER ASSUME WHAT YOU'RE TRYING TO PROVE,
UNLESS YOU'RE TRYING TO PROVE YOU'RE A BONEHEAD.

CASES

Who are these folks and what do they have in common?

Okay, this one is cheating… Hiroko Nagata

Terrorism: common assumptions

1. Terrorists are rational actors (Richardson)

2. Terrorism is largely the result of poverty or unequal distributions of wealth (Crenshaw)

3. Terrorists act to achieve specific political goals (Richardson)

4. Terrorism is a response to external pressures, such as foreign occupation (Pape)

5. Terrorism occurs for nationalist or separatist reasons (Williams)

6. Terrorism is a result of religious extremism (take your pick)

"Fighting terrorism is not unlike fighting a deadly cancer. It can't be treated just where it's visible - every diseased cell in the body must be destroyed." ~ David Hackworth

What are the problems with our common assumptions?

The seven problems identified by Max Abrahms

1. Terrorism fails to achieve the stated goal almost all of the time

2. Terrorism is almost never used as a last resort

3. Terrorist organizations almost always reject compromises despite significant policy concessions

4. Political goals of terror organizations are, without exception, protean

5. Terrorist attacks are usually anonymous

6. Competing terror groups with identical or highly similar goals generally prefer to attack each other than any other target

7. Terror groups seldom disband despite the consistent failure of the tactic to actually accomplish their objectives

"Everybody's worried about stopping terrorism. Well, there's a really easy way: stop participating in it." ~Noam Chomsky

"Noam Chomsky is a dumbass" ~Sebastian Synclair

606

What are the commonalities?

What are the common factors in these groups?

Identity Entrepreneur

Each has a founding member with charismatic appeal, a perception of injustice, and a…

Commitment to Violence

Each group has a core of people who hold a commitment to violence that shapes the goals and philosophy of the group (not the other way around). This core attracts people who seek…

Social Affiliation and Identification

The bulk of these groups are filled out with people who lack the strong commitment to violence, but self identify with the lifestyle, philosophy, social atmosphere, and so forth, of the group.

"I don't worry about terrorism. I was married for two years." ~Sam Kinison

---

Who are these folks and what do they have in common?

What are the commonalities?

What are the common factors in these people?

Traumatic Experience Leading to Anger

Each has had some form of trauma as a child which spurred deep-seated resentment.

Strong Personal Commitment to Violence

Once begun, the commitment to violence against their chosen victims over-rides inhibitions.

Objectification of Victims

To accomplish the over-ride of inhibition to violence, victims are objectified—held apart—from their perception of humanity.

"People don't know me. They think they do, but they don't." ~Andrew Cunanan

---

Wait, What?

?

What are the common factors?

Traumatic Experience Leading to Anger

Strong Personal Commitment to Violence

Objectification of Victims

This is how we kill.

But…

What about the capacity for obsession?

608

Brain Hemispheres

Characteristics:

Narrow Focus

Linear

Non-multi-tasking

Decontextualizing

Abstractive

Specialized
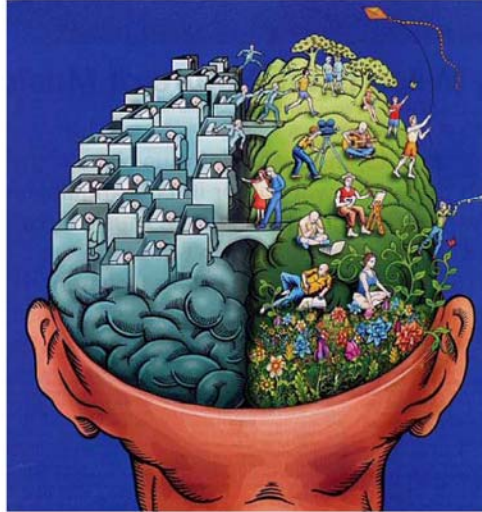
Characteristics:

Broad Focus

Non-Linear

Multi-tasking

Contextualizing

Synthesizing

General

This is a TEST!

Left Brain - Right Brain Conflict

Look at the chart and say the COLOR not the word

YELLOW BLUE ORANGE
BLACK RED GREEN
PURPLE YELLOW RED
ORANGE GREEN BLACK
BLUE RED PURPLE
GREEN BLUE ORANGE

Left - Right Conflict
Your right brain tries to say the color but your left brain insists on reading the word

According to Iain McGilchrist, the hemisphere's of the brain vie for dominance in the individual, and that struggle manifests itself not only in the individual, but in society at large. This is because environmental conditions influence which side is dominant, and there are feedback effects.

609

How Do the Hemispheres Communicate?

Remember Jaynes?
If Bicameral Mind theory is correct, the development of Complex metaphor in language enabled modern consciousness.

What do Linguists think of that?

The Sapir-Whorf Hypothesis

Weak hypothesis: Structural differences between language systems will, in general, be paralleled by nonlinguistic cognitive differences, of an unspecified sort, in the native speakers of the language.

Strong hypothesis: The structure of anyone's native language strongly influences or fully determines the worldview he will acquire as he learns the language.

" Language forces us to perceive the world as man presents it to us." ~Julia Penelope

---

Cracked Pots....

If these observations and theories are correct:

1. There should be a correlation between language and hemispheric dominance in the brain.

2. Strong left-brain dominance in population groups should correlate with higher likelihoods of violent behavior, rational or otherwise.

3. Permissive or non-permissive environments should moderate the levels of violence originating from a particular social or cultural group.

4. These hypotheses should be testable.

5. We should, therefore, be able to identify population groups with higher likelihoods of violent behavior, and be able to apply policy decisions based upon the degree of permissiveness in the environment.

"Follow the path of the unsafe, independent thinker. Expose your ideas to the danger of controversy. Speak your mind and fear less the label of "crackpot" than the stigma of conformity." ~Thomas John Watson, Sr.

"Arnold has had his spokesman call me a crackpot. That was a mistake." ~Warren Beatty

Synthesis…..
This Pot Has Definitely Cracked….

Herein lies the end of the nutty part. The rest is boring.

Thank you for your attention.

Comments are always welcome.

If you want to see the rest, I'm happy to go through after the session.

"Quotation is a serviceable substitute for wit." ~Oscar Wilde

---

CASES
What does the data say?

Incidents with Percapita GDP, GINI, and RPC

A way to conceptualize… The power of metaphor

Gestation	Growth	Maturity	Transition

Membership

Violent Systems Theory: terror groups as living organisms

Entails:

- Life cycle
- System congruence
- Negative entropy

Overriding goal: SURVIVAL

System Properties

Negative Entropy
Required for VNSA survival. Inputs sought until input pool is drained.

Congruence
Subsystems (3rd system level) must function in reinforcing manner, optimizing coordination and information exchange.

Life Cycle

Gestation	Growth
Identity Entrepreneur (Environment Shaping)	Identity Mobilization (Traditional Deterence)
Transition	Maturity
Legitimacy or Death	System Characteristics (Rationality Deterence)

ETA Life Cycle

Gestation · Growth · Maturity · Transition

ETApm – ETAm Split

LEGITIMACY?

ETA founded 31 July 1959

ETA activities focused on Destruction of Spanish Infrastructure and display of Basque national symbols

Attempted Derailment of train carrying Spanish politicians

Guardia Civil Officer killed in 1968, First killing by ETA

Assassination of Meliton Manzanas Chief of Secret Police, San Sebastian in retribution. Shot to the head. 1968

Adm. Luis Carrero Blanco Tapped successor of Franco killed in a car bomb attack by ETA December 1973 First high-profile killing First car-bomb.

Bomb placed in Madrid café 13 September 1973

Democratic National Parliament Formed 1977

'78 – '80 Period of heaviest Levels of violence

Basque Parliament Formed 1980

Series of car bomb attacks in Madrid and Barcelona primarily

Assassination plot on King Juan Carlos fails 1995

"Democratic Initiative" is proposed by ETA Calling for the Spanish Government to Accept absolute Basque autonomy

Basque Councilor Miguel Angel Blanco is abducted and assassinated 1997

ETA continue to use bombs As primary weapon for attacks. Warnings are sometimes given

612

FARC Life Cycle

ETA and Trend Line

FARC

613

A way to conceptualize… The power of metaphor

Counter VNSA Strategies

| Coercive Strategies | Conquer Strategies |

ENVIRONMENT LEVEL   VNSA *qua* SYSTEM LEVEL   VNSA SUBSYSTEM LEVEL

LEVEL OF VIOLENCE

The one size fits all fallacy

The factors mentioned on the previous slides are common among terrorist groups. But they do not necessarily extend to:

Organized Crime

Insurgency

Militias

Civil Wars

…

There is no silver bullet; aggregation breeds inaccuracy.

Non-state violent actors are defined by their acts, not by their intentions.

"Historically, terrorism falls in a category different from crimes that concern a criminal court judge." ~Jurgen Habermas

Tools (bet you don't know who this is)

"Um dem Terrorismus kein weiteres Terrain zu überlassen, ist der Staat von vornherein gezwungen, feste, von keiner Seite überschreitbare Grenzlinien des rechtsstaatlich Möglichen zu ziehen."

In its effort to leave no free ground for Terrorism, the State is, from the outset, firmly hindered by the impenetrable barriers of its legal system." ~ (translation by mom)

Tools: Group

Descriptive:

    Self Similarity

    Power Laws

    Case studies based upon outputs, not intentions

Fractal Geometry?

    See Example Prisoner's Dilemma Game with tit-for-tat strategy

Important Characteristics and Take Aways:

    Equilibrium states exist (power laws)

    Networks are bounded and constrained (self similarity)

But remember: **DESCRIPTIVE DESCRIPTIVE DESCRIPTIVE**

"My master had power and law on his side; I had a determined will." ~Harriet Ann Jacobs

Tools: Individual

Psychological

    Charismatic

    Well delimited Identity

    Strong perception of injustice

Depression Study

    Depression linked to problem solving portion of brain

    Excessive problem obsessions linked to depressive feelings

    Decisions/Actions relieve depression feelings

    Cycle works in both directions

    The bright side of being blue: depression as an adaptation for analyzing complex problems, Andrews and Thomson

But remember: **DESCRIPTIVE DESCRIPTIVE DESCRIPTIVE**

"Man is the only animal for whom his own existence is a problem which he has to solve"
~Erich Fromm

Tools: How do we think about the problem?

"Step Back Thinking"

Move backward from the intractable problems until the issues begin to fall within the capacity to act.

Examples:

1. We don't have to worry about predicting whether or not a sub-national group will use an NBC weapon if they can't get one

2. We don't have to predict the supply routes or behaviors of drug runners if there is no demand for the product

3. I don't have to predict where traffic is likely to be difficult if I telecommute

"When it is obvious that the goals cannot be reached, don't adjust the goals, adjust the action steps." ~ Confucius

Tools: How do we crack the nut?

1. Ruthlessly track and find our own vulnerabilities
2. Catalog, don't predict
3. Requires:
    1. Discover Identity Entrepreneur as early as possible
    2. Ascertain network structure
    3. Identify strategies that
        1. Attack IE
        2. Disrupt congruence of network
        3. Shape environment

"Our limitations and success will be based, most often, on your own expectations for ourselves. What the mind dwells upon, the body acts upon." ~Denis Waitley

---

To Predict or Not to Predict:

The Art of Prediction and the Power of Games

"No one could predict 9/11."

"Life imitates art far more than art imitates life." ~ Oscar Wilde

Tools: Wargaming?

Catalog, don't predict means:

1. Large n wargames (MS Flight Simulator)

2. Virtual world or virtual reality

3. Environmental, self adjudication

4. Catalog strong outliers (learn vulnerabilities, ours and theirs)

Second Life with guns!

1. Real world analogs

2. Real world consequences

3. Put skin in the game!

"If some unemployed punk in New Jersey, can get a cassette to make love to Elle
McPherson for $19.95, this virtual reality stuff is going to make crack look like Sanka."
~Dennis Miller

Thank you for your attention.

Comments welcome.

"Quotation is a serviceable substitute for wit." ~Oscar Wilde

5.6   Design and Evaluation of a Cross-Cultural Training System

Design and Evaluation of a Cross-Cultural Training System

Thomas Santarelli & Kevin C. Stagl
CHI Systems, Inc.
{tsantarelli, kstagl} @chisystems.com

Cross-cultural competency, and the underlying communication and affective skills required to develop such expertise, is becoming increasingly important for a wide variety of domains. To address this need, we developed a blended learning platform which combines virtual role-play with tutorials, assessment and feedback. A Middle-Eastern Curriculum (MEC) exemplar for cross-cultural training U.S. military personnel was developed to guide the refinement of an existing game-based training platform. To complement this curriculum, we developed scenario authoring tools to enable end-users to define training objectives, link performance measures and feedback/remediation to these objectives, and deploy experiential scenarios within a game-based virtual environment (VE). Lessons learned from the design and development of this exemplar cross-cultural competency curriculum, as well as formative evaluation results, are discussed. Initial findings suggest that the underlying training technology promotes deep levels of semantic processing of the key information of relevant cultural and communication skills.

1.0 INTRODUCTION

Modern political confrontations involving the use of military force are being conducted by multinational coalition partners in multiple, concurrent foreign theaters. The parties to those arrangements, and the contexts in which those operations are conducted, have led military forces to be increasingly concerned with cross-cultural training as they pertain to the execution of warfare, counterinsurgency, peacekeeping and reconstruction efforts (see Field Manual (FM) 3-07 Stability and Support Operations; FM 3-24 Counterinsurgency Operations). For example, recent military operations in Afghanistan, Bosnia and Iraq have highlighted the centrality of cross-cultural acumen to securing and sustaining stability and strategic relationships.

The lessons learned from those global deployments are invaluable. Recent multinational, urban-based operations, however, have added a new layer of complexity; requiring soldiers to proactively anticipate, interpret and influence the thoughts, behaviors and affect of their coalition partners, host sponsors and heterogeneous populations [1] Moreover, transformational changes such as network-centric warfare have also increased the complexity of operations; mandating increased inter-service and inter-agency coordination and thereby greater adaptation to fluid organizational cultures [21].

Recent attention has focused on the creation of military training systems utilizing gaming technologies. Many such systems have been developed to address domains such as Arabic language training, intelligence data collection, squad-level tactics, cultural familiarization, and leadership training [12, 3, 16]. An instructional trend in military game development has followed the paradigm of experiential learning [14] which emphasizes the role of experience in learning.

This has shaped the development of training games to be highly interactive and free-form, with trainees being allowed to explore the virtual environment, to gain experience through interactions with virtual characters and the performance of domain tasks, and to develop knowledge as the net effect of all interactions with the training system. Learning is largely opportunistic and no systematic approach has been applied to insure adequate exposure of trainees to all training objectives of a structured curriculum [20].

This problem is further compounded by the general lack or poor implementation of typical training mechanisms (e.g., articulation of training goals and performance standards, detailed feedback, performance remediation, performance appraisal, and explicit coaching prior to learning opportunities) within the current generation of military, game-based training

619

applications. These deficiencies within many contemporary VE-based training solutions are critical gaps as research suggests that if a scenario is linked with training objectives, then trainees are more likely to learn the underlying objectives [5]. These deficiencies were addressed in the development of the Middle-East Curriculum (MEC) discussed below.

2.0 BODY
The MEC design began with a training analysis of a number of sources in order to derive a set of knowledge/skill/abilities (KSAs). From those KSAs, a training objective taxonomy was derived. An underlying cultural training model was developed and a set of assessments and scenarios were created. This process is briefly described in the following sub-sections.

2.1 Knowledge/Skills/Abilities
A rich source of KSA candidates were identified in a Defense Equal Opportunity Management Institute (DEOMI) research report, entitled 'Toward An Operational Definition Of Cross-Cultural Competence From Interview Data' [18]. This data was derived from a corpus of interview data with Non-Commissioned Officers (NCOs) with recent Iraq deployment experience. The survey established a cross-cultural competency (CCC) definition and posits a number of factors impacting efficacy with respect to real-world CCC. These factors were validated through qualitative survey of subject participants. Ross' CCC-based KSA taxonomy included:

- Ethnocultural empathy
- Experience
- Flexibility
- Interpersonal and communication skills
- Mental-model/perspective taking
- Willingness to engage
- Low need for cognitive closure
- Relationship building
- Self-efficacy

- Self-regulation/emotion-regulation

These factors were subsumed within the MEC training design and used to identify important cross-cultural KSAs.

2.2 Training Objective Derivation
The Middle East curricular content was developed based on analyses of widely-available cultural training materials. This included the Defense Language Institute (DLI) cross-cultural training materials (see http://www.dliflc.edu/products.html) as well as other available sources of Middle Eastern reference content. The full set of sources for cross-cultural familiarization which were drawn upon included: TRADOC Culture Center Arab Cultural Training Curriculum, TRADOC DCSINT Handbook No. 2: Arab Cultural Awareness, TRADOC DCSINT Contemporary Operation Environment: Actors & Role, DLI Countries in Perspective (Iraq) and Iraqi Familiarization, Peacecorps Culture Matters workbook, State Department Iraq Fact Sheets, and the CIA World Fact Book (Iraq/Afghanistan).

A training analysis was conducted with these materials and a training objective hierarchy was created to guide the development of specific curricular content such as didactic materials, assessments, and VE-based scenarios. More importantly, this training objective analysis was used to develop a cultural training model, as described below.

2.3 Cultural Training Structure
There have been a number of initiatives conducted to better understand the nature of effective cross-cultural training for multinational operations. For example, the U.S. Army Training and Doctrine Command (TRADOC) requested the Director of Center for Army Leadership, Combined Arms Center and the Chief of the Leader Development Research Unit at ARI to host and conduct the Cultural Understanding and Language Proficiency (CULP) research initiative. A workshop supporting CULP was

620

conducted in July 2007 at Ft. Leavenworth, Kansas in order to provide a clearer understanding of the cultural and linguistic capabilities required by soldiers [2].

CULP participants were guided by a tri-component framework of cultural capability. The tenets of this framework suggest cultural capability consists of: (1) cross-cultural competence, and to a lesser extent: (2) culture-specific knowledge and (3) language proficiency. CULP participants were charged with defining cross-cultural competence, as it was deemed the most important factor to mission success, and identifying the knowledge, skills and attitudes required to achieve intercultural effectiveness. In response to this agenda, CULP participants posited that cross-cultural competence is the key aspect of this tri-component cultural training model because "Leaders will inevitably encounter situations in another culture that do not meet their expectations. Even if provided with highly accurate region- or culture-specific information in training, leaders will not be able to anticipate every impact of cultural differences" [2, p.5]. Moreover, multiple, concurrent engagements executed in widely diverse areas of operation, coupled with limited pre-deployment training time, often preclude soldiers from fully mastering any single national, regional or local language or dialect. Therefore, honing linguistic skills is increasingly less important relative to training time for the development of inter- or intra-cultural competence.

Based on these assumptions, coupled with the analysis of the identified KSAs and training objectives previously noted, a cultural training structure was created along three conceptual categories: Culture General, Culture Applied, and Culture Specific. Culture General includes a meta-cognitive-based overview of topics which introduce the learner to broad definitions of Culture, American Culture, and the importance of Religion. Each of these three categories was further sub-divided into

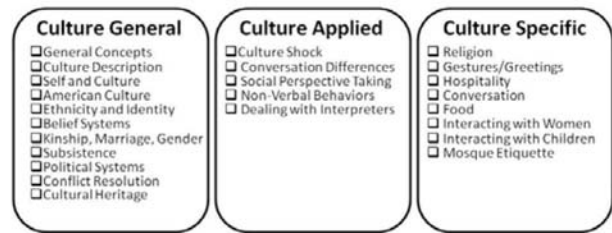specific content themes based on the nature of each, as shown in Figure 1

Figure 1. MEC Structure

This design supports the MEC instructional design as each theme can have a variable number of elements (i.e., assessments, tutorials, and scenarios), incrementally moves from most-general (CG) through to most-specific (CS), and is granular enough to enable users to digest each given theme as an individual lesson within small periods of training time.

2.4 Cultural Assessments
Assessments were created based on the culture-assimilator concept [10, 15, 17]. The Culture Assimilators developed for the MEC were modeled after a set of Arab culture assimilators previously developed through the Defense Language Institute [11]. The DLI Arab culture assimilator content is divided into five specific books and organized around clustered themes. This includes hospitality and conversion, religious practices and history, behaviors toward food and women, greetings and nuances of thought, and traditions versus progress. The five-book culture assimilator was a rich source of material, as it contains more than sixty hypothetical cross-cultural narratives within four-hundred and fifty pages. This corpus was analyzed to derive topical sub-dimensions across the five books.

2.5 Scenario Definition
Scenarios were developed along a number of themes in support of the training objectives identified above. These scenario themes are briefly describe below. Negotiation of cooperation between Arab tribal leader and U.S. Forces regarding

humanitarian support for infrastructure rebuilding: This type of multi-party negotiation is quite common and stresses cross-cultural differences that can often impede or negate positive interactions and outcomes. In Arab cultures, decision making is typically not a diffuse "meeting of the minds"; on the contrary, decision-making typically is vested in the hands of a tribal leader. The determinant of negotiation outcomes is largely a function of the cultural competency of the U.S. negotiation party.

Negotiation of cooperation between Arab tribal leader and U.S. Forces diffusing rioting in the city after U.S. forces mistakenly fired on a wedding party after celebratory gun-fire: An important cultural practice within many Arab cultures is that of celebration of key life events (e.g., funeral, wedding etc.) is that of weapon-firing. This can create tensions and misinterpretation of the intent of such celebrations as they can be perceived as threatening rather than non-threatening events.

U.S. military personnel training indigenous security and military personnel: U.S. forces are commonly tasked with training indigenous populations' security and military forces to increase their tactical and military capabilities for self-protection. This scenario entails a U.S. Lt to interact with and train a small ten-person ISF squad. These contexts typically place an emphasis on cross-cultural interactions involving a small minority of U.S. military personnel in contact with a larger majority of indigenous personnel. Clashes between cultures are common including cultural differences related to authoritarianism, forced-compliance, tolerance, and protection of honor in a group setting.

Entry into a Mosque through demonstration of proper Mosque etiquette: Given that the Mosque is a ubiquities part of Arab religious and cultural life, U.S. military personnel have a large degree of contact with, and need to enter, Mosques. U.S. military personnel are in the tenuous position of having to balance the somewhat conflicting goal of performing MOUT-based patrols, sweeps, and apprehension of persons of interest, with the need to show great deference for required etiquette and practices associated with entry to, and prescribed and proscribed behaviors within, a Mosque. This scenario involves a U.S. Lt needing to gain entry into a Mosque, demonstrate appropriate cultural understanding to remain there, and elicit intelligence information from the Mosque religious leadership regarding a person-of-interest.

Gathering of intelligence indicators through interaction with local populace: Garnering the support of the local populace is key in many military operations but particularly so for intelligence operations in support of Commanders Critical Information Recruitments (CCIR). Respecting conventional cultural norms is a key determinant of success in this context. This scenario involves a U.S. Lt interacting with members of the local populace in order to elicit intelligence indicators regarding the presence of unknown individuals within the city, observed loitering (possible surveillance) near key infrastructure, reports of threats or intimidations toward local business, and perceptions of corruption of local law-enforcement.

3.0 DISCUSSION

We approached the task of instantiating the MEC content described above with a hypothesis that skills involving person-to-person communication, cross-cultural dynamics, an understanding of non-verbal behavior and other broad types of interpersonal skills, require extensive experiential practice before they can be reliably and independently applied by the learner in a broad range of everyday situations. However, providing only experiential training via a game-based virtual environment encounter is likely to be just as limiting and ineffective as providing only lecture-based presentation of abstracted information on cultural sensitivity

or on a specific (sub-)culture. This is because the game-based experiences address only one aspect of the skill- and knowledge-development process – that of experiential application or practice. While essential to learning, practice must be combined with three other broad functions to achieve effective training. Specifically, practice must be:

- Supplemented with didactic instruction such as demonstration, lecture or presentation;
- Guided with scaffolding to provide coaching and/or feedback that directly or indirectly promotes deliberative learning and introspection; and
- Managed through formative (pre) and summative (post) assessments that guide the learner's progress toward the learning objectives.

The last function in the above list points to the explicit purposiveness of training, based on explicit learning objectives. In the MEC, the learning objectives drive not only the assessment process, but also the sequencing and management of the didactic instruction and practice process as well. The learner is systematically paced through a cyclic curriculum of instruction, practice, and assessment in a way that takes the trainee systematically toward the goal of achieving and demonstrating competence in the specific objectives of the training. Thus, the learning objectives strongly constrain the design of the practice environment (i.e., game—the two terms are used interchangeably here) to ensure that it provides clear opportunities for practice and assessment of the various actions and knowledge that the trainee must acquire. Whatever features are designed into the training game to make it interesting, challenging and engaging, they must be subordinated to these needs to provide opportunities to perform and be assessed in the performance of the target knowledge and actions. The impact of the objectives on the MEC practice environment flows into three separate aspects – the tutorials and

assessments outside the game, the scenarios and the dynamics in the game environment itself, and the behavior of the NPCs with whom the trainee interacts in the game. The curriculum is objectives-based and includes well integrated didactic instruction, game-based practice, coaching and feedback, and an ongoing assessment process that drive the cycle of learning forward on an individualized basis.

3.1 Instantiating Content

This Middle East curriculum was implemented using the game-based cultural training architecture VECTOR® [9] previously created by the research team because it provided a great degree of flexibility while requiring only incremental modifications. This includes a didactic learning component implemented using the commercial product Toolbook, a game-based practice component, a suite of authoring tools for creating and extending scenarios, and a data-base back-end for storing trainee performance results.

In the practice game, the trainee can progress through a series of scenarios, each of which involves interacting with a specific physical avatar or Non-Player Character (NPC) that possesses a specific set of cultural-behaviors and sensitivities. The interaction between the trainee and the NPCs in the scenario is organized into transactions, in which each party each say one thing in a turn-taking fashion. For the trainee, each turn is represented by a pre-defined set of utterances from which the trainee must select one. The progress through the scenario depends completely on the trainee's choices and the NPCs react differently on each path based on their pre-defined cultural sensitivities. Figure 2 shows an example of the dialog choices available to the trainee during an initial-meeting encounter with a specific NPC named "Nabil".

Figure 2. Interacting with an NPC

This fidelity required of the NPC avatars required a range of affect and expressivity using expressive non-verbal behaviors. The requirements for cross-cultural training dictated that the scenarios needed to include voice-acted speech coupled with avatars capable of a range of para-linguistic expressivity. Because these features were not required in the original creation of the VECTOR® system, this presented a technology gap. To address this gap we integrated a high-fidelity character-animation and lip-syncing tool, FaceFX [13], in order to provide highly interactive avatars capable of conveying subtle non-verbal cues. The use of FaceFX provides a smooth pipeline for processing voice-acted wav files against avatar speech (i.e., dialog) and produces character asset files which are then used to drive high-realistic game avatars.

During interactions with game NPCs, the trainee is expected to maintain trust with the avatars by communicating in ways that show deference for the modeled cultural norms and communication expectations. The NPC speaks via a voice-actor while the trainee selects responses via text presented on the screen. One of the forms of scaffolding and feedback provided dynamically to the trainee is a "trust meter" based on trainee responses (in the top left) and is an aggregate measure of NPC trust. Additional measures of performance are calculated and stored in the trainee

database for off-line use by an instructor or training administrator. Note that in the trust meter, Nabil's trust of the trainee is average. As the trainee interacts with any given NPC, the trust-meter will increment or decrement as a result of choices the trainee makes while interacting as a function of success relative to underlying training objectives. Additionally, an after-action review (AAR) summarizes all learning objectives measured against trainee performance, per scenario, as seen in Figure 3.

Figure 3. After Action Review Example

3.2   Scenario Authoring

Despite successes in applying simulation and serious-games to interpersonal skills training, scenario content generation remains an obstacle to the cost-effective use of the technology. In fact, a common criticism of game-based training has been the lack of a systematic approach to linking learning objectives to scenario content.

To this end, an important capability was the inclusion of an authoring facility. Such a facility would provide three advantages in that it:

- Allows for systematic and repeatable manipulation of existing scenario to support experimentation;
- Provides the ability for third-party end-users to add content based on changes in cultural conditions
- Positions scenario creation in the context of training-objective articulation, performance measurement, feedback and assessment.

624

The existing VECTOR® scenario editor, depicted in Figure 4, enabled the efficient creation of the Middle East game-based scenarios. The integration of instructional design principles into the authoring process promoted effective training scenario instantiation [4].

Figure 4. Scenario Editor

To facilitate consistent scenario creation, a workflow model for scenario authoring was included within the existing scenario authoring tool. The overall approach to make scenario authoring more accessible to a wider audience (i.e., beyond "game" engineers), was to use a cinematic metaphor to create the design of the authoring tool interface. The use of cinematic metaphors has been successfully used in similar virtual environments [19, 7]]. Scenario authoring within VECTOR® encompasses a number of training aspects, including:

- Training objective specification: Includes a library of training objectives which can also be expanded using the objective editor.
- Scenario information: This includes specifying high-level scenario information such as authorship tracking (critical when scenarios are created and maintained by multiple authors), target trainee population details, and other aspects of the overall scenario learning goals.

- Environment specification: Includes the designation of specific environment/setting within which a scenario will take place to support the identified training requirements.
- Plot organization: Involves the creation and arrangement of an overall scenario "story" which supports the selected training objectives and conveys a complete, coherent scenario to the trainee.
- Vignette creation: Encompasses the process of creating detailed dialog-based interactions and trainee response options, linking those interactions to training objectives, specifying feedback and coaching, and other measurement details.
- Scenario generation: Process for reviewing and validating the scenario before export to the game-engine "player".

3.3 Experimentation

As a continuation of on-going collaborations with the USMA at West point, a student-executed study using an earlier version of the VECTOR® system and Middle East content was conducted by Cadets at the USMA in 2008 [6]. The goal was to address the issue of how an interactive game-based training experience might influence the retention of training content in accordance with the "depth of processing" theory of Craik & Lockhart [8]. The findings of this field experiment regarding trainee reactions, knowledge acquisition and knowledge retention are presented in the three respective subsections below. All results reported are statistically significant at $p <$ .05, one-tailed.

A multidimensional training effectiveness framework was leveraged to guide the planning and conduct of this field experiment to evidence the effects of VECTOR® relative to another cultural and negotiations training system, known as ELECT BiLAT, on trainee reactions, learning and knowledge retention. A between-subjects design was used to

625

document the differential effects of the two training simulator systems on trainee knowledge of cultural negotiation techniques.

Trainee procedural knowledge of cultural negotiations was measured after trainees completed a given simulator-delivered scenario-based training (SBT) event set (viz., VECTOR® v. ELECT BiLAT). Trainee reactions, operationalized herein in terms of affective, utility, engagement and difficulty reactions, were also measured post-training. Cadet retention of procedural cultural knowledge was also measured post-training.

3.3.1 Setting and Procedure
Trainees (N = 28) participating in the study were Second Class Cadets enrolled in the PL300 Military Leadership course at the United States Military Academy at West Point. The cultural training simulations, and trainee assessment tests and measures, were administered in a classroom setting as part of a project led by two Cadets attending West Point. Trainee participants ranged in age from 20-years to 22-years.

Cadets attending the leadership course were selected on a convenience basis by the researchers and randomly assigned to complete a simulator-delivered SBT event set. More specifically, trainees read and signed an informed consent form and were randomly assigned to leverage either the VECTOR® or Elect BiLAT cultural training simulation system to complete cultural negotiations training. Upon completion of the simulator-delivered SBT event set, trainees completed a post-training test of cultural knowledge and training reaction measures. After finishing training, trainees also completed a measure of cultural knowledge to index knowledge retention.

3.3.2 Trainee Reactions
In terms of trainee reactions to the simulator-delivered cultural training experience, trainees completing the VECTOR® SBT event set reported being more engaged during training than those trainees exposed to ELECT BiLAT-based training. More specifically, relative to those Cadets exposed to ELECT BiLAT, those trainees exposed to VECTOR® reported more favorable engagement reactions to the simulator-delivered cultural training ($t(23) = -2.02$, $p \leq .05$). No statistically significant differences were observed between the two simulator system conditions in terms of trainee affective, utility and difficulty reactions to training.

3.3.3 Trainee Knowledge Acquisition
Trainees assigned to complete the VECTOR® SBT event set also acquired more procedural knowledge of cultural negotiation techniques than those trainees exposed to the ELECT BiLAT training system. More specifically, the results of an independent-samples t-test analysis suggest that relative to Cadets exposed to ELECT BiLAT, those trainees exposed to VECTOR® had greater knowledge of cultural negotiation techniques at the completion of the training session ($t(24) = -1.90$, $p \leq .05$).

3.3.4 Trainee Knowledge Retention
VECTOR® trainees also retained more knowledge of how to effectively negotiate in a multicultural context. The results of an independent-samples t-test analysis suggest trainees participating in VECTOR® delivered training maintained more procedural knowledge of cultural issues than those trainees subject to the ELECT BiLAT simulator system ($t(24) = -2.17$, $p \leq .05$). Moreover, a more fine-grained analysis of those items comprising the retention test which were designed to test trainee's culture-specific knowledge also evidenced a stronger impact of the VECTOR® system on Cadet learning. Relative to Cadets exposed to ELECT BiLAT, those Cadets exposed to VECTOR retained more culture-specific procedural knowledge ($t(24) = -2.64$, $p \leq .01$).

4.0 CONCLUSIONS

The shift in DoD focus from high intensity conflicts to the preparation for Stability, Security, Transition, and Reconstruction (SSTR) operations has led to the desire to increase the availability of cultural familiarization training for U.S. forces. One approach that has generated interest is the use of game-based solutions. In theory the use of game-based VEs should permit game-based cultural-training to be practiced by a greater audience.

The work described here investigates the ability of a game-based solution to effectively utilize a VE for an effective cultural training experience. We have concluded that four key elements are responsible for an effective application of this technology towards this goal. These four elements include:

1. *Scalability* allows for an application created for an initial small group training framework to be enlarged to a much greater 'N' of trainees without concomitant increase in cost.
2. *Extensibility* provides for addition of new types of virtual cases within previously-designed VE's, allowing new and different forms of norms and culture to be imparted and assessed.
3. *Evaluability* allows for the direct application of comparative-effectiveness metrics to a system. Such a feature prevents the simplistic aspect of some training programs' "show-it-and-trust-it" approach to any domain knowledge.
4. *Authorability* gives tools to non-technical domain experts, such as clinicians, permitting them to populate cases without having to supply code or otherwise contend with excessively technically-constrained requirements:

The VECTOR® system has the potential to reduce the cost of developing interpersonal skills-based training applications, allow such applications to be easily disseminated to a large numbers of trainees, and permit the applications to be executed on a wide variety of computer hardware. We expect that the Army and Marine Corps will obtain the greatest benefit from similar applications of cultural training because personnel in those services have the greatest need for direct interaction with the local populace in deployment areas. The VECTOR® platform will enable the development of training applications in other interpersonal skills domains and therefore will be a valuable training application for Government and private sector use.

5.0 REFERENCES

1. Abbe, A. (2008). Building Cultural Capability for Full-Spectrum Operations (SR 2008-04). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

2. Abbe, A., Gulick, L.M., & Herman, J.L. (2008). Cross-Cultural Competence in Army Leaders: A Conceptual and Empirical Foundation (SR 2008-01). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

3. Barba, C., Deaton, J., Santarelli T., Kerr, B. Singer, M., Belanich, J., (2006a). Virtual Environment Composable Training for Operational Readiness (Vector). Proceedings of the 25th Army Science Conference 2006 November 27 – 30, 2006, Orlando, FL.

4. Barba, C., Santarelli T., Glenn, F.A., Bogert, D. & Belanich, J. (2006b) An Approach to Scenario Authoring for Virtual Environment Training. In Proceedings of the 2006 Conference on Behavior Representation in Modeling and Simulation (BRIMS), May, 2006, Baltimore, MD.

5. Belanich, J., Sibley, D., & Orvis, K. L. (April, 2004). Instructional characteristics and motivational features of a PC-based game (ARI Research Report 1822). U.S. Army Research Institute for the Behavioral and Social Sciences: Alexandria, VA.

6. Carter, G., & Zwick, G.D. (2008). Negotiation Technology for the Evolving Soldier. Research Report for PL300 Leadership, March, 2008. United States Military Academy, West Point, New York.

7. Cavazza, M. & Charles, F. (2002). "Character-based interactive storytelling," ACM Joint Conference on Autonomous Agents and Multi-Agent Systems, Intary.

8. Craik, F. I. M., & Lockhart, R. S. (1972) Levels Of Processing: A Framework For Memory Research. Journal Of Verbal Learning And Verbal Behavior, 11, 671-684.

9. Deaton, J., Barba, C., Santarelli, T., Rosenzweig, L., Souders, V., McCollum, C. Seip, J, Kerr, B, & Singer, M. (2005) Virtual Environment Cultural Training for Operational Effectiveness (VECTOR) The Journal of Virtual Reality. 8(3), 156-157

10. Fiedler, F. E., Mitchell, T. R.,& Triandis, H. C. (1971). The culture assimilator: An approach to cross-cultural training. Journal of Applied Psychology, 55, 95-102.

11. Fiedler, F. (1966). Saudi Arabic Language And Culture Familiarization Course: Arab Cultural Assimilator, Books 1-5. Defense Language Institute, Monterey, Ca

12. Johnson, W.L., Beal, C., Fowles-Winkler, A., Lauper, U., Marsella, S., Narayanan, S., Papachristou, D., Vilhjalmsson, H. (2004). Intelligent Tutoring Systems 2004. Berlin: Springer-Verlag. 2004

13. King, S. (2008). Animating Speech in Games. Lect. Notes in Comp. Sci, 234-245, SpringerLink.

14. Kolb, D.A. (1984). Experiential learning: Experience as the source of learning and development. New Jersey: Prentice-Hall.

15. Lievens, F., Harris, M., Van Keer, E., Bisqueret, C. (2003). Predicting Cross-Cultural Training Performance: The Validity of Personality, Cognitive Ability, and Dimensions Measured By an Assessment Center and a Behavior Description Interview. Journal of Applied Psychology, Volume 88, Issue 3, June 2003, 476-489.

16. Raybourn, E. M. (2004). Designing Intercultural Agents for Multicultural Interactions. In Sabine Payr & Robert Trappl (Eds.), Agent Culture: Human-Agent Interaction in a Multicultural World, Lawrence Erlbaum, 267-285.

17. Rosenthal, D., Wadsworth, L.A., Paullin, C., Hooper, A., Mathew, J., Bhawuk, D. (2009). Navigating The Human Terrain: Development Of Cross-Cultural Perspective Taking Skills. Arlington, Va. U.S. Army Research Institute.

18. Ross, K. G. (2008, May). Toward an operational definition of cross-cultural competence from interview data. Patrick AFB, FL: Defense Equal Opportunity Management Institute (DEOMI).

19. Seif El-Nasr, M. (2005). Desktop 3-D Interactive Drama – Applying Design Principles from the Performance Arts, Human Computer Interaction International. NV, July 21-27.

20. Sheldon-Biddle, E., Stretton, M., & Burns, J.: (2003). PC Games: A Testbed for Human Behavior Representation Research and Evaluation. Proceedings of the 12th Computer Generated Forces and Behavior Representation Conference Scottsdale. Arizona, May 2003.

21. Stagl, K.C., Salas, E., Rosen, M.A., Priest, H.A., Burke, C.S., Goodwin, G.F., and Johnston, J.H. (in press). Distributed team performance: a multilevel review of distribution, demography, and decision-making. In F. Yammarino and F. Dansereau (Eds.,), Multi-Level Issues in Organizations, Amsterdam/New York: Elsevier.

6.0 ACKNOWLEDGMENTS

5.7   Technophiles to Newbies: The Challenge of Supporting Distributed Teams to Maintain Engagement in Virtual Worlds

Technophiles to Newbies: The Challenge of Supporting Distributed Teams to Maintain Engagement in Virtual Worlds

Karen Griffith
MYMIC LLC Educational Solutions Manager
karen.griffith@mymic.net

The purpose of this paper is to look for links in a virtual trainee's interest and self-efficacy in a simulated event as it relates to their previous self-reported technical skill level. Ultimately, the idea would be to provide the right amount of support at the right place at the right time to set the conditions for maximum transfer of the skill sets to the work place. An anecdotal recap of a recent experiment of a medium-scale training event produced in a virtual world will provide examples for discussion. In July 2010, a virtual training event was produced for the Air Force Research Lab's Games for Team Training (GaMeTT) at the Patriot Exercise at Volk Field in Wisconsin.   There were 29 EMEDS participants who completed the simulated OCO event using the OLIVE gaming engine. Approximately 25 avatars were present at any given time; including role players, observers, coordinators and participants.

1.0 INTRODUCTION

There is a growing number of high fidelity, well-developed, multi-player gaming engines available to organizations for the purpose of training individuals and geographically dispersed teams. These gaming engines provide a variety of options and capabilities, from which detailed simulated events can be conceptualized and executed virtually. This trend toward virtual world training opens up a plethora of options for organizations to accomplish learning objectives and design experiences that allow team functioning and practice in a safe environment.

Occasionally these objectives can become overwhelming to participants with little or no previous gaming experience. At best, individual, experienced gamers can interact in the virtual world with little interruption, but participant teams are seldom homogeneously technophiles. The task of supporting the accomplishment of training objectives executed in a virtual environment with participants of wildly varying technical skill sets can become a barrier to the achievement of the objectives.

With this idea in mind, participants of an experiment with a simulated event in a virtual world were recently asked questions related to their technical skill level, interest level in virtual training and self-efficacy.

The purpose of this paper is to report upon the responses and look for links in a virtual trainee's interest and self-efficacy in a simulated event as it relates to their previous self-reported technical skill level. Ultimately, the idea would be to provide the right amount of support at the right place at the right time to set the conditions for maximum transfer of the skill sets to the work place. An anecdotal recap of a recent experiment of a medium-scale training event produced in a virtual world will provide examples for discussion. .

2.0 BODY

The experiment took place at the close of the above mentioned Patriot Exercise on July 20th and 21st at building 533 at Volk Field near Tomah Washington.   This particular building is on the flight line of the field and consists of several large rooms connected by halls with smaller office spaces.   The rooms were equipped with ample seating, though minimal desk space

for computer set up. The building contained a SIPRINET network which was unavailable for this unclassified experiment. However, the sponsors of the event were able to secure 2 DSL connections which were temporarily routed to the building and became the uplink for the Ethernet network engineers stood up for the event.

Since wireless connections were not allowed, a router was connected to the DSL lines and Ethernet connections were established for each of 13 participant laptops used for participant and exercise support personnel. Two basic configurations were used during the experiment. The first configuration was a simple semi circle, with a trainer station in the center and a screen behind the trainer stand.
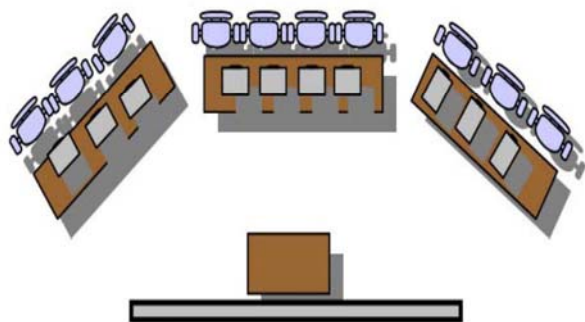
The second configuration was needed in order to simulate a geographical distance between players, as would be the case in ultimate use of the GaMeTT Training System. So, after some initial training was completed, participants were scattered throughout the facility in various rooms and offices. To make this happen, the network configuration above was broken down and moved to create various access points The new seating/network arrangement was designed to produce as much space as possible between participants, mimicking the way in which users would connect from

work or home. See the diagram below for this configuration:
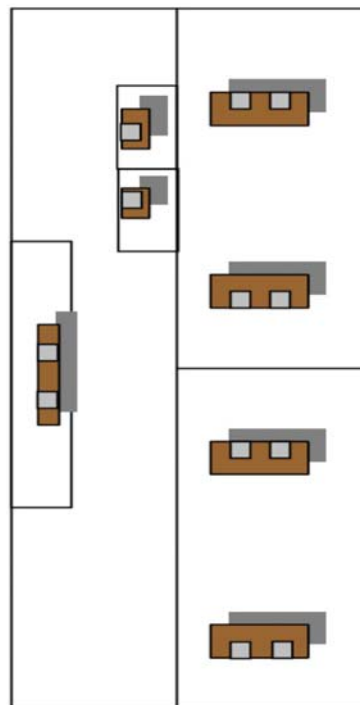
2.1 Details of the Event

There were three groups (Wednesday PM, Thursday AM and Thursday PM) brought in by van at the close of the second week of a two week live training event for forward medical teams. It is noteworthy that each of the 29 participants arrived in the building after almost two weeks of live exercises which continued 24 hours a day for 10 days. Each of the three groups was composed of individuals ready to play the following roles in the simulated event:

Nurse (2)
Doctor (1)
Administrative Officer (1)
Administrative Technician (1)
Medical Technician (3)

Each of these medical personnel was given a 30 page user guide with quick reference charts and a live one hour intensive

630

instructional session on the basics of operating their avatar and functionality of the virtual world. This training was conducted in the semi-circle shown above. Additionally, one support person for every 3 participants was available to answer questions.

Operation of the simulation required the movement of an avatar, movement of objects in-world and the use of a series of menus to access the medical treatment model for trauma cases that were a part of the simulation.

After the initial training event, the participants were separated into the second configuration in order to simulate a distributed environment. In this setup – a minimum 1 live support person was provided for every 4 participants.

During the execution of the event, it must be noted that the server housing the virtual environment went down twice and the DSL Internet connection feeding the experimental network service was interrupted a minimum of three times.

2.2 Evaluation Criteria

The overall evaluation methodology was based on Gagne's Nine Instructional Events. Gagne proposed that if learning content contained 9 significant elements, the optimum conditions for learning would be created for the transfer of learning from the training environment to the real world. The graphic below represents these 9 events (Gagne,1985).

This approach to the assessment and ultimately evaluation of the assessment results takes into account the fact that the transfer of training knowledge, skills and abilities can be difficult to achieve and measure in any setting. Furthermore, it acknowledges that the use an inquiry-based approach provides a holistic and iterative developmental process for solving the

multiple variables of creating a successful simulated event in a virtual world
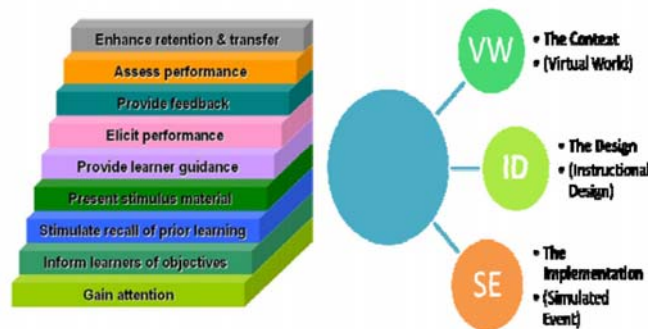
Gagne's hierarchical model builds from the lowest level (Gaining attention) and works its way up to "Enhanced retention and transfer" so that simulation architects can ensure that the optimum conditions for learning have been created. For example, it is important to build in environmental objects and training cues that get the learner's attention before attempting to share the learning objectives. Simulation architects, engineers, and instruction designers can use the methodology's guidelines and checklists to assist with prioritization and to mark the distinction between desired from required elements. Tools such as checklists and guidelines can help streamline concurrent development occurring in three or four related but distinct design fields: engineering, instructional, graphical and logistics. (Gagne, 1977)

The idea is to use a short cycle of feedback and assessment to influence the development/execution of the simulated event as opposed to an independent examination that is conducted only once at the close of the project. Virtual world development is too intricate and multi-variable to be reduced to a single snap shot in time that produces and yes or no answer.

Meeting the requirements of the Gagne's Nine Instructional Events can be more complex than it may seem because even though their effect is hierarchical, the

creation may be neither contiguous or chronological.

The survey responses included in this paper are part of this nine layer approach to evaluating simulated events implemented in virtual worlds. In the Pre-Event Survey, the participants were asked to self-report their technical skill level directly as well as in a series of questions designed to give a relative sense of their technical abilities in relationship to others. Participants were also asked to rate their interest level in the concept of conducting training in a virtual context.

In the Post-Event Survey, participants were asked if they felt they would apply their newly acquired skill sets on the job. Research shows that this concept of self-efficacy is a strong indicator to the transfer of skill sets from a training environment to the workplace.

Learner characteristics that support transfer are self-efficacy, pre-training motivation and perceived utility. (Bandura, 1994) In fact, a number of recent studies indicate that self-efficacy is the primary indicator of whether or not participants will experience increased performance once returning to the work place. Further, the complex task of measuring performance improvement on the job and attributing that performance to causal factors related to training events can be short circuited by simply asking the participant how useful they found the learning. (Grille, 2000) Similarly, asking learners about their reasons for participating in a learning experience (are they motivated to learn) and whether or not the material is applicable to their work (will they use the skill on the job) is a positive indicator that transfer will occur.

3.0 DISCUSSION
Among all of the before mentioned details, there are several caveats. The first is that if the experiment were extended to include participants who are dispersed geographically, all training and support would be done using the web and/or VoIP connection. Training could possibly be conducted via webinar, which still means participants would need to ready their equipment with the appropriate downloads and hardware to run the virtual environment.

This particular experiment was conducted live because the participants gathered for a live training exercise and were a "captive" audience on whom the tool could be tested.

Results of the participant Pre-Survey show that 64% reported themselves as technically proficient and 69% appear to be tech-savvy to their peers. This similar number indicates some agreement among the direct and less direct questions regarding technical skill level. Also, there were 5 individuals who rated themselves the lowest possible number on the technically proficient scale and 4 individuals who rating themselves the highest possible number.

When asked whether or not they were interested in learning more about virtual worlds, 13 individuals or 46% said they highly agreed, agreed, or were neutral. Likewise, 57% of the participants believed that virtual training can be effective for their team. Forty-six percent said they would be willing to train virtually when they return from the live exercise and 64% indicated they have high expectations for simulated events in virtual worlds.

On the Post-Event Survey, a consistent 64% said they would apply the skills learned in the simulated event to their work.

4.0 CONCLUSION(S)
These results are incredibly consistent and remarkably unremarkable considering the diversity of the group. The numbers may be interpreted to say that about the same number of highly tech-savvy individuals have high expectations and plan to apply their skills learned in the virtual world to the real world. These results span the Pre- and Post-Event surveys, therefore show little or no change in attitude.

There are two additional questions from the Post Event Survey that may reveal some interesting attitudes surrounding this simulated event in a virtual world.

When asked "During the training session, I was provided with enough support to be able to adequately use the technology", only 29% of the participants highly agreed or agreed with the statement. Further, when asked to agree with the statement "Someone was available to answer my questions about the virtual world used in this training session." a mere 5 individuals or 17% highly agreed or agreed.

These low numbers beg the question – if the simulated event was executed, and participants were able to learn skill sets that would be applied to the work place, how is it that the users did not feel supported? Another interesting question would be, what types of support would be needed if participants were geographically dispersed and there was no tech coach to stand over their shoulder?

In the end, the argument could be made that the low numbers for support indicate the correct amount was provided for an onsite event, since eventually the model calls for the skills to be learned either virtually or from a tutorial. Perhaps the low numbers were a good thing because if there had been too much hand holding, it would not have been replicable in a virtual environment.

In conclusion, providing the right amount of support for live and virtual events can be a complex at a minimum. Deploying a virtual environment for training can require providing a replicable model for support that addresses a number of skill levels and learning styles. Numbers can be deceiving; high agreement with a support question could mean that the level of support cannot be replicated in a distributed environment.

This delicate balance is a necessary one to achieve to avoid creating a barrier with the virtual technology rather than a tool.

5.0 REFERENCES

Bandura, A. (1994). Self-efficacy. In V.S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4pp. 71 – 81). New York: academic Press.

Gagne, Robert (1985). *The Conditions of Learning (4th Edition).* New York: Holt, Rinehart & Winston.

Grillie, Tony (2000). *Evaluating Training.* The Gillie Partnership LTD London, UK www.TheGilliePartnership.co.uk.

5.8 A Systematic Approach for Engagement Analysis under Multitasking Environments

A Systematic Approach for Engagement Analysis under Multitasking Environments

Guangfan Zhang, John Leddo, and Roger Xu
Intelligent Automation, Inc.
gzhang@i-a-i.com; jleddo@i-a-i.com; hgxu@i-a-i.com
Carl Richey and Tom Schnell
University of Iowa
richeyc@ccad.uiowa.edu; thomas-schnell@uiowa.edu;
Frederick McKenzie and Jiang Li
rdmckenz@odu.edu; jli@odu.edu

Abstract. An overload condition can lead to high stress for an operator and further cause substantial drops in performance. On the other extreme, in automated systems, an operator may become underloaded; in which case, it is difficult for the operator to mainta sustained attention. When an unexpected event occurs, either internal or external to the automated system, a disengaged operat may neglect, misunderstand, or respond slowly/inappropriately to the situation. In this paper, we discuss a systematic approach monitor for extremes of cognitive workload and engagement in multitasking environments. Inferences of cognitive workload ar engagement are based on subjective evaluations, objective performance measures, physiological signals, and task analysis result The systematic approach developed in this paper aggregates these types of information collected under the multitaskir environment and can provide a real-time assessment of engagement.

1.0 INTRODUCTION

Human operators play an important role in aviation and other safety critical missions. In existing aviation systems, the Operator Functional State (OFS) is usually not monitored and remediation is not implemented. In practice, two types of hazardous states of awareness are likely to lead to human errors [1]: a stress state due to high cognitive workload (we do not consider physical workload in this research) or a complacent/bored state in extremely low workload situations for a prolonged period of time [2]. It has been found in existing research that proper assessment of the cognitive workload and appropriate task mitigation in overload conditions offers potential to improve mission effectiveness and aviation safety [3]-[5].

On the other hand, the disengaged state developed in low workload conditions has not received equal attention. Disengagement is usually accompanied by poor situational awareness, which can lead to severe consequences in the multi-tasking

aviation domain. This is especially true in typical commercial flight scenario, whic has periods of high workload during pre flight preparations, takeoff and landing wit long periods of very low workload as th pilot cruises enroute toward the destinatio with the aircraft on autopilot. Pilots ca easily get disengaged during the enrout phase as they may be less attentive unde low workload. When unexpected event occur, the disengagement in the tasks bein performed could lead to operational error Such events could include unexpecte changes in weather (turbulence, fc example), equipment failure/malfunctio (such as hydraulic pump failure) or potenti collisions with other aircraft.

Therefore, the primary focus of thi research is to provide a real-tim engagement/disengagement assessmen mechanism. For this purpose, we will sta with a study of the relationship amon workload, engagement and performanc and identify the causes of low engagemen status (low workload) and its effect

634

(impaired performance). To better train an engagement assessment model, we will design a mechanism to identify the engagement ground truth based on different sources of information (performance measures, subjective evaluation, physiological signals, and task/workload analysis), followed by a committee machine-based real-time assessment model technique, and demonstrate the concept with a hypothesized dataset.

The remainder of the paper is organized as follows. Section 2 describes the relationship among workload, engagement, and performance. Section 3 presents a mechanism to determine the ground truth for engagement modeling. Section 4 describes an enhanced committee machine-based real-time engagement assessment model. Section 5 shows preliminary simulation results. Section 6 concludes the paper.

2.0 WORKLOAD, ENGAGEMENT, AND PERFORMANCE

The relationship between task engagement and performance has been an active area of research for over 20 years [6]. Researchers describe a state of high performance called "flow," which occurs when people are performing challenging tasks for which they have high skill. When a person's skill exceeds the task challenge, it is very likely he/she may become bored and disengaged. Performance can suffer if the imposed workload is greater than the resource that an operator can afford.

Workload and engagement are closely related. An optimal workload, one in which an operator performs challenging tasks within his or her abilities, leads to high levels of engagement and, accordingly, high levels of performance. If workload exceeds an operator's capacity, he/she will be overloaded and the performance will drop eventually. By the same token, if an operator is under a low workload condition for a prolonged period of time, he/she will usually drift into a disengaged state and the

performance will accordingly decrease after a certain amount of time, which will be the focus of this paper.

Therefore, performance can usually be affected by both workload and engagement. We define performance as a function of imposed workload ($WL_{imposed}$), workload capacity (WL_C), engagement (E), and efficiency (E_{ff}). These terms are defined as follows:

- *Imposed workload* is typically what is provided to the operator and consists of what objectives need to be met, a period of performance (e.g., a deadline or a length of time an activity must continue), criteria for success or quality (i.e., how the work may be evaluated), and other constraints that apply (e.g., what resources or people a person has available or whether a failure occurs in the system and how an operator is qualified).

- *Workload capacity* of an individual can change due to physical fitness (sleep loss, sickness, *etc.*) or training.

- *Engagement* is how much attention an operator puts in a task.

- *Efficiency* is usually determined by how efficiently he/she accomplishes a task.

The ratio between the imposed workload and the workload capacity basically determines whether an operator is either overloaded or underloaded:

$$\alpha = \frac{WL_{Imposed}}{WL_C} \qquad \text{(Equation 1)}$$

If α is beyond 1, the task requirements are greater than the person's processing capacity, and his or her workload is high; whereas if α is at or around 1, workload is appropriate for the worker. However, if α is well below 1 (for example < 0.5), workload is too low for the operator.

With the terms defined, performance can be derived based on the difference between

the imposed workload and the effective & engaged processing workload (WL_{EE}), which is determined by the capacity, engagement and efficiency:

$$WL_{EE} = WL_C * E * E_{ff} \qquad \text{(Equation 2)}$$

If the imposed workload $WL_{imposed}$ is greater than the effective and engaged processing workload WL_{EE}, the performance can be low due to the insufficient instantaneous processing power to meet the task requirements; otherwise, performance can be satisfactory.

Although performance can be affected by these four factors (capacity, engagement, efficiency, and imposed workload); in practice, real-time performance variation is not likely due to efficiency changes since experienced operators generally use efficient strategies. Also, there is a plethora of research on workload capacity (such as fatigue) and overload conditions. Therefore, in this research, we only focus on the study of engagement under low workload and task performance is used as an indicator of the engagement state.

3.0 ENGAGEMENT GROUND TRUTH

Before an engagement assessment model can be deployed, it needs to be trained based on the engagement ground truth and corresponding input information (physiological signals, performance, and others). However, there does not exist a sensor to provide engagement ground truth; instead, engagement ground truth is often derived based on all available information, including workload analysis, subjective evaluations, performance measures, and physiological measures. In each of the four types of information, different characteristics exist.

First, as discussed before, disengagement usually occurs under low workload conditions, and therefore, workload information shall be utilized to assess the engagement state. Cognitive workload analysis under multi-tasking environments can provide a direct and continuous

measure of the tasks being performed. We can hypothesize that an operator in a low workload condition for a prolonged period of time may become disengaged. It is important to note that the cognitive workload analysis is an objective measure of the task requirements and it cannot account for many other factors, such as individual variations and environmental conditions.

Second, subjective evaluation during missions is not suitable for identifying engagement ground truth (real-time assessment also), since in-mission subjective evaluation requires interaction with the operator being monitored, which would affect the operator's state artificially. Instead, a subjective evaluation after the mission can be directly utilized to assess the engagement state when he/she was performing the task. For example, recall of task/scenario events and a question of "did you feel you were engaged while doing the task?" can provide the information whether the operator was engaged during the task.

Third, performance measures can reflect the effects of individual characteristics and contextual information (including system setup, hardware/software issues, etc.) on engagement. Similar to in-mission subjective evaluation, intrusive performance measures, such as the fatigue-related Psychomotor Vigilance Test (PVT), a sustained attention task requiring subject response to an isolated target, is also not suitable for engagement assessment. However, non-intrusive performance measures, such as reaction time to Air Traffic Controller (ATC) communications, can be used as a good engagement indicator; slow response times, requests for clarification, and errors in readback could be associated with a disengaged state.

Finally, selected physiological measures can indicate when an operator is in a disengaged state. For example, a disengaged pilot during the enroute phase may have longer fixation durations and/or increased saccade length due to decreased

workload [7]. Other physiological measures include EEG readings, facial analysis, body posture, pressure readings from a pressure sensitive mouse or other equipment (to measure stress levels) and use of a wristband to measure stress as well. Previous research using facial analysis, posture, the mouse and the wristband shows that these physiological measures can correlate up to .78 with subject reports of engagement during a task (every five minutes [8]. Adding EEG could potentially strengthen this measure and by adding eye tracking, we can get extra evidence as to whether the pilot is actually attending to the unexpected event.

Based on the characteristics of these kinds of information, the ground truth finding procedure can be described as follows:

1) Analyze cognitive workload for the task(s) being performed:

   Outcome: $WL_{imposed}$; a continuous measure of the cognitive workload (0-100) induced by the tasks being performed.

2) Performance evaluation: we will derive a performance-based engagement score based on collected performance measures (continuous and/or discrete; for example, a relatively long reaction time to the ATC communications would probably indicate a disengaged state). If only discrete performance measures are available, interpolation can be used to derive the performance-based engagement scores in between.

   Outcome: a performance-based engagement score (0-100).

3) Fusion of cognitive task analysis results and performance-based engagement scores. Different fusion techniques can be adopted to combine the cognitive task analysis and performance evaluation results. A simple example is a set of fuzzy rules to fuse the imposed cognitive workload and the performance loss, such as

   a. If $WL_{imposed}$ is high, performance is high, the engagement score is high;

   b. If $WL_{imposed}$ is high, performance is low, the engagement score is medium; and

   c. If $WL_{imposed}$ is low for certain duration and performance-based engagement score is low, engagement is low.

   Outcome: E_O; a continuous objective measure of engagement (0~100)

4) Utilize critical physiological signals and features to indicate a disengaged state. Please note that this step only identifies potential disengaged state indicators during a mission, such as yawning and long fixation duration.

   Outcome: a discrete objective engagement score (E_{DO}, 0-100) representing how well an operator is engaged in a task based on critical disengaged physiological signs.

5) Analyze subjective evaluation results. There may be more than one subjective evaluation measures. In this case, we will first fuse these different discrete subjective evaluation results.

   Outcome: E_{DS}; a discrete subjective measure of the engagement (0~100)

6) Calibrate the continuous objective engagement (E_O) with the subjective engagement assessment ($E_S = \{E_{DO}, E_{DS}\}$). If at the same time instant, both E_{DO} and E_{DS} are available, they will first be combined before calibration (weighted sum, for example).

   Outcome: Engagement (E); a continuous overall measure (0~100)

$$E = E_o + weight * (E_s - E_o) * e^{-F*(t-t_0)}$$

(Equation 3)

where t_0 is the time instant, when the subjective engagement result is available; *weight (0~1)* is the confidence level of the subjective assessment result relative to the objective assessment result; and F is a forgetting factor that controls how long the subjective evaluation assessment result will impact the final outcome.

The goal is to use the subjective assessment to enhance the overall engagement assessment accuracy. In real operations, objective assessment results can be obtained frequently depending on the computer processing speed. On the other hand, subjective assessment results can only be available at longer time intervals because the assessment process is intrusive. In other words, the overall engagement mechanism is multi-variate. The basic idea behind Equation 3 is as follows. In the absence of subjective assessment results, we will solely rely on the objective results, which are more frequently available. When the subject assessment is available at t_0, we will use Equation 3 to calibrate the final engagement assessment result. Depending on our confidence level of subjective assessment, we will choose a proper weight or the weight can be trained with the training data. If weight is 0, $E = E_o$, which means that the subjective assessment of engagement is totally discounted and the engagement result fully relies on the objective measurements; On the other hand, if the weight at the other extreme of 1, $E = E_s$, meaning the engagement is determined solely by subjective assessment at the time the subjective assessment is introduced. Even with a confidence level defined (*weight*), to reduce the bias from the subjective assessment, we introduce an exponential term in Equation 3, with which the bias is exponentially discounted (controlled by the forgetting factor F).

With the above procedures, a continuous engagement measure can be derived considering different sources of information.

4.0 REAL-TIME ENGAGEMENT ASSESSMENT MODEL

The basic procedure for real-time OFS assessment is shown in Figure 1.
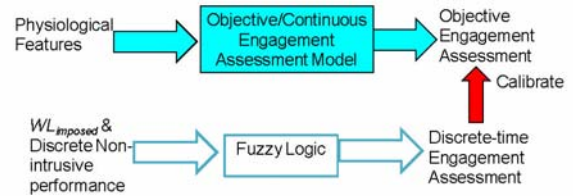
$$Engagement = E_o + weight * (E_s - E_o) * e^{-F*(t-t_0)}$$

Figure 1: Real-time OFS assessment procedure

It is similar to the engagement ground truth finding procedure described in Section 3. However, in real-time aviation applications, we cannot rely on manual selection of physiological features that can indicate a disengaged state, such as eye fixation duration and Heart Rate Variability (HRV). Instead, the physiological signals are being continuously monitored and the variation of engagement is automatically determined by a model trained with the physiological features and the identified engagement ground truth using a set of training data. The output of the real-time assessment model is an objective assessment of engagement.

Two sources of discrete information are utilized to "calibrate" the objective engagement assessment: the imposed workload (especially low workload for a prolonged period of time) and non-intrusive performance measures (mostly discrete in commercial aviation applications, such as reaction time and errors associated with ATC communications). Fuzzy rules similar to those used in ground truth finding can be applied to derive a discrete-time evaluation of engaged/disengaged state.

638

Again, the final engagement assessment is based on a calibration of the objective evaluation using the difference between objective and subjective evaluation results modulated by a forgetting factor.

It is worth noting that an enhanced committee machine method has been proposed by the authors in [9]. In this research, we will apply the same technique to build the objective/continuous engagement assessment model. The enhanced committee machine method is able to address large OFS individual variations by selecting the committee members and features that are the most sensitive to the OFS of each individual. The method has been successfully verified and validated with a driving test data set with a mean squared error of OFS estimation being significantly decreased (by around 20%) comparing to that without individualization [9].

5.0 SIMULATION RESULTS

In this paper, we generated a simulation dataset based on the flight information of AAL1238 on 05/12/2010, from Seattle to Chicago O'Hare, to illustrate the developed engagement assessment method. The basic flight information was extracted from the link from [10]. The altitude change along the flight is shown in Figure 2.
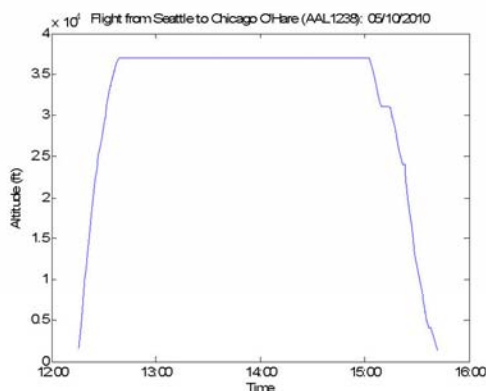

Figure 2: Flight AAL 1238 on 05/10/2010: altitude vs. time

Several assumptions were made to adopt this flight for the proof-of-concept purpose:

1) The imposed workload was constantly high during take-off and landing
2) The imposed workload was low during cruising (altitude above 30k ft).
3) Additional assumptions:
   a. Both an engaged pilot and a disengaged pilot reported disengaged from 1:45PM to 2:15PM.
   b. Reaction time of an engaged pilot is shorter; four reaction times are assumed: 1, 2, 2.4, 8, and 2 seconds vs. 3.2, 6.5, 8.3, 5.5 seconds for a disengaged pilot.
   c. Simulated fixation duration is used as a performance indicator for engagement assessment

Figure 3 shows the imposed workload, low workload more than 2 minutes, and subjective evaluation of disengagement.
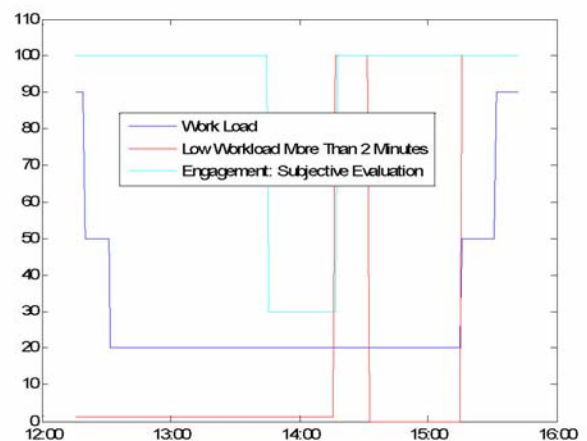

Figure 3: Workload, low workload more than 2 minutes, and subjective evaluation of disengagement

For an engaged pilot, the fixation duration is usually smaller than that of a disengaged pilot, who may be in a state of day dreaming or high fatigue. Also, the reaction time of the engaged pilot is usually shorter than that of a disengaged pilot. As an example, Figure 4 and Figure 5 show the physiological signals (normalized fixation duration) and the physiological indicators of a disengaged

pilot and an engaged pilot, respectively. It can be seen that a few more physiological indicators of disengagement are found for a disengaged pilot (shown in black).
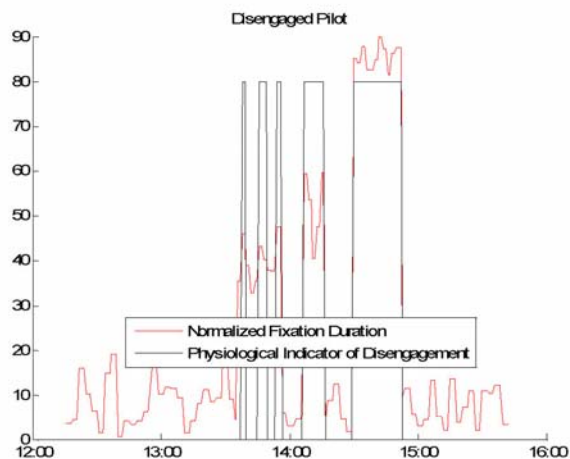

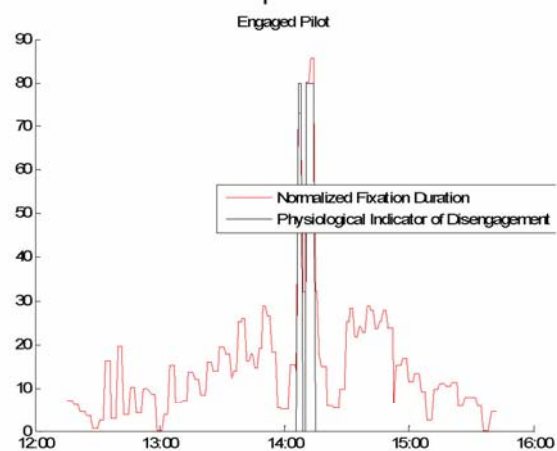Figure 4: Physiological signals and physiological indicators of a disengaged pilot

Figure 5: Physiological signals and physiological indicators of an engaged pilot

By combining with the performance indicators derived from reaction time, we derived the engagement score for an engaged pilot vs. a disengaged pilot with the method described in this paper. An example plot of the final engagement scores is shown in Figure 6.
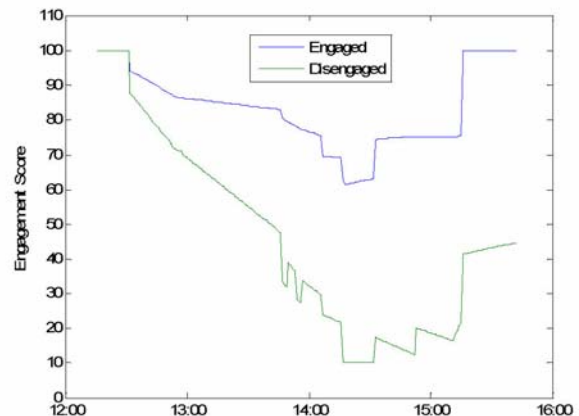

Figure 6: Disengagement score of an engaged pilot vs. a disengaged pilot

Clearly, from the example shown above, we can see that engagement/disengagement state assessment cannot solely rely on the subjective evaluation results, and low workload does not necessarily indicate a disengaged state. Physiological indicators of disengagement and selected non-intrusive performance measures, although discrete, can usually provide a better estimation of disengagement.

6.0 CONCLUSION

In this research, we have successfully developed a systematic approach for engagement assessment. The approach is based on a thorough understanding of the relationship among performance, workload, and engagement. To train a real-time engagement assessment model, we have developed a systematic approach to identify the engagement ground truth based on different sources of information: workload, non-intrusive performance measures, physiological indicators, and subjective evaluations. The ground truth identification approach was demonstrated using a simulation data derived from the AAL1238 flight on 05/12/2010.

One of the future tasks is to further implement the proposed real-time assessment technique on the enhanced committee machine-based model and is to verify and validate its performance with experimental data. Another important task is

640

to continue addressing the individual variation in the enhanced committee machine-based real-time engagement assessment model.

7.0 ACKNOWLEDGMENTS

8.0 REFERENCES

[1]. Pope, A. T.; and Bogart, E. H. (1993). Identification of Hazardous Awareness States in Monitoring Environments. SAE Technical Paper No. 921136, SAE 1992 Transactions: Journal of Aerospace, Section 1 - Volume 101, pp. 449-457.

[2]. Paul C. Schutte, et.al. "The Naturalistic Flight Deck System: An Integrated System Concept for Improved Single-Pilot Operations," NASA report, December 2007

[3]. Schmorrow D, Kruse AA. Improving human performance through advanced cognitive system technology. Interservice/ Industry Training, Simulation and Education Conference, Orlando, FL; 2002.

[4]. Schmorrow D, Stanney KM, Wilson GF, Young P. Augmented cognition in human-system interaction. In: Salvendy G, ed. Handbook of human factors and ergonomics, 3rd ed. New York: John Wiley; 2005.

[5]. Karl F. Van Orden; Erik Viirre; David A. Kobus, "Augmenting Task-Centered Design with Operator State Assessment Technologies," Foundations of Augmented Cognition: Lecture Notes in Computer Science, 2007, Volume 4565/2007, 212-219.

[6]. Csikszentmihalyi, M. and Csikszentmihalyi, I.S. (1988). Optimal experience: Psychological studies of flow in Consciousness. New York: Cambridge University Press.

[7]. Ellis, Kyle Kent Edward, "Eye tracking metrics for workload estimation in flight deck operations" (2009). Theses and Dissertations. Paper 288.

[8]. Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D. and Picard, R. (2009) 'Affect-aware tutors: recognising and responding to student affect', Int. J. Learning Technology, Vol. 4, Nos. 3/4, pp. 129–164.

[9]. Guangfan Zhang, Roger Xu, Wei Wang, Jiang Li, Tom Schnell, Mike Keller, Individualized Cognitive Modeling for Closed-Loop Task Mitigation, ModSim World Conference, 2009, Virginia Beach, VA

[10]. *http://flightaware.com/live/flight/AAL 1238/history/20100510/1610Z/KSEA/KO RD/tracklog*

MODSIM WORLD
Conference & Expo

A Systematic Approach for Engagement Analysis under Multitasking Environments

Guangfan Zhang, John Leddo and Roger Xu
(Intelligent Automation, Inc.)
Carl Richey and Tom Schnell (University of Iowa)
Frederick McKenzie and Jiang Li (Old Dominion University)

Outline

- Background

- Engagement Analysis

- Enhanced Committee Machine-based Engagement Assessment

- Simulation Results

- Conclusion & Acknowledgement

Background

- Definition of Engagement

  Attentional state of an operator during execution of a given task

- Two types of hazardous states of awareness that may lead to human errors:
  - A stress state due to high cognitive workload
  - A disengaged state due to low workload, poor physical fitness, etc.

- An operator in a low engagement level may neglect, misunderstand, or respond slowly/inappropriately to unexpected events
  - For example, commercial flight pilots can easily get disengaged during enroute phase under low workload

Definitions

- Imposed workload, $WL_{imposed}$
  - The workload that is assigned to successfully accomplish the given task.

- Workload capacity, WL_c
  - The maximum workload an operator can handle in the task

- Effective workload, WL_{EE}
  - The workload that the operator actually delivered toward the given task

Engagement and Performance

- $WL_{EE} = WL_C * E * E_{ff}$

  where E denotes the engagement (a score from 0 to 100), W_{EE} the effective workload, WL_C the workload capacity, and E_{ff} stands for *efficiency*, which is determined by how efficiently the operator accomplishes a task

- Performance:

$$perf = \begin{cases} 1 & if \quad WL_{EE} \geq WL_{imposed}, \quad succeed \\ WL_{EE}/WL_{imposed} & otherwise, \qquad failed \end{cases}$$

Note that

$$\frac{WL_{EE}}{WL_{imposed}} = \frac{WL_C \cdot E_{ff}}{WL_{imposed}} E$$
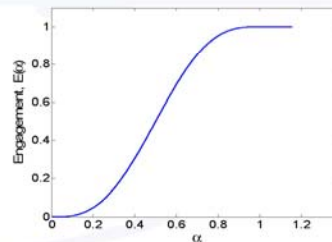
Engagement Analysis

- Engagement is determined by multiple factors

  - Task challenging level, $\quad \alpha = \dfrac{WL_{imposed}}{WL_C}$
  - Physical fitness, e.g., sleep loss, sickness
  - Environmental conditions

- We focus on the relationship between the engagement and imposed workload

$$E(\alpha) = f_B(\alpha)$$

α denotes the challenging lever, and B denotes other factors that affect engagement


644

Engagement Assessment

- Awareness evaluation
  - e.g., reaction time to the Air Traffic Controller (ATC) communications
  - Continuous and/or discrete

- Physiological signals and features
  - e.g., EEG readings, eye fixation durations, heart rate variability (HRV), etc.
  - Continuous

- Subjective evaluation
  - e.g., fatigue-related Psychomotor Vigilance Test (PVT)
  - Discrete
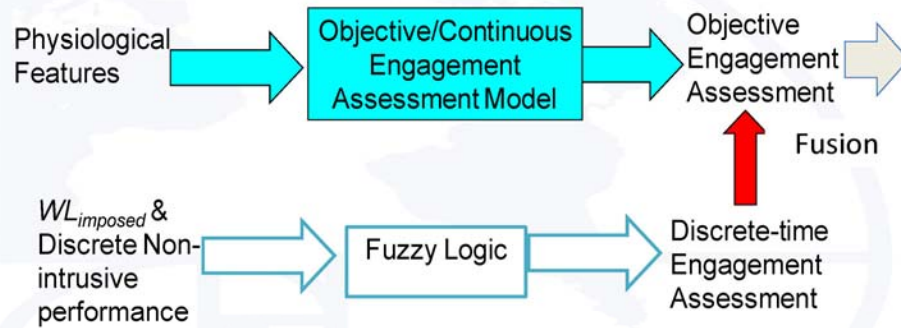
Overall Engagement Assessment

- Fusion of continuous objective engagement measure $E_o(t)$ with discrete subjective measure $E_s(t_i)$

$$E(t) = w_o E_o(t) + w_s g(E_s, t)$$
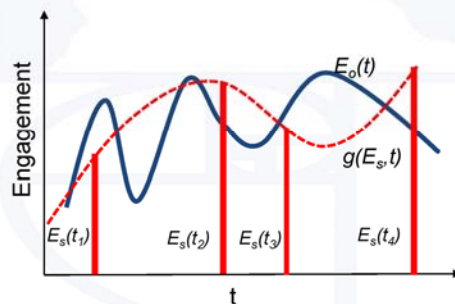
- $g(E_s, t)$ is a prediction function based on the previous discrete measures $E_s(t_i)$
- $g(E_s, t)$ can be estimated using a data-driven or parametric model method in the prediction theory

645

Real-time Engagement Assessment Model

Physiological Features → Objective/Continuous Engagement Assessment Model → Objective Engagement Assessment → Fusion

$WL_{imposed}$ & Discrete Non-intrusive performance → Fuzzy Logic → Discrete-time Engagement Assessment
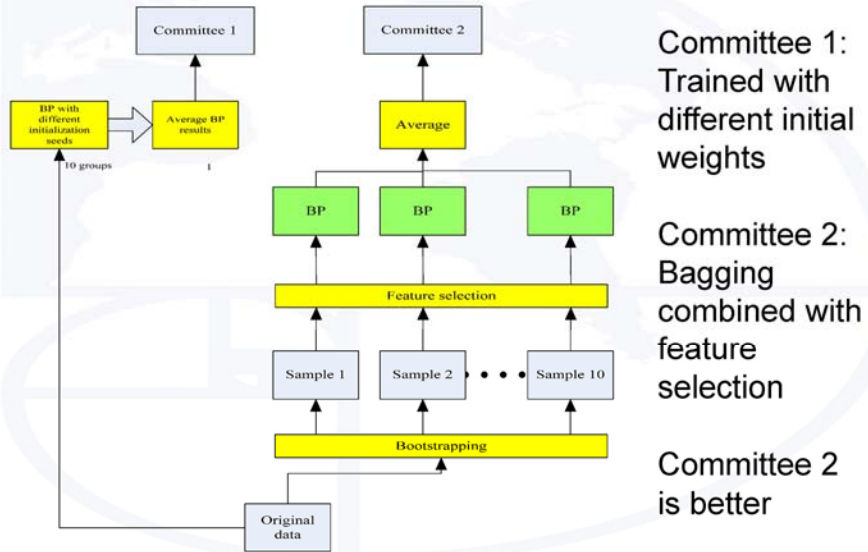
$$Engagement = w_o E_o(t) + w_s g(E_s, t)$$

Enhanced Committee Machine for Objective/ Continuous Engagement Assessment

Enhanced Committee Machine [1]

- Feature selection + Bootstrapping
- Advanced Feature Selection: Piecewise Linear Orthogonal Floating Search (PLOFS) [2]
  - Computationally Efficient
    - Performance: Wrapper type
    - Speed: Filter type
  - Select from Original Features
    - No transformation needed like PCA
  - Consider interactions among features
  - Generate a list of combinations
- Bootstrapping: resubmission

Enhanced Committee Machine Architecture

Committee 1:
Trained with
different initial
weights

Committee 2:
Bagging
combined with
feature
selection

Committee 2
is better

Simulations Results

AAL1238 on 05/12/2010, from Seattle to Chicago O'Hare [3]

Time	Latigude	Longitutde	Heading	Direction	KTS	MPH	Feet	Rate	Location
12:16PM	47.42	-122.31	178°	South	170	196	1,600		Seattle Center
12:17PM	47.35	-122.31	147°	Southeast	209	241	3,800	2,220	Seattle Center
12:18PM	47.31	-122.26	64°	Northeast	237	273	6,400	3,000	Seattle Center
12:19PM	47.34	-122.16	55°	Northeast	195	224	9,900	2,520	Seattle Center
12:20PM	47.38	-122.07	59°	Northeast	246	283	11,500	1,620	Seattle Center
12:21PM	47.43	-121.94	91°	East	324	373	13,200	1,860	Seattle Center
12:22PM	47.43	-121.81	92°	East	339	390	15,300	2,280	Seattle Center
12:23PM	47.43	-121.64	90°	East	357	411	17,800	2,340	Seattle Center
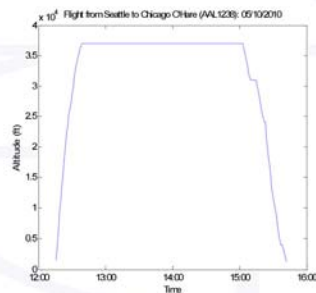
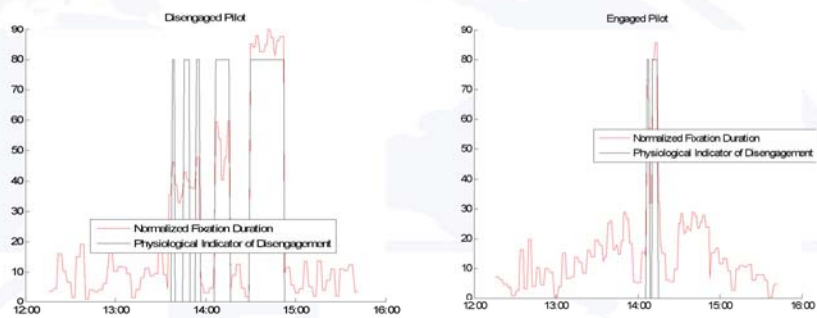Flight AAL 1238 on 5/12/2010
altitude vs. time

647

Simulation Results (cont')

•Simulated datasets
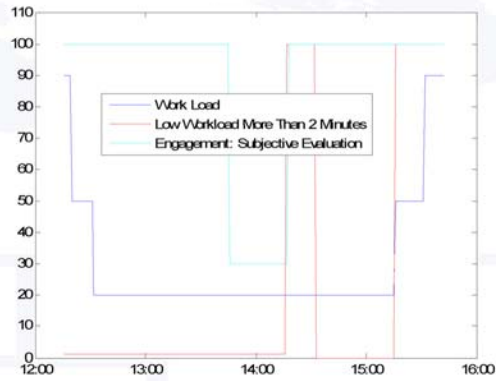
Workload, low workload more than 2 minutes,
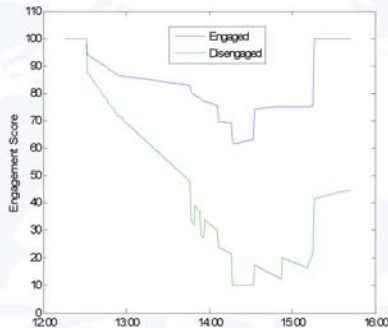and subjective evaluation of disengagement

Simulation Results (cont')

Normalized eye fixation duration
and physiological indicators
of a disengaged pilot

Normalized eye fixation duration
and physiological indicators
of an engaged pilot

Simulation Results (cont')

Engagement score of
an engaged pilot vs.
a disengaged pilot

- Engagement assessment cannot solely rely on the subjective evaluation
- Low workload does not necessarily indicate a disengaged state
- Physiological indicators and selected non-intrusive awareness measures usually provide a better estimation

Summary

- Established the relationship among performance, workload and engagement
- Developed a systematic approach for engagement assessment
- Demonstrated the feasibility of the proposed approach with simulations
- Future work:
  - Verify and validate the proposed system with experimental data
  - Address the individual variation in the enhanced committee machine-based real-time engagement assessment model

Acknowledgement

- Co-Authors:
  - Carl Richey & Tom Schnell (University of Iowa)
  - Frederick McKenzie & Jiang Li (Old Dominion University)
- Funded by the NASA
  - Contract No: NNX10CB27C
- NASA COTR: Dr. Alan T. Pope
- NASA PM: Dr. Kara Latorella

References

1. Guangfan Zhang, Roger Xu, Wei Wang, Jiang Li, Tom Schnell, Mike Keller, Individualized Cognitive Modeling for Closed-Loop Task Mitigation, ModSim World Conference, 2009, Virginia Beach, VA
2. Jiang Li et al., "Feature selection using a piecewise linear network", IEEE Transaction on Neural Network, vol. 17, no. 5, pp. 1101-1115, 2006.
3. http://flightaware.com/live/flight/AAL1238/history/20100510/1610Z/KSEA/KORD/tracklog

5.9   Modeling Pilot State in Next Generation Aircraft Alert Systems

Modeling Pilot State in Next Generation Aircraft Alert Systems

Alan S. Carlin, Amy L. Alexander, Nathan Schurr
Aptima, Inc.
acarlin@cs.umass.edu, aalexander@aptima.com, nschurr@aptima.com

Abstract. The Next Generation Air Transportation System will introduce new, advanced sensor technologies into the cockpit that must convey a large number of potentially complex alerts. Our work focuses on the challenges associated with prioritizing aircraft sensor alerts in a quick and efficient manner, essentially determining when and how to alert the pilot. This "alert decision" becomes very difficult in NextGen due to the following challenges: 1) the increasing number of potential hazards, 2) the uncertainty associated with the state of potential hazards as well as pilot state, and 3) the limited time to make safety-critical decisions. In this paper, we focus on pilot state and present a model for anticipating duration and quality of pilot behavior, for use in a larger system which issues aircraft alerts. We estimate pilot workload, which we model as being dependent on factors including mental effort, task demands, and task performance. We perform a mathematically rigorous analysis of the model and resulting alerting plans. We simulate the model in software and present simulated results with respect to manipulation of the pilot measures.

1.0   INTRODUCTION

The introduction of Next Generation Air Transportation System (NextGen) technologies into the cockpit is expected to dramatically increase the responsibilities of the pilot (JPDO, 2007). In particular, additional aircraft alerting systems will be introduced, and the pilot will need to adapt to the increase in both the number and types of possible hazard alerts. NextGen will also introduce additional automation technologies into the cockpit, capable of addressing alerts with minimal assistance from the human pilot. However, interfacing these technologies with both the human pilot as well as the large number of possible hazard alerts introduces a set of research challenges, including how to prioritize the alerts, how to plan the interaction between human and automation to address the prioritized hazards, and how to adjust the plan according to the state and capabilities of the pilot.

In order to address these challenges, Aptima, Inc., in cooperation with SAIC and under the supervision of NASA, is developing a NextGen aircraft system called ALARMS (ALerting And Reasoning Management System). The system has four parts: Bayesian reasoning to determine type and priority of existing hazards, a Time Dependent Markov Decision Process (TMDP)-based planner to address the hazards in a timely fashion, a human performance estimator to inform the planner as to the state and capabilities of the pilot, and an interface to inform the pilot of alerts in the best possible manner.

In this paper, we concern ourselves with the third item, how to estimate pilot state and capabilities in order to inform a plan for the human and automation to cooperate. Defining a plan to cooperate has been the subject of empirical research (Galster, 2003; Galster & Parasuraman, 2003; Parasuraman & Riley; Parasuraman, Sheridan, & Wickens, 2000). One approach is to describe a level of automation in a continuum between fully automated hazard response, to fully manual hazard response (Wickens, Mavor, Parasuraman, & McGee, 1998; Sheridan & Verplank, 1978). An extended method, proposed by Parasuraman et al. models human information processing in four stages: Sensory Processing, Perception/Working Memory, Decision Making, and Response Selection (Parasuraman, Sheridan, & Wickens, 2000). Differing circumstances may call for differing stages of automation.
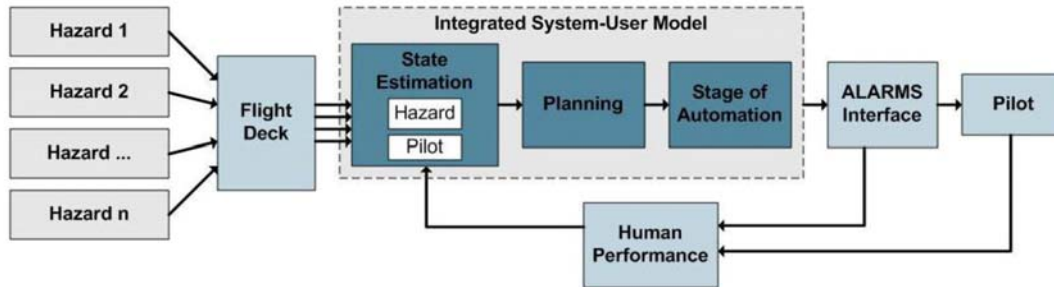
Figure 1: ALARMS System Diagram

But which stage of automation is appropriate may depend on several variables, including the characteristics of the hazards, as well as "pilot state," the ability of the pilot to perform under a given stage of automation. In this paper we introduce a model for estimating pilot state on the aircraft, for the purposes of informing the hazard alerting system. The model leverages existing literature on pilot workload as well as pilot performance. The model in this paper uses three stages of automation, each of which corresponds to a flight deck display. In Stage 1 of automation, increasing workload will greatly decrease quality and increase duration for high workload conditions as compared to low workload conditions. In Stage 2, increasing workload will decrease quality and increase duration. In Stage 3, the effects will be negligible.

The outline for the rest of the paper follows: First, we introduce the ALARMS system architecture and briefly outline its components, including a hazard state estimation module and a planning module for stages of automation. Next, we outline the pilot state estimation module, which estimates pilot workload. We show how the pilot state can be used as input for the ALARMS planning module. Finally, we show modeled results for how changes of pilot state will change temporal plans for stages of automation, and conclude with a summary and a discussion of future steps.

2.0 BODY: ALARMS SYSTEM ARCHITECTURE

The ALARMS system architecture is shown in Figure 1. Proceeding from left to right, multiple hazards exist in the environment and result in alerts on the flight deck. The hazards themselves, and the sensor alerting systems, are external to the ALARMS system. The alerts are issued to the ALARMS Integrated System User Model. The system alerts are treated as *evidence,* and from this evidence the ALARMS state estimation module estimates the actual hazards. A more detailed description of this estimate can be found in a companion paper to this work (9). To summarize, a Dynamic Bayesian Network (DBN) is used. DBN's have been found in similar systems, notably in medical diagnosis (Shwe & Cooper, 1991). Our use of a DBN in the ALARMS system is analogous, if the system alerts are treated as "symptoms" to estimate the "disease" of the actual hazard.

The hazard state is combined with the estimated state of the pilot (which we will return to in a moment), to form a complete state estimate. This pilot and hazard estimate is fed into a Planning module. The Planning module recommends the stage of automation for the hazards, which is fed into the ALARMS interface for display. The result is displayed to the pilot.

In this work, we focus on the Human Performance module in the diagram. This module estimates the status of the pilot,
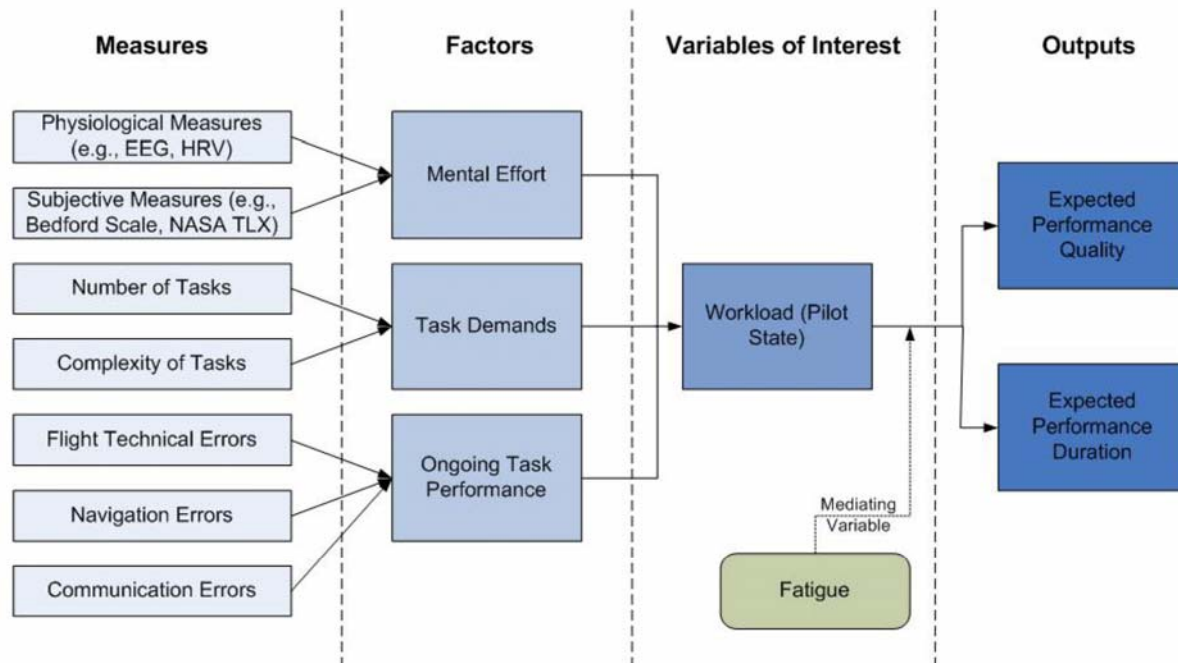
652

Figure 2: Human Performance Module

which in turn is used to estimate the performance of the pilot at each stage of automation. The planner will then select the appropriate stage of automation that maximizes the effectiveness of the combined pilot and automation.

2.1 Human Performance Module

We model human performance as shown in Figure 2. The Output from the module is an estimate of expected pilot performance, in terms of duration and quality of pilot handling of hazards. The Variable of Interest, **Workload**, is the key parameter representative of pilot state that changes over time and directly impacts performance. The Mediating Variable, **Fatigue**, influences the relationship. Other variables of interest (e.g., situation awareness) or mediating variables may be considered in future work.

In environments with task demands, workload affects the mental resources that a pilot can access to address the demands (Wickens & Hollands, 2000). Specifically, the effect can be modeled through a performance resource function, or PRF

(Norman & Bobrow, 1975). When cognitive resources are unavailable or unused for a task, performance will be diminished. As more resources are dedicated, performance will improve, until the task becomes limited by data and not resources. When multiple tasks must be accomplished, such as is the case when a pilot must supervise multiple systems in the cockpit, resource limitation becomes an issue (Kantowitz & Casper, 1988). The workload of the pilot will define the availability of a pilot's resources to handle alerts.

It is possible to assess workload as an index, and several criteria have been specified to compute the index (Wickens & Hollands, 2000; O'Donnell & Eggemeier, 1986). Among these criteria: a satisfactory workload index is *sensitive* to changes in task demands, *diagnoses* the cause of workload variation, is *selective* in that factors that do not affect workload are not included in the index, is *unobtrusive* in that the computation of the index does not affect workload itself, and is *reliable*.

For ALARMS, we identify three factors that predict workload: mental effort, task

653

demands, and ongoing task performance. We also identify relevant measures of these factors from the literature.

2.1.1 Mental Effort

We follow the literature by specifying mental effort as a contributing factor to workload. High levels of performance can be achieved under conditions of normal mental effort while extremely high mental effort situations tend to result in decreased performance. Measures of mental effort include both subjective and physiological measures (Veltman, 2001).

Subjective information in our model includes potential measures such as the NASA TLX scale (Hart & Staveland, 1988), which allows the operator to specify mental demand, physical demand, temporal demand, performance, effort, and frustration level. The Bedford Workload scale (Roscoe, 1984), on the other hand, is a decision tree, and the leaves of the tree provide a workload score on a single dimension.

Physiological information can also be obtained. Examples of potential measures include electroencephalography (EEG) or heart rate variability (HRV). It has been shown that heart rate can differentiate between phases of flight (which require different levels of mental effort) for pilots and co-pilots (Bonner & Wilson, 2002), even when subjective measurements do not.

2.1.2 Task Demands

In the prior subsection, mental effort is described as being necessary to accomplish tasks. The level of effort demanded will depend on the task. Simple tasks will require smaller amounts of resources, while complex tasks will require a higher degree of mental effort. Measures of Task Demands include both the complexity of tasks and the number of tasks.

Task complexity can affect workload; specifically, complex tasks will result in a higher workload. For example, the landing phase of flight produces higher workload than the enroute phase (Bonner & Wilson,

2002). As a second example, more automated tasks consume fewer resources than less automated ones (Schneider & Fisk, 1982).

Number of tasks affects workload as well, in two ways. First, the presence of additional tasks adds to workload. Second, there is a cost to switching among tasks (Rogers & Monsell, 1995). Thus, the contribution of tasks to workload exceeds the sum of the tasks complexities.

2.1.3 Ongoing Task Performance

Workload contributes to the model insofar as it is predictive of pilot performance. Thus, a well-accepted manner of estimating workload is to examine performance directly. Potential measurements include **Flight Technical Errors**, **Navigation Errors**, and **Communication Errors**. These errors can be measured by the ALARMS system at run-time.

2.2 Interface to ALARMS Planner

The goal of estimating pilot Workload is to predict the quality and duration of pilot actions so that a joint pilot/automation plan can be formed to address hazards. In this section, we describe the details of the interface to the planner. We begin by summarizing the planner itself, as introduced in (Carlin, Marecki, & Schurr, 2010) and adjusted in this paper to account for pilot state.

2.2.1 ALARMS TMDP Planner

We model the ability of the pilot and system to address hazards with a Time-Dependent Markov Decision Process (TMDP) (Boyan & Littman, 2000). A TMDP is a tuple $<S,A,P,D,R>$, where S is a set of states, A represents a set of actions, P is a transition matrix, D is a set of probability density functions, and R is a reward. Assume a finite set S of discrete states and a finite set A of actions. When the state s in S, and action action a in A is executed, the process transitions with probability $P(s,a,s')$ to state s' in S. The transition consumes t units of

time with probability *d(s,a,s',t)*, where *d(s,a,s',t)* is a probability density function over *t* for a given *s,a*, and *s'*. Similarly, the reward *R(s,a,s')* depends on *s*, *a*, and *s'*. Reward occurs when an action terminates. A deterministic TMDP policy is a mapping $S \times [0,\Delta] \rightarrow A$ where Δ (the deadline) is the earliest point in time after which all rewards are zero.

Let Ψ be a set of alert levels (e.g. "Nominal" (N), "Advisory" (A), "Caution" (C), "Warning" (W), or "Directive" (D)), Φ be an ordered set of hazards, and Ω be a set of autonomy stages. A TMDP problem in ALARMS is defined as follows:

- **States** A state *s* in *S* is a mapping from the hazards to their aleret levels. For example, given three hazards, state *s* = <*C,N,W*> defines that the first hazard is at Caution level, the second hazard is at Nominal level, and the third hazard is at a Warning level.
- **Actions** The actions of the ALARMS system represent the different ways in which the system displays the information about the hazards on the pilot's GUI. Each component of an action represents a stage of autonomy. For example, given three hazards, action *a* = <*1,3,2*> will present three hazards at stages 1,3, and 2 of autonomy. It is possible to represent different hazards at the same stage of autonomy (e.g. <*2,2,2*>).
- **Transitions:** ALARMS assumes that all hazards will eventually be addressed (their alert levels will return to N as a result of human or autonomy actions). An exception is when the action is not to address the hazard at any stage of automation (e.g. <*0*>), in which case the state remains the same for that hazard.
- **Durations:** ALARMS models action duration distributions by assuming that actions at differing stages of

automation take different durations. The specific durations of actions will be affected by pilot state, as we will specify in the next section.

- **Reward:** Reward is achieved for addressing the hazard and transitioning back to a nominal state. Each hazard can have a different reward associated with it. Reward will also depend on the pilot state, as we will see in the next section.

2.2.2 Pilot State in ALARMS TMDP

As shown in Figure 2, Workload affects the duration and quality of pilot actions in the ALARMS model. This is accomplished by performing a two step process. First, a workload score is computed from measurements of factors. This is accomplished through a linear weighting of the factors:

$$Workload = \alpha *ME + \beta *TD + \gamma *TP$$

where *ME* represents Mental Effort, *TD* represents Task Demands, and *TP* represents Task Performance. α, β, and γ represent linear weights that allow the prioritization of the factors to be varied.

In the second step, the workload score is used to modify the Duration and Reward function of the ALARMS TMDP. We use the Workload estimate to feed information into the Integrated User Module about the expected capabilities of the pilot, specifically the expected performance quality and the expected duration of pilot actions. The effect of Workload varies according to the stage of automation. In Stage 1 of automation, increasing workload in our model will greatly decrease quality and increase duration for high workload conditions as compared to low workload conditions. In Stage 2, increasing workload will decrease quality and increase duration. In Stage 3, we make the effects negligible.

The specific quantities attached to these terms "greatly decrease", etc, are
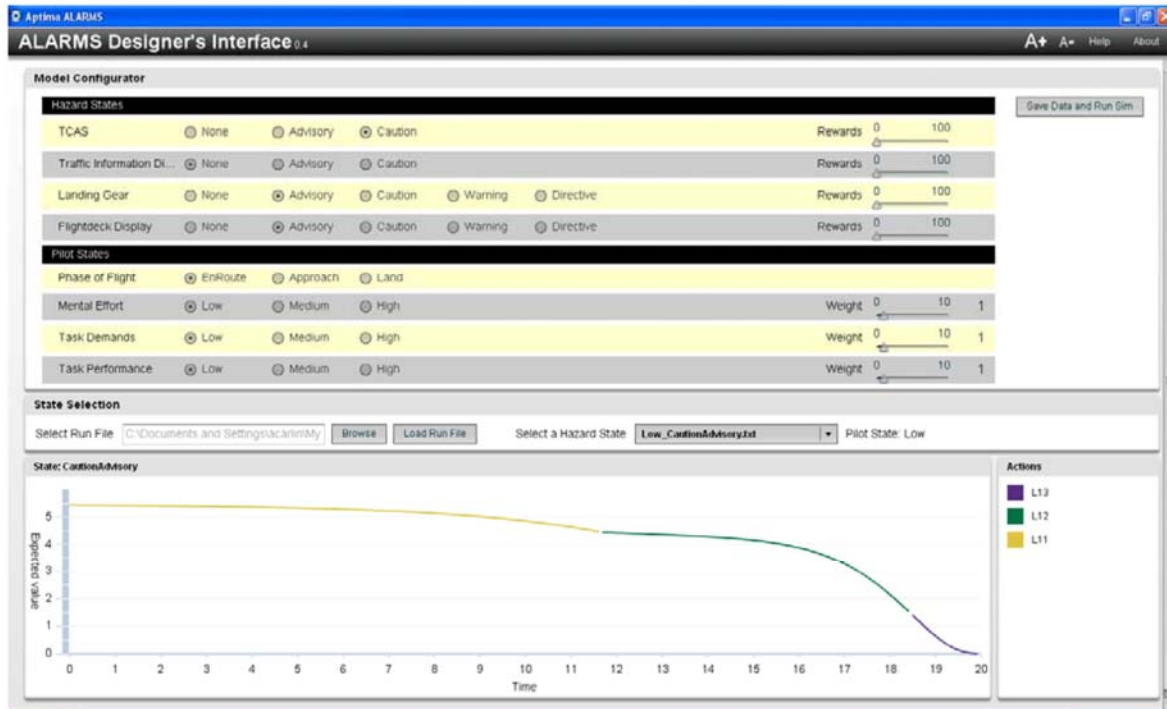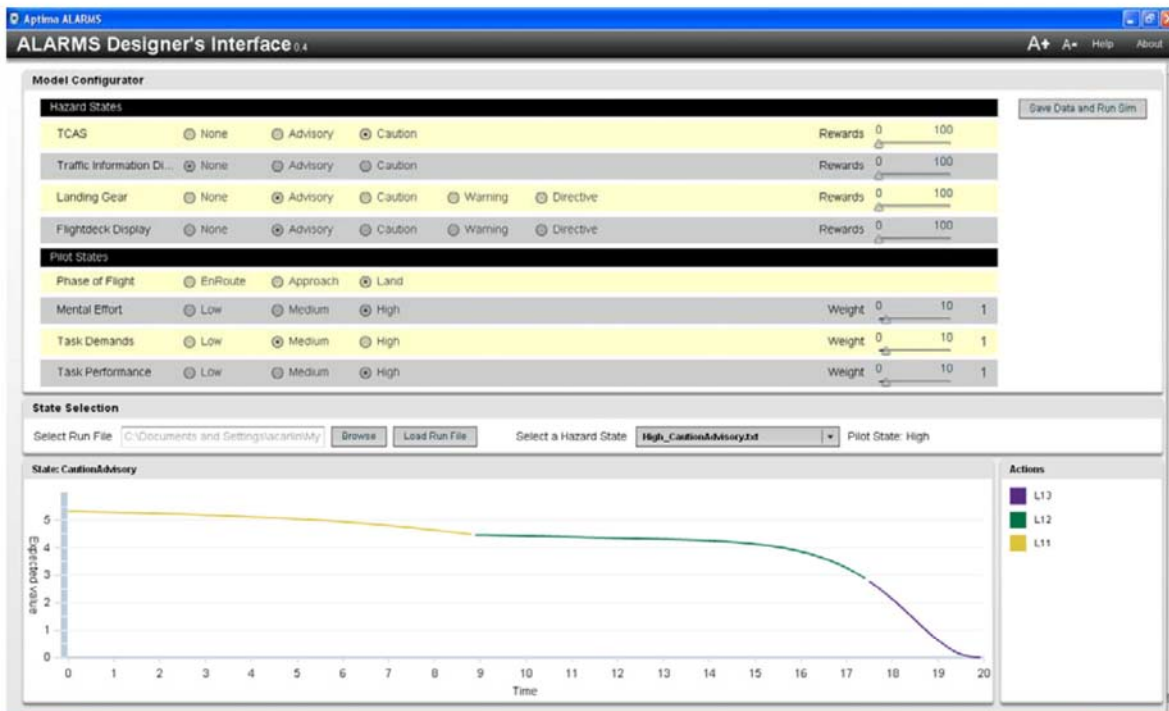
Figure 3: ALARMS System Designer Interface

Figure 4: High Workload Plan

parameters in our model. At present, we set quality and duration to halve and double, respectively, in Stage 1, when workload is changed from Low to High. Similarly we set quality and duration to decrease and increase 25% in Stage 2, and to decrease and increase 5% in Stage 3. Medium workload is currently simulated by

interpolating between the high and low workload conditions.

3.0 DISCUSSION/EVALUATION

In order to evaluate the effect of the factors on workload and on existing plans, an ALARMS System Designer Interface was developed. The interface is shown in Figure 3. On the top, multiple system alerts can be specified at various levels of alert. Directives are the highest priority of alert, followed by Warnings, Cautions, and Advisories. The entries "TCAS" and "Traffic information display" are indicative of a loss of separation hazard, and the entries "Landing Gear" and "Flight deck display" are indicative of a system failure hazard encountered while landing. Thus, we see in the figure that there is a caution-level alert for loss of separation, and a lower-priority advisory for a system failure hazard. Below the hazards, the factors affecting pilot state are specified, including Mental Effort, Task Demands, and Task Performance. Each factor is given a weight (corresponding to α, β, γ), in this case the weights are all 1.0. The figure shows that factors are all selected as "Low," and thus the pilot is under low workload conditions.

Below, we see a time-dependent plan for addressing the hazards, as computed by ALARMS. The x-axis is in time units. Without loss of generality, it is assumed that the tasks have a "deadline" at the 20 unit mark on the x-axis, and the plan works backwards from that mark. The y-axis shows the utility of the plan (on a relative scale to the ALARMS planning problem). As expected, utility decreases as the deadline approaches. The figure shows that ALARMS produces a 3-part plan for addressing the hazards under these conditions. Each part of the plan consists of the letter "L" followed by a stage of automation for each hazard, thus "L11" denotes that both hazards are handled at a stage of automation of 1, "L12" indicates

that the loss of separation hazard is handled at stage 1 and the system failure hazard is handled at stage 2. The figure shows that when there are more than 8 time units remaining, the hazards are both handled at Stage 1 of automation. As the deadline approaches, the recommended stage of automation transitions to "L12," that is, the lower priority hazard is handled at a higher stage of automation. Within 1.5 time units of the deadline, the stage of automation transitions to "L13" the lower priority hazard is handled at Stage 3 of automation.

Figure 4 shows a second plot. Here, the Phase of Flight has been changed to Land, Mental Effort and Task performance are labeled as indicating "High" workload conditions, and Task Demands are "Medium." This is a higher workload condition than the first example. As a result, the ALARMS planner is informed that low stages of automation will be less effective. The resulting plan at the bottom of the figure shows that higher stages of automation are selected at earlier points in time.

4.0 CONCLUSION

In this paper, we introduced a model designed to predict pilot performance in the cockpit, proposed to be implemented as a component of NextGen alerting systems. The larger ALARMS system design consists of Bayesian reasoning to determine type and priority of existing hazards, a Time Dependent Markov Decision Process-based planner to address the hazards in a timely fashion, a human performance estimator to inform the planner as to the state and capabilities of the pilot, and an interface to inform the pilot of alerts in the best possible manner. In this paper we focused on a model to contribute to the human performance estimator.

Key components of the model are that it estimates workload, it predicts the duration and quality of pilot performance, and it can be used to recommend what information will be displayed to the pilot, and what

information processing stage will be supported.

Future work consists of several directions. First, we will focus on the real-time nature of the measures, and how such measurements can be integrated into the cockpit in an unobtrusive manner. Second, we will embellish the model further. For example, the literature on task switching as well as issues related to attention (Yerkes & Dodson, 1908) can be added to the model.

5.0 BIBLIOGRAPHY

Bonner, M. A., & Wilson, G. F. (2002). Heart rate measures of flight test and evaluation. *International Journal of Aviation Psychology, 12,* 63-77.

Boyan, J., & Littman, M. (2000). Exact solutions to time dependent MDPs. *Proceedings of Neural Information Processing Systems*, (pp. 1026-1032).

Carlin, A., Marecki, J., & Schurr, N. (2010). ALARMS: Alerting and Reasoning Management System for Next Generation Aircraft Hazards. *The 26th Conference on Uncertainty in Artificial Intelligence.* Catalina Island, CA: UAI.

Galster, S. (2003). *An examination of complex human-machine system under multiple levels and stages of automation.*.(Technical Report No. AFRL-HE-WP-TR-2003-0149). WPAFB, OH.

Galster, S., & Parasuraman, R. (2003). The application of a qualitative model of human-intereaction with automation in a complex and dynamic combat flight task. *Proceedings of the 12th International Symposium on Aviation Psychology.* Dayton, OH: Wright State University.

Hart, S., & Staveland, L. (1988). Development of NASA-TLS (Task Load Index): Results of empirical and theoretical research. In P. H. (Eds.), *Human mental workload.* Amsterdam: North Holland.

JPDO. (2007). *Concept of Operations for the Next Generation Air Transportation System.* Joint Planning and Development Office.

Kantowitz, B., & Casper, P. A. (1988). Human workload in aviation. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 157-187). San Diego, CA: Academic Press.

Norman, D., & Bobrow, D. (1975). On data-limited and resource limited processing. *Journal of Cognitive Psychology*, 7, 44-60.

O'Donnell, R., & Eggemeier, F. (1986). Workload assessment methodology. In L. K. K. Boff (Ed.), *Handbook of perception and performance (vol 2).* New York: Wiley.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39, 230-253.

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics*, 30, 286-297.

Pritchett, A. R., Vándor, B., & Edwards, K. (2002). Testing and implementing cockpit alerting systems. *Reliability Engineering and System Safety*, 75, 193-206.

Rogers, R., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, 207-231.

Roscoe, A. H. (1984). Assessing pilot workload in flight. *Conference Proceedings No. 373. Flight Test Techniques.* Paris: AGARD.

Schneider, W., & Fisk, A. (1982). Concurrent automatic and controlled visual sarch: Can processing occur without cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 261-278.

Shwe, M., & Cooper, G. (1991). An empirical analysis of likelihood-weighting simulation on a large, multiply connected medical belief network. *Computers and Biomedical Research*, 5, 453-475.

Sheridan, T., & Verplank, W. (1978). *Human and Computer Control of Undersea Teleoperators.* Cambridge, MA: MIT Man-Machine Systems Laboratory.

Veltman, H. (2001). *Mental Workload: Lessons learned from subjective and physiological measures..*

Wickens, C. D., & Hollands, J. (2000). *Engineering Psychology and Human Performance.* Upper Saddle River, NJ: Prentice Hall, Inc.

Wickens, C., Mavor, A., Parasuraman, R., & McGee, J. (1998). *The Future of Air Traffic Control: Human Operators and Automation.* Washington D.C.: National Academy Press.

Yerkes, R., & Dodson, J. (1908). The relation and strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology* , 459-482.

6.0 ACKNOWLEDGEMENTS

659

Modeling Pilot State in Next Generation Aircraft Alert Systems

Alan S. Carlin, Amy L. Alexander, Nathan Schurr

APTIMA®
HUMAN-CENTERED ENGINEERING

NASA   SAIC   IBM

NextGen Aircraft Alerting Systems

APTIMA HUMAN-CENTERED ENGINEERING

- Future flight deck systems will require sophisticated **information management**

Integrated Intelligent Flight Deck Technologies
AVIATION SAFETY PROGRAM

- Need integrated alerting and notification (IAN) function to:
  – Continuously monitor info from various sources to evaluate hazart potential

  – Consider immediate hazards (current) and situations requiring re-planning or coordination (future)

  – Provide caution/alert/warning automated notifications

  – Provide context-relevant decision support to pilot and other aircraft automation functions

- Interdisciplinary effort

Levels of Automation*

- 10 – Computer Decides everything, acts autonomously
- 9 – Informs human if it decides to
- 8 – Informs human only if asked
- 7 – Executes then informs
- …
- 4 – Suggests one alternative
- 3 – Narrows selection to a few
- 2 – Offers complete set of alternatives
- 1 – No computer assistance

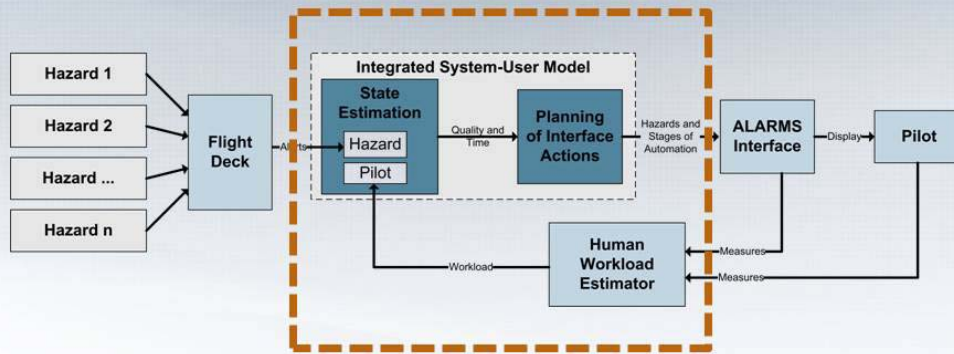*Wickens 1998, based on Sheridan and Verplank 1978

3

Stages of automation

- Example 4 stage model
  - Sensory Processing
  - Perception/Working Memory
  - Decision Making
  - Response Selection

- Four classes of functions
  - Information acquisition
  - Information analysis
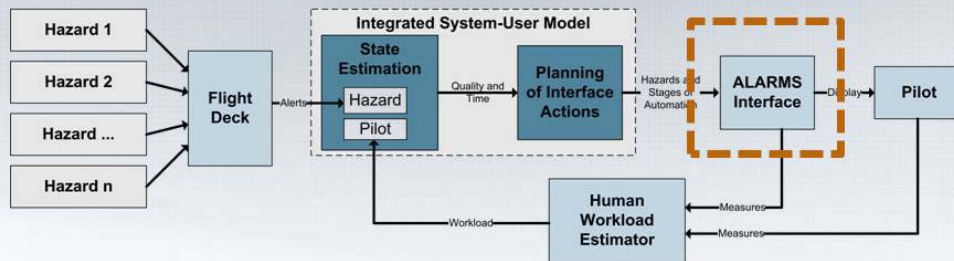 - Decision and action selection
  - Action implementation

4

Integrated User Model

Integrated User Model
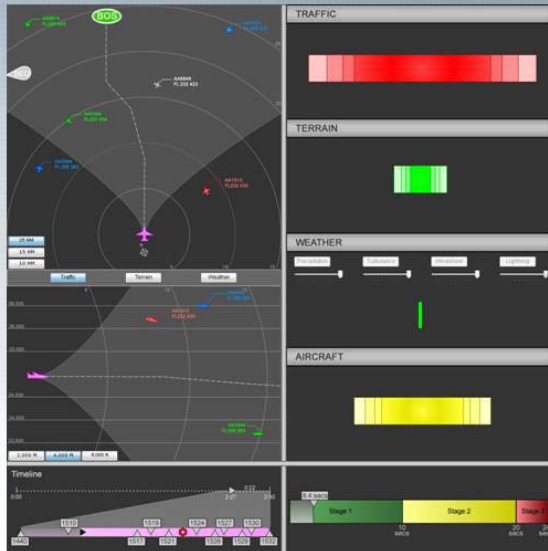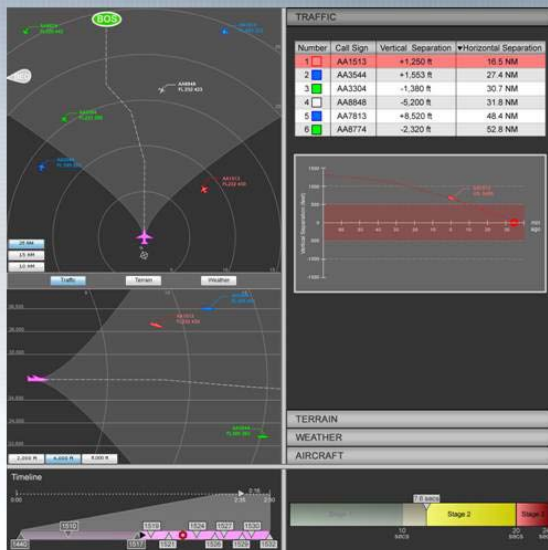
Mock-ups designed by Andy Chang M.S. and Dr. Amy Alexander

Mock-ups designed by Andy Chang M.S. and Dr. Amy Alexander

Stage 3 Automation
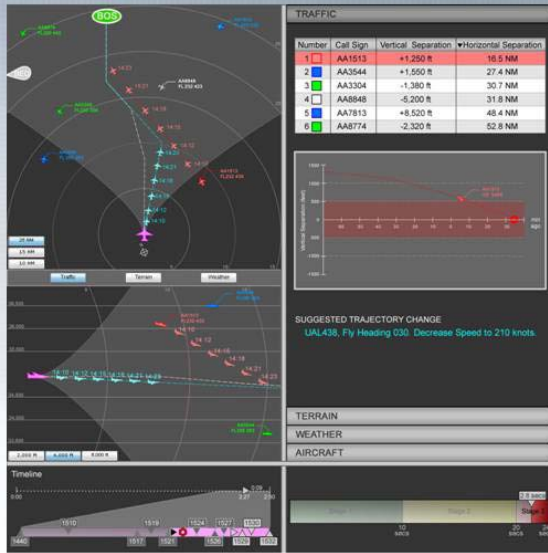
Decision Support

Mock-ups designed by Andy Chang M.S. and Dr. Amy Alexander

Integrated User Model

Pilot State Refinement

Measures / Factors / Variables of Interest / Outputs
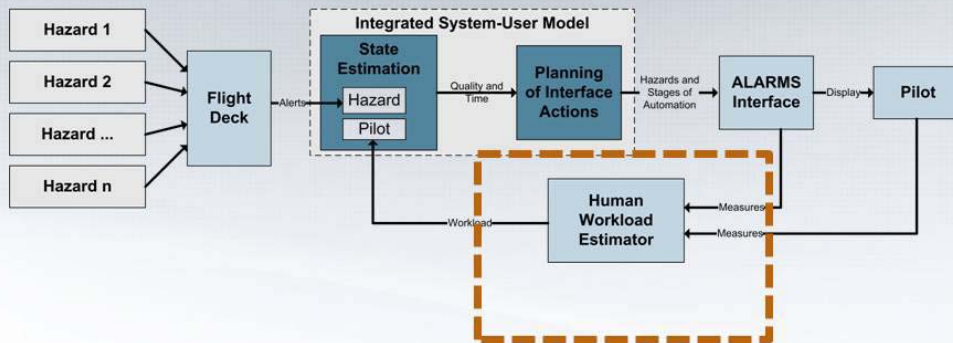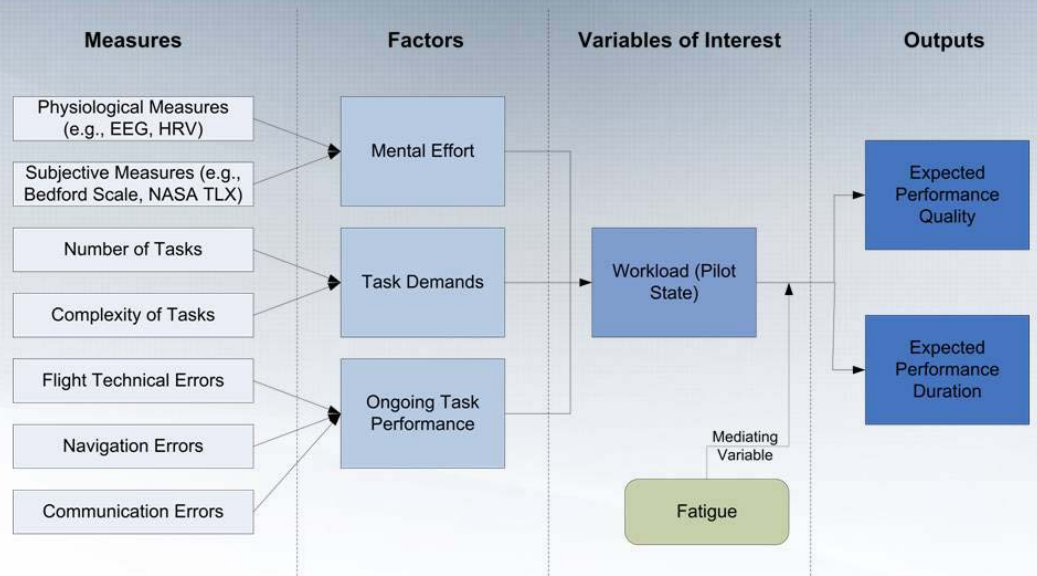
Workload

- Workload
  - Performance Response Function (PRF)
    - Finite amount of resource able to be allocated to problem
    - Time Sharing
    - Automaticity
  - Satisfactory workload Index has these properties:
    - Diagnoses cause of variation
    - Selective of factors
    - Unobtrusive
    - Reliable

Factors

- **Mental Effort**
  - Subjective Information
    - NASA TLX scale
    - Bedford Workload Scale
  - Physiological Information (EEG, HRV)
- **Task Demands**
  - Task Complexity
  - Number of Tasks
- **Ongoing Task Performance**

Pilot State Refinement

- **Mental Effort, Task Demands, and Ongoing Task Performance are used to compute Workload**
  - We model these as Low/Medium/High, take linear combination
  - Workload = $\alpha * ME + \beta * TD + \gamma * TP$
  - Workload is combined with information about hazard state and stage of automation to determine the quality and duration of action.
- **Increasing Pilot Workload leads to:**
  - Greatly decreasing quality, increasing duration in Stage 1
  - Decreasing quality, increasing duration in Stage 2
  - Very small effects in Stage 3
- ***Thus, increasing Pilot Workload will tend to increase the stage of automation.***

666

Integrated User Model

15
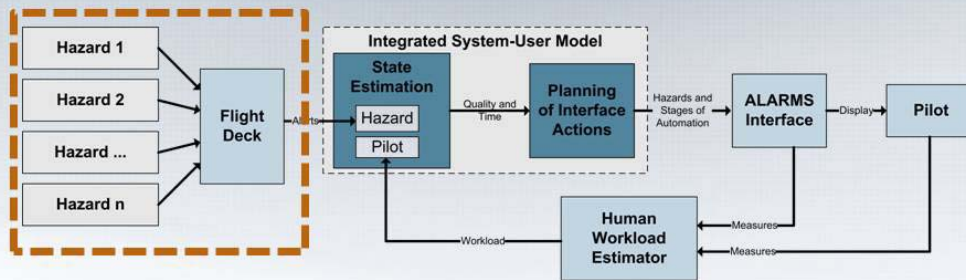
Identify CWA (Caution/Warning/Alerting) Applications For NextGen and Develop a System Architecture

- **Design Approach –**
  - Define Alerting Levels

23 Existing Systems Reviewed
13 NextGen Planned Systems Reviewed

Alert	Timeframe
Directive	<10 seconds
Warning	10--15 seconds
Caution	< 40 seconds
Advisory	non-critical

  - Define Hazard List
    - System Failure
    - System Performance Compromised
    - Loss of Separation
    - Adverse Weather Encounter
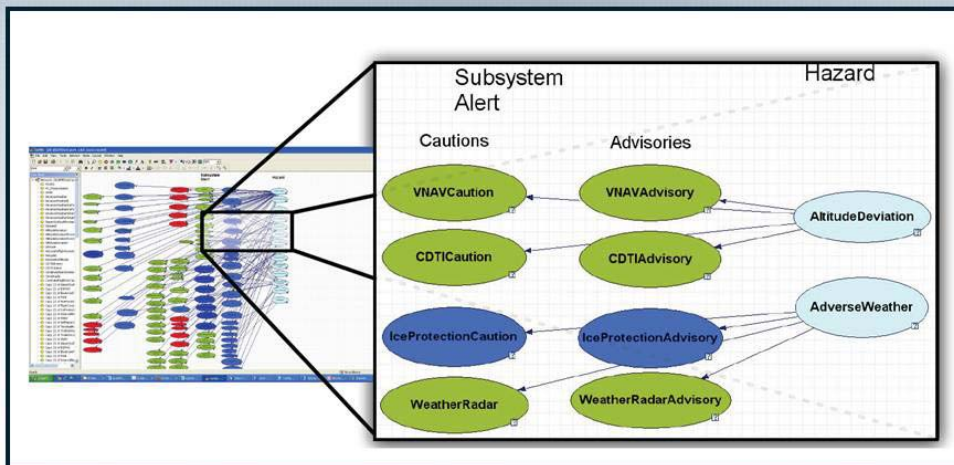   - Altitude Deviation
    - Navigation Deviation

- Controlled Flight Into Terrain
- Crew Incapacitation
- Flight Performance Compromised
- Structural Failure
- Life Support
- Protected Airspace Incursion
- Loss of Communication
- Runway Incursion

16

APTIMA
HUMAN-CENTERED
ENGINEERING

Tools/Technology/Systems	Sub-Systems	System Failure	System Performance Compromised	Loss of Separation	Adverse Weather Encounter	Altitude Deviation
Electrical (3.1.11)		(C)	(W)			
Hydraulic (3.1.15)		(C)	(C)			
Fuel (3.1.14)		(C)	(C)			
Landing Gear (3.1.17)		(W)	(C)			
Air Conditioning/Pressurization (3.1.21/22)		(W)	(C)			
Ice Protection (3.1.16)		(W)	(W)		(C)	
Fire Protection (3.1.12)		(C)	(C)			
Enhanced Ground Proximity Warning (EGPWS) (3.1.20)		(A)	(A)		(W)	
Navigation Radio (Nav Radio)						
	Enroute (3.1.1)	(A)	(A)			
	Approach (3.1.1)	(A)	(A)			(A)
Flight Management System (FMS)						
	RNAV (3.1.18)	(A)	(A)			

A=Advisory, C=Caution, W=Warning
Blue=Aviation, Green=Navigation, Purple=Communication

©2010, Aptima, Inc.   17

ALARMS Bayesian Network

APTIMA
HUMAN-CENTERED
ENGINEERING

Subsystem Alert   Hazard

Cautions   Advisories

VNAVCaution   VNAVAdvisory   AltitudeDeviation

CDTICaution   CDTIAdvisory

IceProtectionCaution   IceProtectionAdvisory   AdverseWeather

WeatherRadar   WeatherRadarAdvisory

©2010, Aptima, Inc.   18

668

Integrated User Model
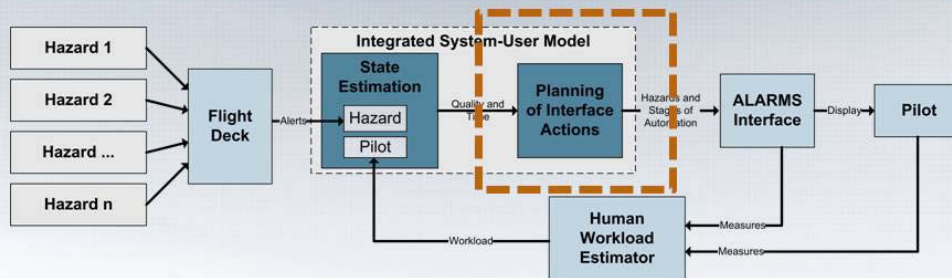
TMDP model

- **TMDP model** (Boyan and Littman 2000)
  - $\langle S,A,P,d,R \rangle$
  - Set of States: Mapping from hazards to alert levels
    - Example: $\langle N,A,W \rangle$
  - Set of Actions: Stage of Automation for each hazard
    - Stage 1 = human-centric, Stage 3 = highly automated
    - Example $\langle 1,3,2 \rangle$
  - Probabilistic Transitions: $P(s,a,s')$
  - Action Durations: Probability Density Function $d(s,a,s',t)$
    - Deadline Δ after which there is no reward
    - More automated actions: Quicker
  - Reward for addressing hazard and returning to nominal state
    - More automated actions: Lower reward
  - Policy: Mapping from state and time to action

Pilot State Refinement

- Mental Effort, Task Demands, and Ongoing Task Performance are used to compute Workload
  - We model these as Low/Medium/High, take linear combination
  - Workload = $\alpha*ME + \beta*TD + \gamma*TP$
  - Workload is combined with information about hazard state and stage of automation to determine the quality and duration of action.
- Increasing Pilot Workload leads to:
  - Greatly decreasing quality, increasing duration in Stage 1
  - Decreasing quality, increasing duration in Stage 2
  - Very small effects in Stage 3
- *Thus, increasing Pilot Workload will tend to increase the stage of automation.*

CPH Planner (Marecki et al. 2007)

- Define values V(s)(t)
- Expected Reward (for Bellman Backup)

$$\arg \max_{a \in A(s)} \left(\sum_{s' \in S} P(s'|s,a) \int_0^t p_{s,a}(t')(R(s,a,s') + V^*(s')(t-t'))dt' \right)$$

- Since continuous, need approximate value functions
  - Convolution becomes intractable
  - Solution: phase-type distributions
  - Convert to MDP with uniform action durations

$$\int_0^t p(t') V(s')(t-t') \, dt'$$
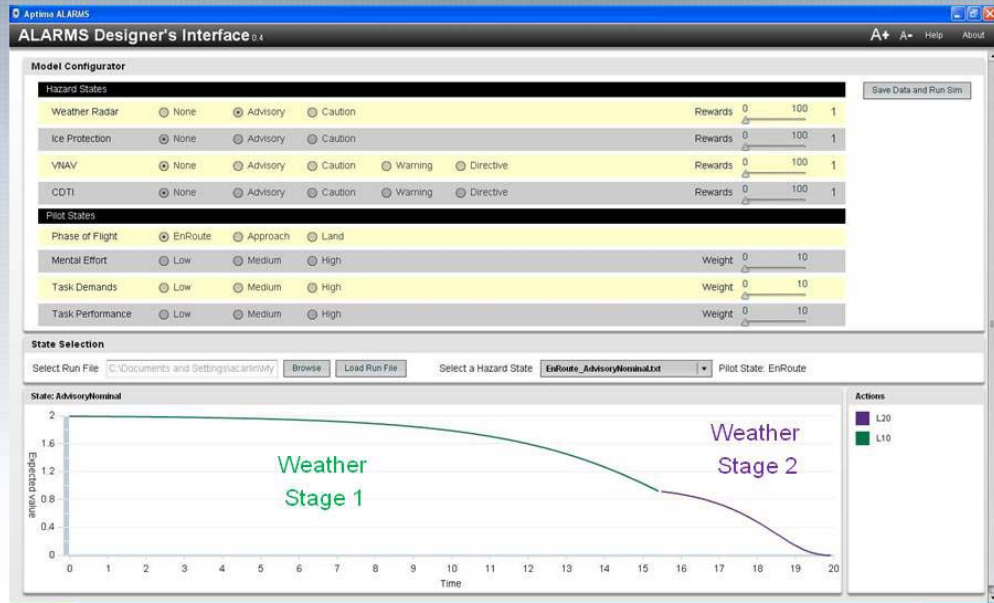
*Interface GUI programmed in FLEX by Gilbert Mizrahi at Aptima

*Interface GUI programmed in FLEX by Gilbert Mizrahi at Aptima

671
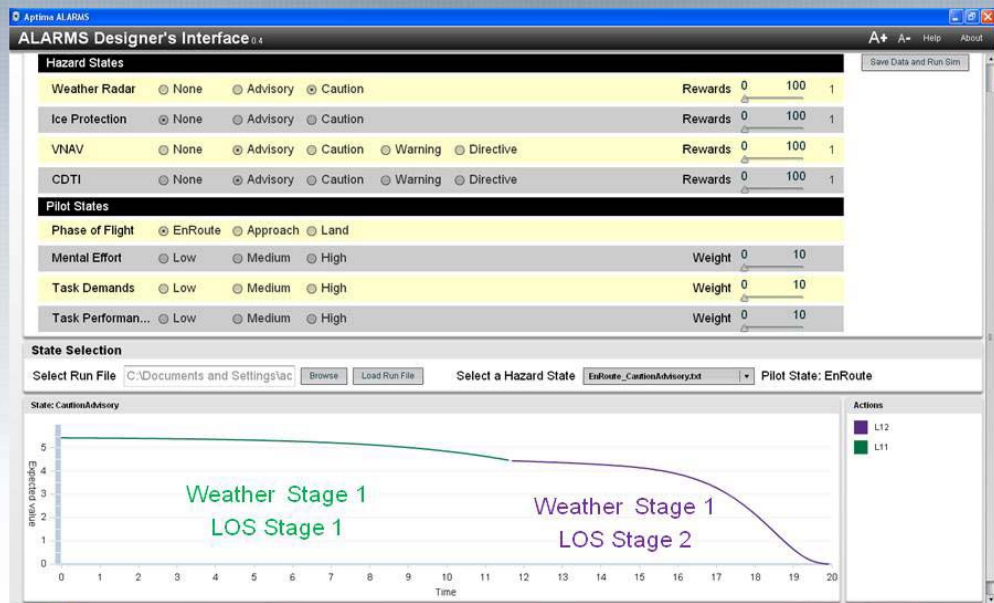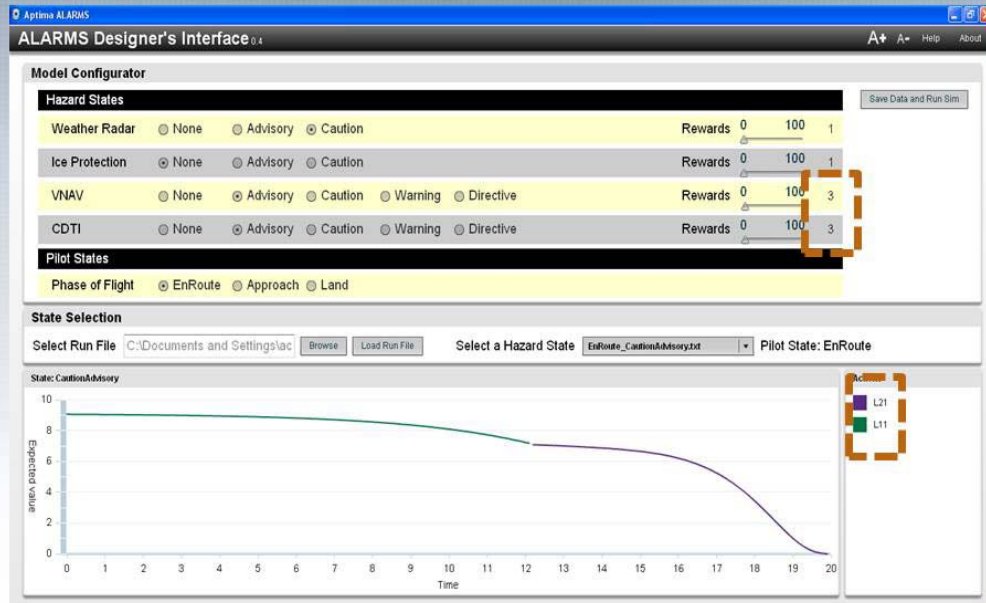
*Interface GUI programmed in FLEX by Gilbert Mizrahi at Aptima
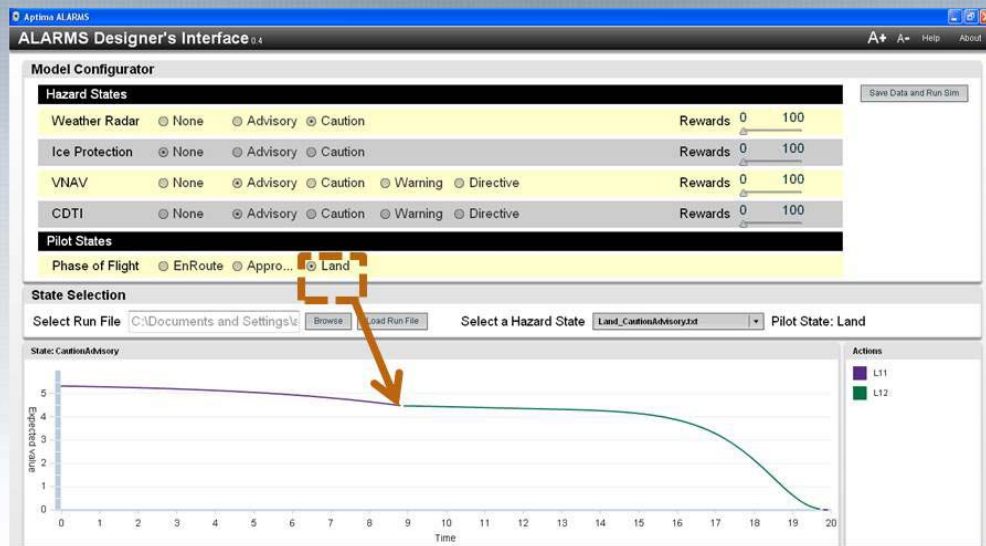
*Interface GUI programmed in FLEX by Gilbert Mizrahi, M.S., at Aptima

672

Conclusions

- **Contributions of ALARMS**
  - Introduced system architecture
  - Developed interface mockups
  - Identified current and future hazards
  - Bayesian reasoning over uncertain hazard state and sensor systems
  - TMDP planning of interface stages of automation
    - TMDP planning reasons about time duration uncertainty
    - TMDP model is adaptable to empirical findings
  - Accounts for pilot state
  - Developed system designer interface for understanding system behavior
- **Future work**
  - System integration and empirical evaluation

27

Questions?

Questions?

28

5.10 Imbalanced Learning for Functional State Assessment

Imbalanced Learning for Functional State Assessment

Feng Li, Frederick McKenzie and Jiang Li
Old Dominion University
flixx003@odu.edu rdmckenz@odu.edu jli@odu.edu

Guangfan Zhang and Roger Xu
Intelligent Automation, Inc.
gzhang@i-a-i.com hgxu@i-a-i.com

Carl Richey and Tom Schnell
University of Iowa
richeyc@ccad.uiowa.edu thomas-schnell@uiowa.edu

Abstract. This paper presents results of several imbalanced learning techniques applied to operator functional state assessment where the data is highly imbalanced, i.e., some function states (majority classes) have much more training samples than other states (minority classes). Conventional machine learning techniques usually tend to classify all data samples into majority classes and perform poorly for minority classes. In this study, we implemented five imbalanced learning techniques, including random under-sampling, random over-sampling, synthetic minority over-sampling technique (SMOTE), borderline-SMOTE and adaptive synthetic sampling (ADASYN) to solve this problem. Experimental results on a benchmark driving test dataset show that accuracies for minority classes could be improved dramatically with a cost of slight performance degradations for majority classes.

1.0 INTRODUCTION

An Operator Functional State (OFS) refers to a multidimensional pattern of the human psychophysiological condition that mediates performance in relation to physiological and psychological costs [1]. Accurate OFS assessment for human operators plays critical roles in automated aviation systems because it can ensure mission success and improve mission performances [2].

Researchers proposed various modeling tools to assess OFS. In Ref. [3], a stepwise discriminate analysis (SWDA) method and artificial neural networks (ANN) were proposed to perform OFS assessment. As a nonlinear model, the ANN is considered more advantageous in complex task situations, especially if multiple features are used. In Ref. [2], committee machines proved useful in improving the assessment accuracy. Errors of individual committee members can be canceled if the errors are independent. Therefore, improvement can be achieved if individual members have low biases and are less correlation i.e., they are diversified [4]. In addition to the traditional "bagging" technique, which generates multiple versions of prediction based on the bootstrap technique to produce the final prediction [5], performing a feature selection procedure before training can further reduce correlations among committee members [2].

To successfully perform OFS assessment, however, researchers often face the challenge of modeling imbalanced datasets where datasets are not balanced, i.e., some OFS states have much more data samples than others do. In the machine learning community, those OFSs having more data samples than others are named 'majority' classes while those having less samples are called 'minority' classes. Traditional classifiers tend to classify all data samples into majority classes, resulting in poor performances for minority classes [6], which is not acceptable for OFS assessment.

Many imbalanced learning techniques have been proposed to balance performances among majority and minority classes. Those techniques could be divided into four categories [6]: sampling methods, cost-sensitive methods, kernel-based methods, and active learning methods. Sampling methods aim to reduce the imbalance by removing (under-sampling) samples from majority classes or generating (over-sampling) more training samples for minority classes [7]. Cost-sensitive methods improve classification performance by using different cost matrices to compensate for

imbalanced classes [8]. Kernel based methods, such as the support vector machine (SVM), are based on the principles of statistical learning and Vapnik-Chervonenkis (VC) dimensions [9]. Active learning is a type of iterative supervised learning technique, which is used in situations where unlabeled data is abundant. Active learning is often integrated into kernel-based learning methods by selecting the closest instance to the current hyper plane from the unseen training data and adding it to the training set to retrain the model [10].

We have developed an OFS assessment strategy based on a committee machine for a closed-loop adaptive task manage system, where the OFS assessment was treated as a regression problem [2]. In this paper, we redesigned a similar model for the same task; however, we treated the OFS assessment as a classification problem. Because the data sets are highly imbalanced, traditional classifiers failed to classify minority states. We implemented several imbalanced techniques to improve classification performances for those minority OFS states.

The remainder of the paper is organized as follows: Section 2 describes several imbalanced learning techniques implemented in this paper. Section 3 presents the architecture of a committee classifier. Section 4 illustrates our experimental design, including implementation of the imbalanced learning techniques and the design of a committee classifier. Section 5 shows our achieved experimental results. Section 6 provides discussions for the results and Section 7 concludes the paper.

2.0 IMBALANCED LEARNING TECHNIQUES

There exist many imbalanced learning techniques in the literature as described in the excellent review paper [6]. In our study, we implemented five of them as described below.

- Random under-sampling
- Random over-sampling
- Synthetic minority over-sampling technique (SMOTE)
- Borderline-SMOTE
- Adaptive synthetic sampling (ADASYN)

All the methods have been detailed in the Ref. [6], including their implementations, performances and limitations. The overall goal of those methods is to make data samples balanced among classes by dropping some data samples from majority classes and adding samples to minority classes, and to keep roughly the equal number of data samples for all classes.

2.1 Random under-sampling
Random under-sampling was only applied to majority classes. The method randomly selects a number of majority data samples to keep. This method may loss information in the majority classes.

2.2 Random over-sampling
The random over-sampling method was only utilized to minority classes. In contrary to the random under-sampling technique, this method randomly selects data samples from minority classes and duplicates them till the data set is roughly balanced. This method may lead to overfitting because data samples are repeatedly used.

2.3 SMOTE
To overcome the overfitting defect of the random over-sampling method, SMOTE generates or synthesizes new samples for minority classes. To create a new synthetic sample for a given data point (seed) from minority classes, it first randomly selects one of its K-nearest minority neighbors (K is specified by researchers arbitrarily). Then, a random point that is on the line between the seed and the selected neighbor will be synthesized as a new data sample. SMOTE may lead to the problem of over generalization [12]. The following methods, Borderline-SMOTE and ADASYN, are developed to overcome this limitation.

2.4 Borderline-SMOTE

Borderline-SMOTE and SMOTE differ in the ways they select seeds. SMOTE may select any minority sample as a seed while Borderline-SMOTE only considers those who are from minority classes and are on the borderline between minority and majority classes. A minority class sample is considered as on the borderline if majority of its M nearest samples belong to majority classes (M is specified by researchers arbitrarily).

2.5 ADASYN

The difference between ADASYN and SMOTE is the amount of new data samples to be synthesized for each seed. SMOTE generates the same number of data samples for each seed while ADASYN syntheses data samples according to the distribution of seeds. Considering K nearest neighbors of a seed, the more belonging to majority classes, the more new samples will be synthesized for the seed.

3.0 COMMITTEE MACHINE

A committee machine is an ensemble of multiple estimators (committee members), which could be any learning method for classification or regression. The output of a committee machine is fusion of the outputs from all of its members. A theoretic interpretation for the principle of committee machine is that the errors from individual committee members can be canceled to some extent if they are uncorrelated.

Research results show that the performance improvement can be affected by two factors: accuracies of individual committee members and correlations among them [4]. For the first factor, selection of an appropriate individual model is essential, because a better performance will usually be achieved if each of the individual members performs well. For the second factor, several techniques like bagging, boosting, averaging or voting, mixture of experts have proved effective [4]. In this paper, we use the following techniques to build the committee machine.

- Use the bootstrapping technique to generate multiple 'copies' of the training data.

- Apply an advanced feature selection algorithm, Piecewise Linear Orthogonal Floating Search (PLOFS) [11], to diversify the committee members such that their performances are not highly correlated.

- Train a Multi-Layer Perceptron (MLP) by the standard Back Propagation (BP) algorithm as a base classification model

- Delete the committee members having high biases (accuracy < 50%).

- Utilize the majority vote scheme to fuse decisions from committee members. For example, if majority of the 15 total committee members predict class 1, the final output of the committee is class 1.

The system diagram of the committee machine is shown in Figure. 1.
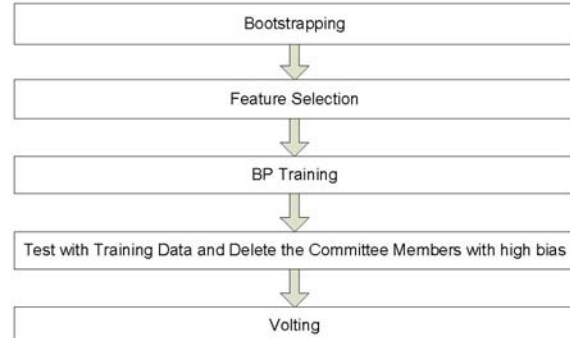
Figure 1: Diagram of the Committee Machine

4.0 EXPERIMENT DESIGN

4.1 The driving test dataset

We utilized a driving test dataset to validate our proposed method for OFS assessment. The dataset was collected by participants performing a driving test over the course of two hours. The collected information includes description of the driving task, system dynamics related information, performance measures, physiological signals (128-channel EEG, ECG,

677

respiration, etc.), and eye tracking. The workload was also analyzed according to the driving conditions (city-driving, stopped, highway passing, etc.), and seven OFSs, which indicate seven workload levels, were defined.

Six subjects participated in the driving test and data was recorded in a separate file for each participant, resulting in six individual datasets. Each dataset has seven operator functional states (workload) that are considered as seven classes by our committee classifier. In the dataset, the number of data samples in each class is not balanced. Four classes (minority class) have much less data samples than other three majority classes do. Table 1 and Figure 2 show data distributions for all classes.

Data distributions are similar for all subjects. Class 2 has the largest number of samples (about 35% of the whole data). Class 3 and 4 have the second largest number of samples (about 20%). Therefore, around 75% of samples belong to those three classes. Class 7 has the smallest number of samples accounting for less than 1% of the whole data, and subjects 2, 4 and 6 even have no data for class 7. Class 6 is the second smallest class having about 3% of the whole data samples. Both class 1 and 5 account for 5% of the data samples.

4.2 Imbalanced learning techniques

To implement the five imbalanced techniques, we first compute a desired percentage of data samples per class as,

$$N_d = 100 \text{ / no. of classes} * 100\%$$

$$= 100/7 * 100\% = 14.29\%$$

We then calculate a high threshold (T_H) and a low threshold (T_L) for the number of data samples in each class as,

$$
\begin{aligned}
T_H &= N_d * (1 + 0.1) \\
&= 14.29\% * 1.1 = 15.71\% \\
T_L &= N_d * (1 - 0.1) \\
&= 14.29\% * 0.9 = 12.86\%
\end{aligned}
$$

Table 1: Data Distribution among Classes

Class	Data set 1 (%)	Data set 2 (%)	Data set 3 (%)	Data set 4 (%)	Data set 5 (%)	Data set 6 (%)
1	6.17	8.70	6.29	5.59	3.52	3.86
2	38.34	39.24	33.83	39.66	32.65	39.87
3	19.96	21.42	24.56	32.94	26.39	20.16
4	23.55	19.40	21.07	16.43	31.24	27.05
5	8.03	8.25	11.30	3.03	2.99	6.10
6	3.89	2.98	2.67	2.35	2.99	2.96
7	0.06	0.00	0.28	0.00	0.22	0.00
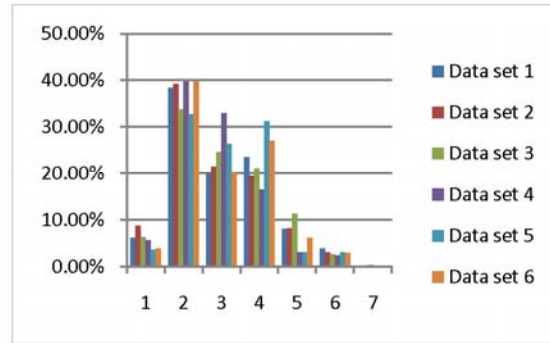
Figure 2: Data Distribution among Classes

Classes having data samples more than T_H are considered as majority classes while classes with data samples less than T_L are considered as minority classes and others are treated as medium classes.

As such, there are seven classes and N_d, T_L and T_H are 14.29%, 12.86% and 15.71%, respectively. Referring to Table 1, it is clear that classes 2, 3 and 4 are majority classes. Class 1, 5, 6 and 7 are minority classes and there is no medium class in our datasets. In order to achieve a balanced dataset, the data portions in both majority and minority classes are made roughly the same as N_d. We apply the random under-sampling technique to the majority classes and four over-sampling methods to the minority classes, resulting in four balanced datasets as shown in Figure 3. For each participant, the balanced dataset shares the majority classes' data samples but has different data samples from minority classes, depending upon which oversampling method is used.

678

Figure 3: Generation of Balanced Datasets
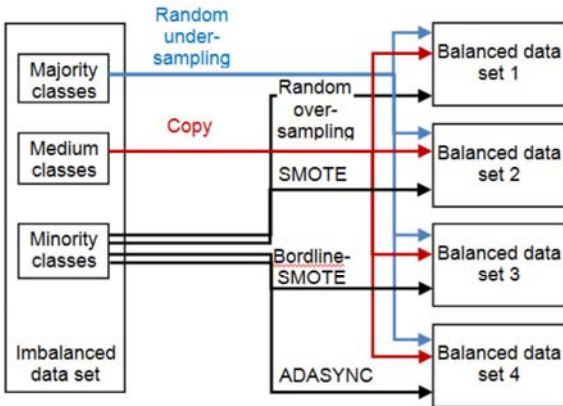
4.3 Committee classifier

The committee classifier consists of a bootstrap procedure, a feature selection process and a majority voting scheme (see Figure. 4). A MLP trained by the BP algorithm was implemented as the base classification model. Basic procedures performed by the committee classifier are as follows:

1. Randomly divide a subject's dataset into two parts with equal number of data points, one for training and another for testing.

2. Generate M bootstrapped datasets for the training dataset.

3. Apply one of the imbalanced learning techniques to the bootstrapped datasets. A balanced dataset is then obtained for each of the M datasets.

4. Select a set of most effective features for each of the balanced datasets using the PLOFS algorithm. Selected features for different datasets maybe different.

5. Train a MLP classifier for each of the datasets using the features selected for that dataset.

6. Apply the trained MLP to the training and testing datasets.

7. Generate the final classification result by majority voting. MLPs having training

accuracies greater than 50% are used only. Repeat the above procedures by exchanging the role of training and testing datasets.

8. Repeat the above steps for each of the imbalanced learning techniques described in Section 3.
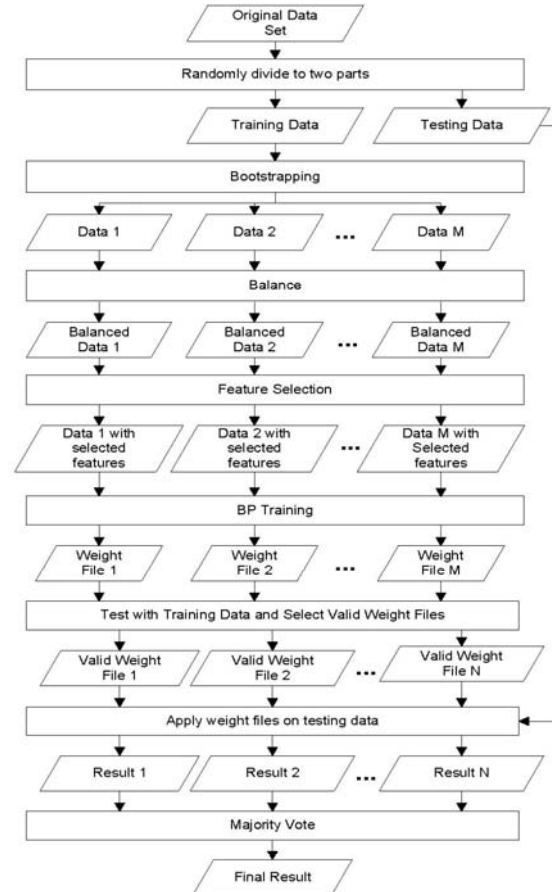
Figure 4: Design of the Committee Classifier

5.0 RESULTS

We trained a committee classifier for each of the six participants (datasets) and results are shown in Tables 2 - 7 and Figs. 5 - 10.

In the Tables, the 'Untreated' column illustrates results achieved on the original data sets. Other four columns present accuracies (in percentage) for each class achieved by applying the four imbalanced learning techniques to the minority classes. The last row shows the average (overall)

679

accuracies achieved by each of the techniques.

6.0 DISCUSSION

It is observed that the classification accuracies are highly imbalanced if no imbalanced learning technique is used. For instances, the minority class 7 always has 0% accuracy for all subjects but good performances are usually achieved for majority classes 2, 3 and 4. Classification accuracies have been balanced among minority and majority classes by applying the four imbalanced learning techniques to minority classes. Accuracies often have been significantly improved for minority classes while those for majority classes have been decreased slightly. As a result, the overall performance has been slightly degraded. Note that different sampling algorithms appear to perform similarly, indicating the robustness of the imbalanced learning techniques.

Table 3: Results for Dataset 2

Class	Un treated	Over Sample	Smote	Border	AdaSyn
1	96.23%	96.92%	98.97%	99.32%	99.32%
2	99.92%	98.25%	99.16%	98.56%	98.33%
3	88.46%	75.10%	73.44%	84.98%	79.00%
4	67.43%	57.91%	46.08%	61.14%	63.75%
5	10.47%	55.60%	74.01%	23.10%	24.19%
6	58%	96%	90%	81%	89%
Over all	82.21%	81.76%	80.99%	81.70%	81.17%
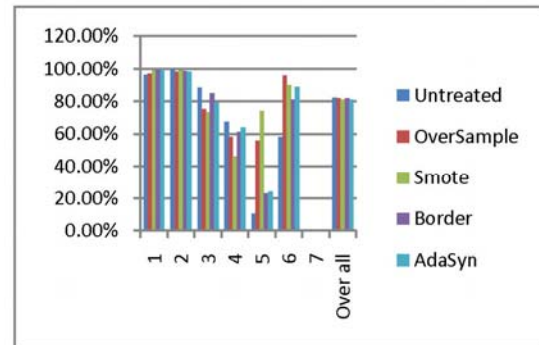
Figure 6: Results for Dataset 2

Table 2: Results for Dataset 1

Class	Un treated	Over sample	Smote	Border	Ada Syn
1	94.27%	97.40%	98.44%	96.35%	97.40%
2	99.75%	98.58%	99.08%	98.49%	98.99%
3	90.82%	85.67%	74.24%	77.78%	82.13%
4	70.53%	57.30%	42.84%	66.85%	59.07%
5	2.80%	38.40%	71.20%	15.60%	26.80%
6	57.02%	89.26%	82.64%	66.12%	73.55%
7	0%	100%	100%	100%	100%
Over all	81.23%	81.01%	77.96%	78.86%	79.34%

Figure 5: Results for Dataset 1

Table 4: Results for Dataset 3

Class	Un treated	Over Sample	Smote	Border	AdaSyn
1	89.83%	93.22%	94.92%	95.48%	94.35%
2	100%	96.74%	96.11%	96.74%	96.64%
3	78.00%	65.70%	61.51%	67.15%	65.56%
4	47.05%	45.19%	36.26%	46.21%	47.05%
5	45.91%	52.52%	68.24%	52.52%	51.89%
6	70.67%	85.33%	78.67%	78.67%	77.33%
7	0%	100%	87.50%	87.50%	87.50%
Over all	75.62%	72.74%	71.29%	73.24%	72.81%

Figure7: Results for Dataset 3

680
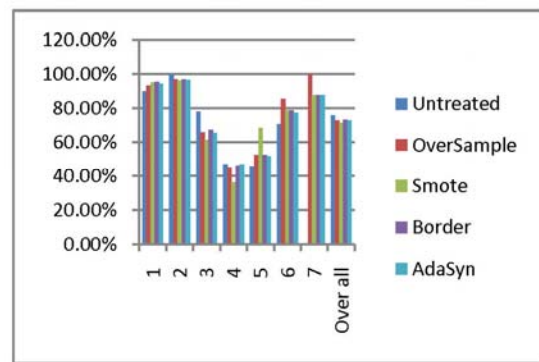
Table 4: Results for Dataset 4

Class	Un treated	Over Sample	Smote	Border	AdaSyn
1	60.49%	88.29%	89.76%	85.85%	89.76%
2	96.97%	93.87%	93.67%	95.94%	95.53%
3	90.39%	72.74%	66.36%	77.13%	75.64%
4	67.11%	70.27%	70.43%	72.43%	71.26%
5	34.23%	63.96%	61.26%	51.35%	53.15%
6	44.19%	96.51%	96.51%	87.21%	95.35%
Over all	84.72%	81.88%	79.72%	83.76%	83.38%
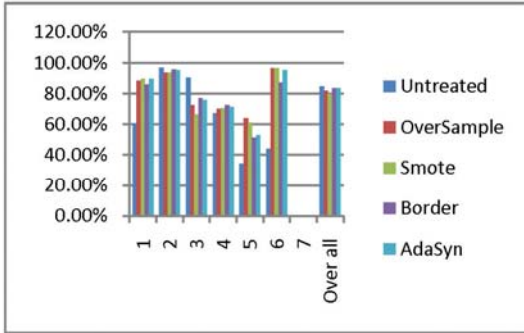
Figure 7: Results for Dataset 4

Table 6: Results for Dataset 6

Class	Un treated	Over Sample	Smote	Border	AdaSyn
1	92.86%	97.32%	99.11%	98.21%	95.54%
2	97.23%	93.78%	93.00%	93.78%	94.90%
3	89.23%	75.04%	71.62%	85.30%	74.36%
4	68.28%	59.24%	49.30%	64.20%	60.13%
5	3.95%	43.50%	55.37%	10.73%	27.68%
6	29.07%	87.21%	83.72%	60.47%	82.56%
Over all	79.91%	77.53%	74.53%	78.19%	76.91%
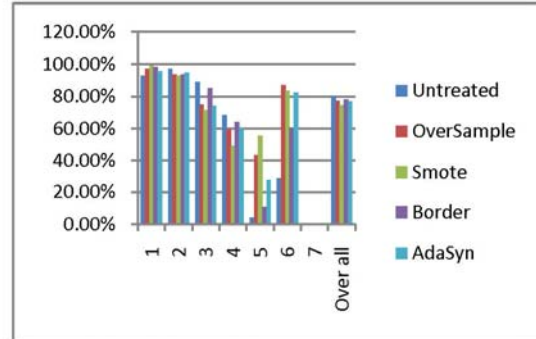
Figure 9: Results for Dataset 6

Table 5: Results for Dataset 5

Class	Un treated	Over Sample	Smote	Border	AdaSyn
1	55.36%	95.54%	98.21%	95.54%	96.43%
2	96.63%	97.88%	97.69%	98.07%	97.50%
3	82.96%	75.92%	73.78%	78.31%	77.83%
4	59.62%	26.99%	30.21%	35.65%	34.14%
5	53.68%	61.05%	77.89%	62.11%	73.68%
6	84.21%	94.74%	94.74%	97.89%	96.84%
7	0%	100%	71.43%	85.71%	71.43%
Over all	78.14%	68.67%	69.58%	72.16%	71.69%
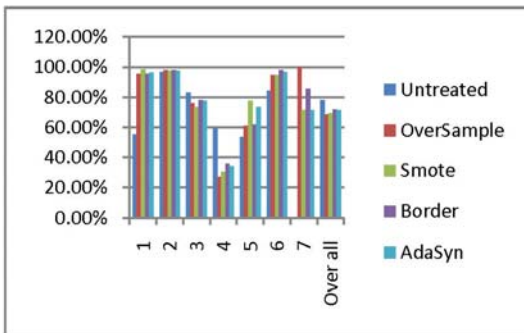
Figure 8: Results for Dataset 5

7.0 CONCLUSIONS

We have implemented five different imbalanced techniques for OFS assessment and validated our methods on driving test benchmark datasets. Experimental results consistently show that classification accuracies for minority classes in the tested datasets are improved dramatically with a cost of slight performance degradations for majority classes, indicating that imbalanced learning techniques could be very useful for OFS assessment.

In a practical setting, an OFS assessment model will be trained offline. We can utilize the imbalanced learning techniques to improve recognition accuracies of the assessment model for minority OFSs without severely decreasing assessment effectiveness for majority OFSs. Once the model is trained, it will then be able to recognize all possible OFSs relatively accurately on the fly. This is critical because some minority OFSs may be highly correlated to aviation safety.

Our future work includes further testing the applicability of more imbalanced learning techniques to the OFS assessment task, validating those methods on more subjects' datasets and integrating the most effective scheme into a real time OFS assessment system.

8.0 ACKNOWLEDGEMENT

9.0 REFERENCES

[1]. NATO Advanced Research Workshop on Operator Functional State and Impaired Performance in Complex Work Environments (2002 : Il Ciocco), "Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks," (NATO Asi) (Vol 355).

[2]. Guangfan Zhang, Roger Xu, Wei Wang, Jiang Li, Tom Schnell, Mike Keller, "Individualized Cognitive Modeling for Closed-Loop Task Mitigation," ModSim World Conference, 2009, Virginia Beach, VA

[3]. Glenn F. Wilson and Chris A. Russell, "Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task," Human Factors: The Journal of the Human Factors and Ergonomics Society 2003 45:381.

[4]. Volker Tresp," Committee Machines," a book chapter in: Handbook for Neural Network Signal Processing, CRC Press, 2001.

[5]. Leo Breiman, "Bagging Predictors," Machine Learning, 24 (2), pp. 123-140, 1996.

[6]. Haibo He, "Learning from Imbalanced Data," IEEE Transactions on Knoledge and Data Engineering, Vol. 21, NO. 9, September 2009.

[7]. A. Estabrooks, T. Jo, and N. Japkowicz, "A Multiple Resampling Method for Learning from Imbalanced Data Sets," Computational Intelligence, vol. 20, pp. 18-36, 2004.

[8]. C. Elkan, "The Foundations of Cost-Sensitive Learning," Proc. Int'l Joint Conf. Artificial Intelligence, pp. 973-978, 2001.

[9]. Corinna Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, 20 (3), pp. 272, 297, 1995.

[10]. Seyda Ertekin, Jian Huang, C. Lee Giles, "Active Learning for Class Imbalance Problem," Annual ACM Conference on Research and Development in Information Retrieval, 2007.

[11]. Jiang Li et al., "Feature Selection Using a Piecewise Linear Network," IEEE Trans. Neural Network, vol. 17, no. 5, pp. 1101-1115, 2006.

[12]. B.X. Wang and N. Japkowicz, "Imbalanced Data Set Learning with Synthetic Samples", Proc. IRIS Machine Learning Workshop, 2004.

MODSIM WORLD
Conference & Expo

Imbalanced Learning for Functional State Assessment

Feng Li, Frederic McKenzie, Jiang Li
Old Dominion University, Norfolk, VA

Guangfan Zhang, Roger Xu
Intelligent Automation, Inc. Rockville, MD

Carl Richey, Tom Schnell
University of Iowa, Iowa city, IA

October 14, 2010

1

MODSIM WORLD
Conference & Expo

Outline

- Introduction
- Imbalanced Learning Techniques
- Experiment Design
- Results & Discussion
- Conclusion

2

Operator Functional State (OFS)

Related Work

- a stepwise discriminate analysis (SWDA) method and artificial neural networks (ANN) were proposed to perform OFS assessment [3];
- committee machine proves to be useful in improving the assessment accuracy [2];
  - Bootstrap (Bagging)
  - Feature Selection

Recent Work

- Data Description
- Committee Machine
- Regression

Final Result

Average BP Results

BP 1 | BP 2 | BP M

Feature selection

Data Sampl | Data Sampl | Data Sample

Bootstrapping

Original Data

5

Data Description

Distribution of data samples
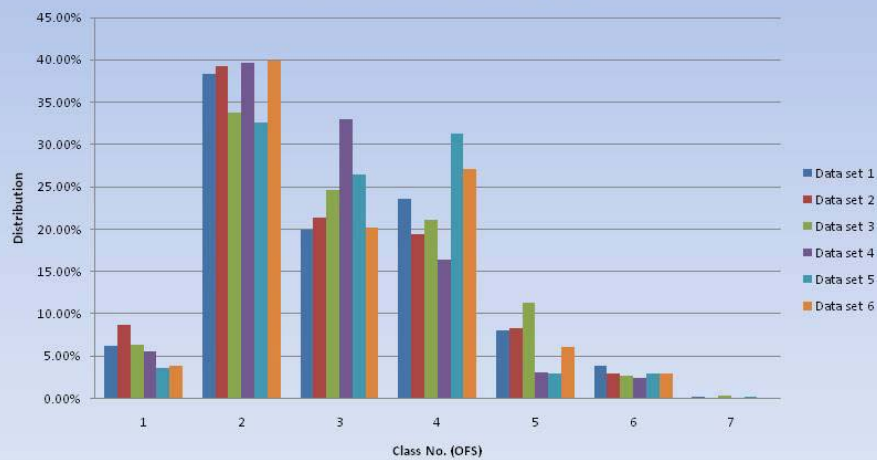
Challenge

- Imbalanced data set
- Problem
- Techniques [6]
  - Sampling methods
  - Cost-sensitive methods
  - Kernel-based methods
  - Active learning methods

Minority OFS/Class

Majority OFS/Class

5%

95%
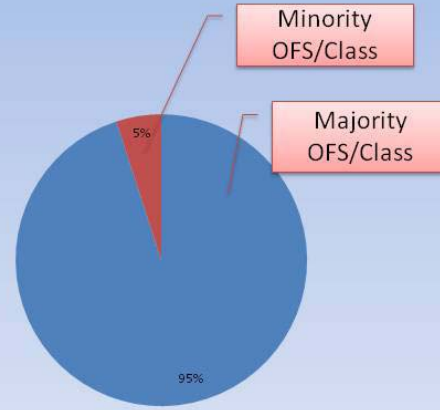
Distribution of 2-class data set

7

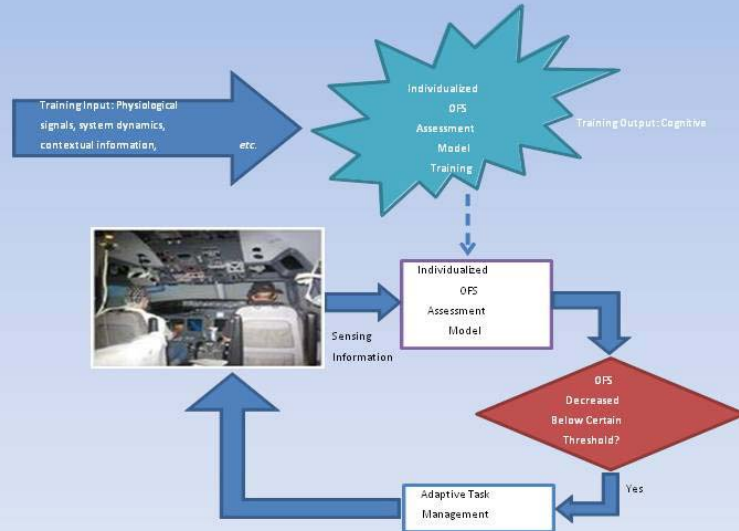# Our Goal

- Classification based on Committee Machine
- Imbalanced Learning Techniques

8

Outline

- Introduction
- Imbalanced Learning Techniques
- Experiment Design
- Results & Discussion
- Conclusion

Sampling methods

- Random under-sampling [6]
- Random over-sampling
- Synthetic minority over-sampling (SMOTE)
- Borderling-SMOTE
- Adaptive synthetic sampling (ADASYN)

Random Under-sampling

- Randomly select a number of majority data samples and remove them.
- May loss important information.
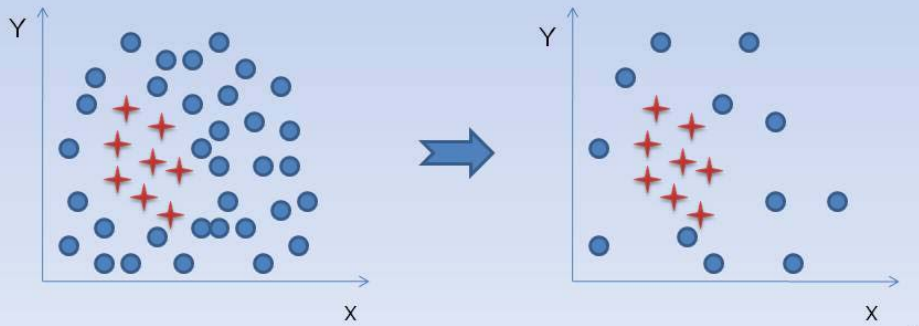
Random Over-sampling

- Randomly select minority data samples and duplicate them to training data set.
- May lead to over-fitting.
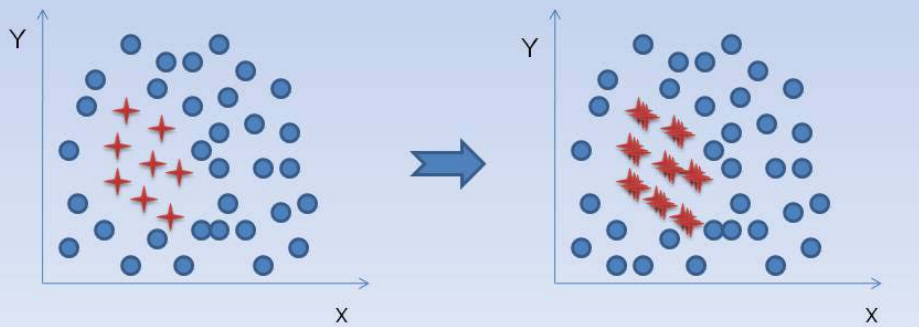
Synthetic Minority Over-sampling Technique (SMOTE)

- Synthesize a random point on the line between the seed and the neighbor

- May lead to overgeneralization
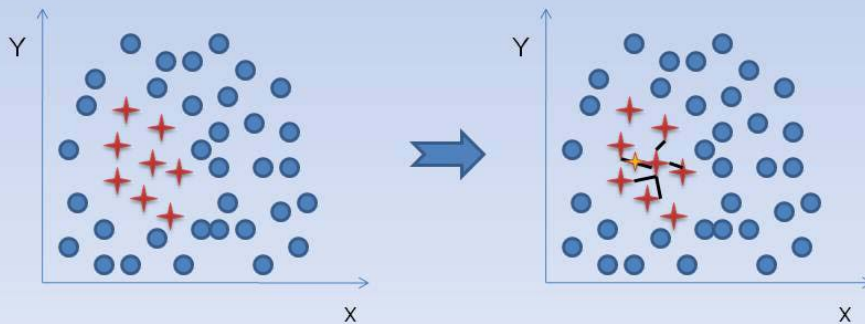
Borderline-SMOTE

- Borderline-SMOTE and SMOTE differ in how they select seeds. Borderline-SMOTE only selects seeds on the borderline.
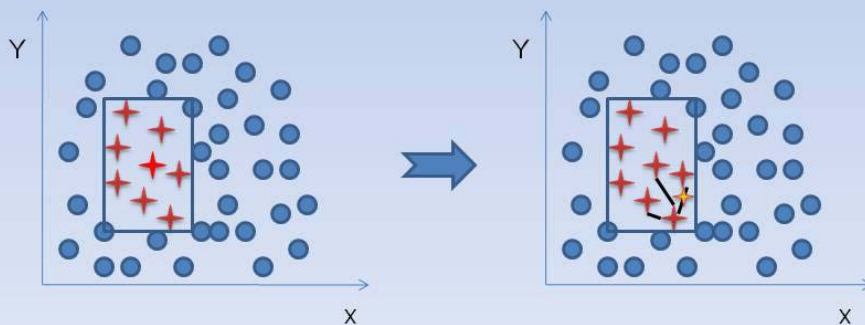
Adaptive Synthetic Sampling (ADASYN)

- ADASYN and SMOTE defer in the amount of new samples that need to be synthesized. The more neighbors belong to majority OFSs, the more samples need to be synthesized.

15

Outline

- Introduction
- Imbalanced Learning Techniques
- Experiment Design
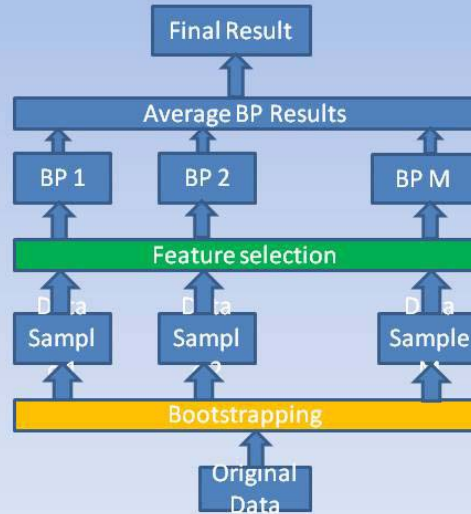- Results & Discussion
- Conclusion

16

690

Bootstrapping

Feature Selection
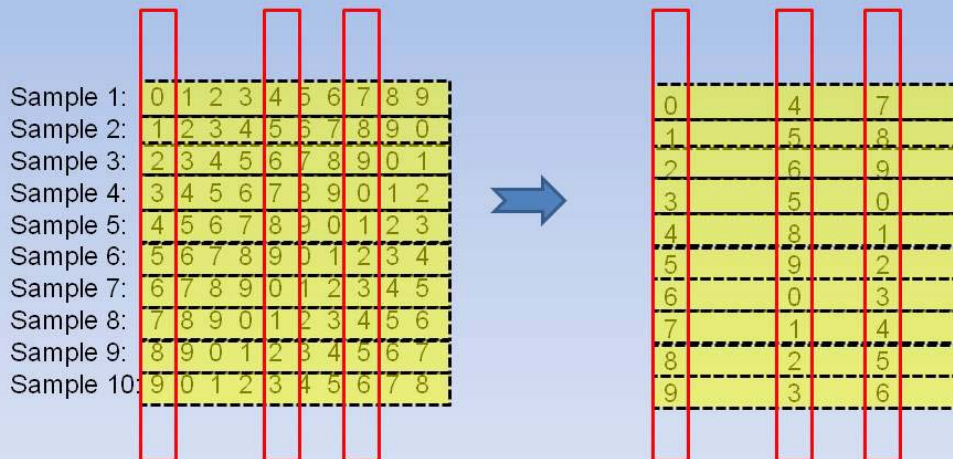
Piecewise Linear Orthogonal Floating Search (PLOFS)

Majority Voting

1   0   1   1

23
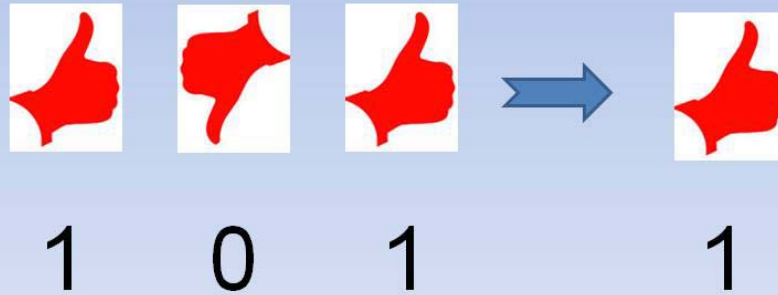
Framework of the OFS Assessment Strategy
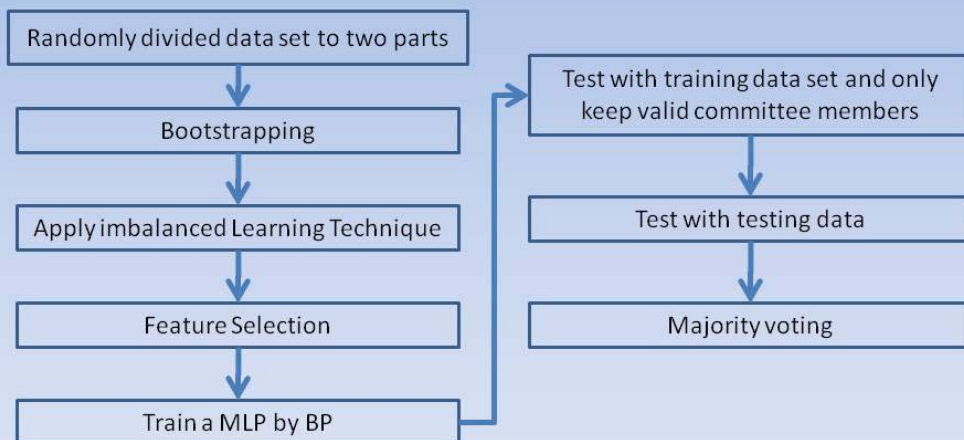
Randomly divided data set to two parts

Bootstrapping

Apply imbalanced Learning Technique

Feature Selection

Train a MLP by BP

Test with training data set and only keep valid committee members

Test with testing data

Majority voting

24

694

Outline

- Introduction
- Imbalanced Learning Techniques
- Experiment Design
- Results & Discussion
- Conclusion

Result for Data Set 1
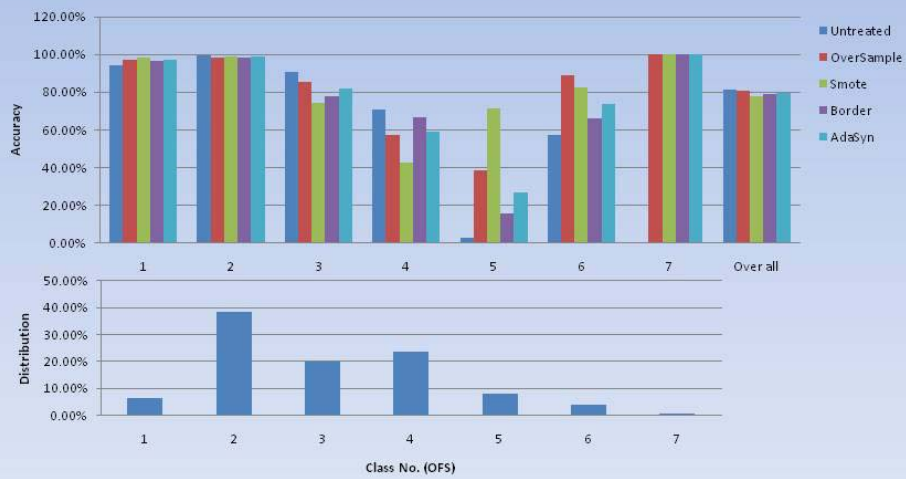
Accuracy of Classification

Result for data set 2

Result for data set 3

696

Result for data set 4

Result for data set 5

697

Result for data set 6

Accuracy of Classification

Legend: Untreated, OverSample, Smote, Border, AdaSyn

X-axis: 1, 2, 3, 4, 5, 6, 7, Over all

Lower chart — Distribution vs Class No. (OFS): 1, 2, 3, 4, 5, 6, 7
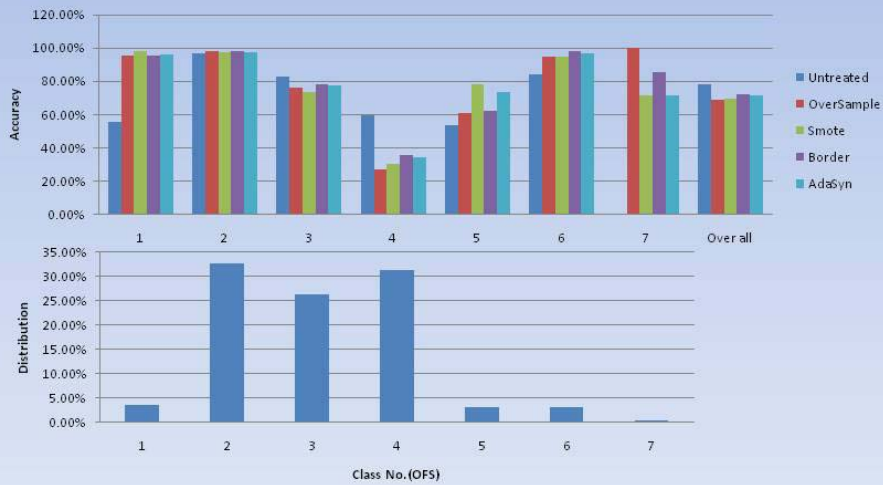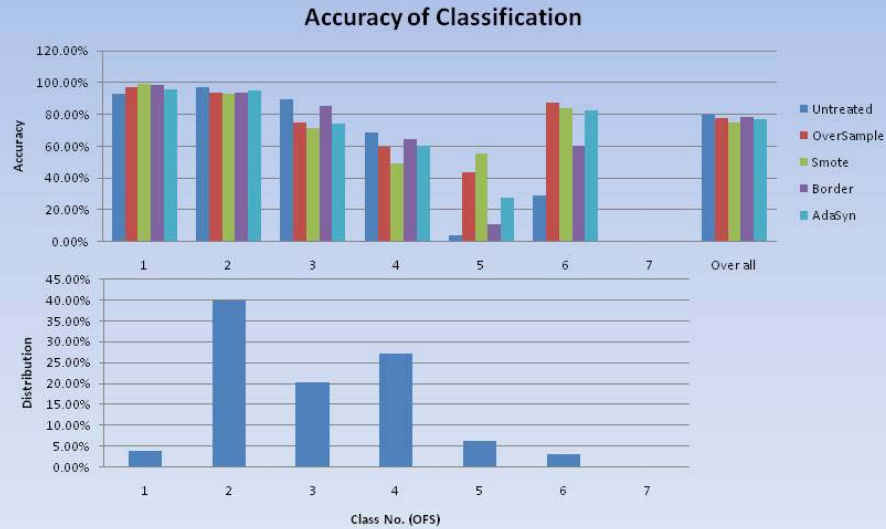
31

Outline

- Introduction
- Imbalanced Learning Techniques
- Experiment Design
- Results & Discussion
- Conclusion

32

698

Conclusions

- By using imbalanced learning techniques, classification accuracies for minority OFSs are improved dramatically with a cost of slight performance degradations for majority OFSs
- Different sampling algorithms appear to perform similarly
- Future work
  - Test more imbalanced techniques
  - Validate those techniques on more subjects' datasets

33

Acknowledgement

34

Reference

1. NATO Advanced Research Workshop on Operator Functional State and Impaired Performance in Complex Work Environments (2002 : Il Ciocco), "Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks," (NATO Asi) (Vol 355).
2. Guangfan Zhang, Roger Xu, Wei Wang, Jiang Li, Tom Schnell, Mike Keller, "Individualized Cognitive Modeling for Closed-Loop Task Mitigation," ModSim World Conference, 2009, Virginia Beach, VA
3. Glenn F. Wilson and Chris A. Russell, "Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task," Human Factors: The Journal of the Human Factors and Ergonomics Society 2003 45:381.
4. Volker Tresp," Committee Machines," a book chapter in: Handbook for Neural Network Signal Processing, CRC Press, 2001.
5. Leo Breiman, "Bagging Predictors," Machine Learning, 24 (2), pp. 123-140, 1996.
6. Haibo He, "Learning from Imbalanced Data," IEEE Transactions on Knoledge and Data Engineering, Vol. 21, NO. 9, September 2009.
7. A. Estabrooks, T. Jo, and N. Japkowicz, "A Multiple Resampling Method for Learning from Imbalanced Data Sets," Computational Intelligence, vol. 20, pp. 18-36, 2004.
8. C. Elkan, "The Foundations of Cost-Sensitive Learning," Proc. Int'l Joint Conf. Artificial Intelligence, pp. 973-978, 2001.
9. Corinna Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, 20 (3), pp. 272, 297, 1995.
10. Seyda Ertekin, Jian Huang, C. Lee Giles, "Active Learning for Class Imbalance Problem," Annual ACM Conference on Research and Development in Information Retrieval, 2007.
11. Jiang Li et al., "Feature Selection Using a Piecewise Linear Network," IEEE Trans. Neural Network, vol. 17, no. 5, pp. 1101-1115, 2006.
12. B.X. Wang and N. Japkowicz, "Imbalanced Data Set Learning with Synthetic Samples", Proc. IRIS Machine Learning Workshop, 2004

35

•Thank you
• Q & A

36

5.11 Predicting the Consequences of Workload Management Strategies with Human Performance Modeling

Predicting the Consequences of Workload Management Strategies with Human Performance Modeling

Diane Kuhl Mitchell & Charneta Samms
Army Research Laboratory
Aberdeen Proving Ground, Maryland
diane.k.mitchell@us.army.mil csamms@arl.army.mil

Abstract. Human performance modelers at the US Army Research Laboratory have developed an approach for establishing Soldier high workload that can be used for analyses of proposed system designs. Their technique includes three key components. To implement the approach in an experiment, the researcher would create two experimental conditions: a baseline and a design alternative. Next they would identify a scenario in which the test participants perform all their representative concurrent interactions with the system. This scenario should include any events that would trigger a different set of goals for the human operators. They would collect workload values during both the control and alternative design condition to see if the alternative increased workload and decreased performance. They have successfully implemented this approach for military vehicle designs using the human performance modeling tool, IMPRINT. Although ARL researches use IMPRINT to implement their approach, it can be applied to any workload analysis. Researchers using other modeling and simulations tools or conducting experiments or field tests can use the same approach.

1.0 INTRODUCTION

As system engineers begin to design a system, it is critical for them to understand how the human operators will interact with the system. This understanding is critical because they are designing the system so the human operators can accomplish specific goals. The humans' ability to accomplish these goals, therefore, determines the effectiveness of the system design. A key component of the human operators' abilities to use the system, in turn, is their mental workload level. Mental workload is a key component because it influences the human operators' performance.

The relationship between human performance and mental workload is often represented as similar to the Yerkes-Dodson (1908) inverted-U relationship as shown in Fig 1. As Fig 1 indicates when mental workload is very low human performance will decline.

Figure 1 Inverted-U relationship between workload and performance (modified from Yerkes & Dodson, 1908).

As workload increases so does human performance. However, at some point workload transitions to a level high enough to overload human mental resources (Wickens, 2008). To manage the high workload, humans employ strategies to reduce workload to manageable levels. These strategies are called workload management strategies (Little, 1993). A strategy, for example, might be to stop an ongoing task, ignore a new task or to perform concurrent tasks sequentially. All of these workload management strategies can result in performance decrements.

For over a decade human performance researchers (Colle & Reid, 2005; Rueb, Vidulich, & Hassoun, 1994; Reid & Colle, 1988; Schlegel, B., Schlegel, R., & Gilliland, 1988; Grier, Wickens, Kaber, Strayer, Boehm-Davis, Trafton, & St. John, 2008) have attempted to refine the inverted-U representation of workload by identifying the point where workload and performance transition from acceptable to unacceptable. They refer to this transition point as the workload redline or threshold (Grier, et al, 2008). Identifying this workload threshold is important. If it could be determined, then human factors researchers could establish a workload level that is considered acceptable for optimum human performance. System engineers, in turn, could use this workload guidance to help ensure their system designs provide effective human performance. Despite the many years of research, there is, however, no consensus among researchers on a workload threshold.

A range of workload threshold values have been proposed by researchers who used the subjective workload assessment tool (SWAT) to estimate

workload. These researchers have proposed SWAT threshold values in the range of 40- 50 (Colle & Reid, 2005; Rueb, Vidulich, & Hassoun, 1994; Reid & Colle, 1988; Schlegel, B., Schlegel, R. & Gilliland, 1988). The SWAT workload range is useful for system engineers conducting system evaluations. In these evaluations human participants can give self-report workload ratings which SWAT requires.

Not all evaluations of system designs, however, include human participants that can give self-report workload ratings. Human performance modeling, for example, is an effective technique for evaluating system designs that includes mental workload evaluation but does not include human participants (Mitchell, 2000). The human operators are simulated in human performance models and, therefore, self-report workload scales, such as, SWAT cannot be used. Using human performance modeling, however, has several advantages over techniques that use human participants.

Human performance modeling is particularly useful early in the system development phase when finding a representative sample of human users of the proposed system can be costly and challenging due to funding constraints. In addition modeling can be used when a representative sample of users is unavailable or only a small sample size of users is available. Finally, it is useful when the design is still a concept and no system mock-ups exist. For human performance modeling techniques that include mental workload prediction as part of the system design evaluation a workload threshold remains critical.

Human performance modelers at the Army Research Laboratory have developed an analytical approach for establishing a workload threshold they can use for evaluation of a proposed system design. Their technique includes three key components. First, they create a scenario containing segments with each segment representing events that change the goals of the operators of the system. Second, they establish a baseline they can use for workload and performance comparisons. Finally, for each of these segments, they select unique workload threshold values for each operator who will operate the system.

In 2009, the ARL modelers implemented this approach in an analysis of the impacts of two conceptual technologies on the workload and performance of a tank crew (Mitchell, in review).

2.0   CASE STUDY

To implement their approach, the ARL modelers used the human performance modeling tool, IMPRINT (Improved Performance Research Integration Tool; http://www.arl.army.mil/IMPRINT). IMPRINT is a stochastic task- network modeling tool that provides modelers with the capability to simulate humans performing tasks. The humans simulated for this project were the tank crewmembers. Specifically, the ARL modelers built a model simulating the tasks performed by each crewmember of a baseline tank. Next, they built a model to represent the tasks performed by the tank crewmembers when the vehicle design was enhanced to include a driver's aid and a loader's situation awareness display.

In addition to simulating task performance, IMPRINT also provides modelers with the capability to predict the mental workload associated with the tasks individuals perform (Mitchell, 2000). The ARL modelers used this mental workload option to predict the mental workload of the crewmembers of the baseline tank as well as the enhanced tank. The theoretical basis for the IMPRINT mental workload option is Multiple Resource Theory (MRT) (Wickens, 2008).

According to MRT, the capacity of human mental resources is limited. Therefore, as an individual performs a task, the task makes demands upon these limited mental resources. Furthermore, when an individual performs two or more tasks concurrently, all the concurrent tasks demand some of the individual's mental resources. Because the mental resources have limits, the demands of the concurrent tasks may exceed or overload the individual's resources. The point where the individual's resources are overloaded is the workload threshold. When this threshold is exceeded, the individual implements workload management strategies which cause the individual's performance to decline.

Because the IMPRINT workload capability is based on MRT, its workload predictions are task-based predictions. Changes in the tank crewmembers workload, therefore, are related to changes in the tasks they perform in the baseline versus modified tank. If the technologies in the modified tank reduce crew workload then the IMPRINT workload predictions should be lower for the modified versus baseline tank model runs.

The IMPRINT tool implements MRT by providing modelers with the capability to enter the mental resources required by each task for the human operators of a proposed system. Furthermore, it provides numerical values for estimating the demands of the operators' tasks on their mental workload. IMPRINT provides these numerical values in the form of scales. There are seven scales, one for each resource. The resources represented by the seven scales are visual, auditory, cognitive, fine motor, gross motor, tactile, and speech.

Using the workload scales in IMPRINT, the ARL modelers selected the appropriate values for each of the resources that a tank crewmember used for each task. The IMPRINT software aggregated these workload inputs across all the tasks the crewmember performed every time a new task was started. IMPRINT then provided an overall workload score. This overall workload score is compared to a workload threshold set by the modelers. If the overall workload number exceeded the threshold than a workload management strategy is triggered within the model. Modelers can then see the impact of the crewmember's workload on performance with the system. Because the workload threshold is the key to determining if a workload management strategy is employed, it was critical for the IMPRINT modelers to select an appropriate workload threshold for the tank crew in their analysis.

As the first step in identifying a workload threshold for the tank crew analysis, the ARL modelers selected a scenario to model with IMPRINT. For the performance to be representative of the typical tank crew, the scenario needed to be one in which the crew performed the majority of their common tank crew tasks. These common tank crew tasks are driving, communicating, searching for targets, and engaging targets (Directorate of Training, Doctrine, and Combat Development Field Manual 3-20.15, 2007).

The ARL modelers needed to include common crew tasks in the scenario because they would build the tank crew tasks into the IMPRINT model based on the scenario. It was critical for the IMPRINT workload analysis to be valid that the crew be performing all the tasks the technologies might influence within the model. Furthermore, it was especially important for the ARL modelers to include in the scenario those common crew tasks the crewmembers perform concurrently. The

inclusion of concurrent tasks in the models was important because workload is typically higher, and performance is typically lower, for concurrent tasks than sequential tasks (Just, Carpenter, Keller, Emery, Zajac, & Thulborn, 2001). To meet these scenario characteristics, the modelers selected a movement to contact mission (Directorate of Training, Doctrine, and Combat Development Field Manual 3-20.15, 2007).

After selecting the scenario, the ARL modelers divided the mission into segments that represented changes in the crewmembers' goals. For example, as a movement-to-contact mission begins, the crewmembers' goal is to detect the enemy, whereas, once they detect the enemy their goal shifts to destroying the enemy. As a consequence of the shift in goals between the two segments, the crew performs different tasks. Because the workload predictions in the IMPRINT model are based on task demands, the crews' workload will change along with the tasks. Therefore, if the crewmembers perform a unique set of tasks in one segment than another segment, it is reasonable to assume that their workload will be very different from one segment to another. For example, in mission segments during which the tank is stationary, the driver could engage a target. In contrast, when the vehicle is moving, the driver is driving and would not be engaging targets. The segments the ARL modelers selected to represent diverse sets of crewmember tasks for the movement-to-contact mission were: movement to contact begins, move via checkpoints to the line-of-departure, precision engagement, and move to defensive position.

As they begin the movement-to-contact mission, the goal of the crewmembers is to be ready for the mission. They perform workstation and communications equipment set-up. As they move via checkpoints, their goals shift to searching for potential enemy. They communicate, drive, search for threats, track the battle and do hasty planning. After the enemy is detected, their goals shift again to destroying the threat. They identify, engage, and destroy the threat. Finally, after the enemy is eliminated, their goals shift to avoiding detection by opposing forces. They back-up the vehicle and drive quickly to a defensive position while avoiding enemy detection.

For each of these segments, the ARL modelers set a unique workload threshold. Each threshold was unique because of the variation in tasks, and,

therefore, workload in each mission segment. Furthermore, each crewmember needed a unique workload threshold because the crewmembers performed very different combinations of tasks. For example, the PL does tactical planning, communications monitoring, and supervisory tasks while the driver drives the vehicle. They obtained the threshold values for each crewmember for each segment from an existing baseline IMPRINT tank model. Mitchell (2009) describes this model and the steps the ARL analyst followed in its development in detail.

After developing the mission segments and selecting thresholds, the ARL analysts included the segments in their task-network models. In the IMPRINT task- network models, the ARL modelers represented the sets of tasks the crewmembers performed in each segment of the scenario as functions. Driving, scanning for threats, and communications, for example, would be functions in the model. Furthermore, the task-network model is hierarchical which means functions, at the higher level, can be decomposed into smaller units called tasks. Thus, the ARL modelers decomposed the functions in each segment into tasks. Examples of tasks for the driving function would be maintain speed, adjust steering, monitor forward terrain, etc.

After creating the hierarchical task-network of functions and tasks for each crewmember in each segment of the scenario, the ARL modelers identified the interfaces or equipment the crewmembers used to perform the tasks. IMPRINT provides modelers with the capability to enter the list of interfaces used by the human system operators for each task. Thus the ARL modelers entered the list of interfaces each crewmember used for each task into the baseline tank model. Then, using the IMPRINT workload scales, the modelers estimated the demands that each task and interface combination placed upon the each crewmember's mental resources (visual, auditory, cognitive, gross motor, fine motor, speech, or tactile).

Once the workload data was entered, the ARL modelers ran the baseline tank model multiple times. The multiple runs represented all the possible combinations of functions and tasks that the crewmembers performed during each segment of the mission. Based on these runs the modelers then identified for each crewmember in each mission segment, the combination of tasks that had the highest overall workload value. In addition,

they calculated the average workload across all the runs for each crewmember for each segment. The maximum workload value and average workload value became the workload threshold for that crewmember for that mission segment for the baseline model.

The ARL modelers then modified the baseline model to represent the crewmembers performing the tasks with the two proposed technologies. Specifically, they modified the interfaces used by two of the tank crewmembers, the driver and the loader. Because the interfaces for these two crewmembers were modified from the baseline, the ARL modelers needed to modify the tasks these two crewmembers performed. For example, in the baseline model, a crewmember needed to open the hatch to do a specific task while in the modified model the loader's display enabled the loader to perform with a closed hatch.

When the modified model was complete, the ARL modelers ran it multiple times and calculated the same workload measures as they had for the baseline model. They then compared the two models to see if the crew workload in the modified tank model was higher than the threshold value established from the baseline model. If the workload was the same or lower, they recommended the technologies for further testing. If it was higher than the baseline they recommended evaluating if the potential for overload was mitigated by an increase in performance.

In addition, to the workload comparison, the ARL modelers compared the performance of the crewmembers in the two models. For example, the loader's workload may have remained the same for both models but the technology may have increased his performance by permitting him to do surveillance buttoned-up rather than out-of-the-hatch. Furthermore, a crewmember performing with an open hatch is at a greater risk of injury than with a closed hatch. Greater risk of crew injury, in turn, for represent a great risk to crew survivability and, therefore, the overall movement-to-contact mission.

The overall conclusion of the analysis was that the new technologies did have the potential to increase mission performance while reducing crew workload.

3.0 DISCUSSION

The ARL modelers found the practical approach to establishing the threshold more effective for justifying their recommendations for system design changes than other threshold techniques they had used over the past decade. In earlier efforts, the modelers had used a single overall workload value of 60 (Mitchell, Samms, Henthorn & Wojciechowski, 2003) or 40 (Mitchell, 2005) or 7 (McCracken & Aldrich, 1984) as the workload threshold value. Although these projects changed system requirements (Mitchell & Samms, 2009), and the results were replicated in experiments (Chen, 2009), the selection of a single workload threshold for all crewmembers across a scenario was challenging for the ARL modelers to defend. Because the threshold was difficult to defend, it made it difficult to convince system engineers to change designs. The single threshold value was difficult to defend because the overall workload values from IMPRINT could vary widely between crewmembers due to variations in the functions and tasks they perform. The driver of the tank, for example, might have a maximum workload value of 200, in contrast to the loader who has a maximum workload value of 60. Thus, with a single threshold value of 40, both crewmembers would be overloaded in the baseline but one would have a much higher workload value than the other. In this situation, the crewmember with the workload that exceeded threshold by the most would be most likely to be the focus for system design changes. In comparison, by identifying a threshold for each crewmember, the ARL modelers had more capability to focus attention equally across crewmembers and influence system design changes for each crewmember.

Another challenge confronting the ARL modelers was that the functions, tasks, and workload that a single crewmember performed changed significantly from segment to segment. For example, the highest workload value in an IMPRINT model for a tank driver within a mission might be 200 and the average across the mission might be 100. This high workload is associated with the driving function and tasks. The workload, therefore, would not be representative of the driver's workload when the platform is stationary. During this mission segment, the driver's workload would be 31, a much lower value because the driver is not driving but is scanning for threats. If the mission were not divided into segments this difference in workload would not be apparent

because average workload across the mission would be 115.5 and high workload 200. The practical threshold approach solves this problem by divided the operational scenario into segments representing changes in functions and tasks for crewmembers and the associated workload value changes.

4.0 CONCLUSION

ARL modelers recommend the practical approach to setting a workload threshold be used to evaluate system designs. Although they implemented their approach with the human performance modeling tool, the practical threshold approach can be applied to any workload analysis. Researchers conducting experiments or field tests can use the same approach. To implement the approach in an experiment, the researcher would create two experimental conditions: a baseline and a design alternative. Next they would identify a scenario which includes all the goals of the participants with the system. They would divide this scenario into the segments that represent these goals. They would then have the test participants perform all their representative concurrent interactions with the system in each segment. They would collect workload values during both the control and alternative design condition for each segment and compare workload and performance of the participants in the two conditions. They would then make recommendations based on the workload comparisons.

As a result of this analysis, two enhancements to IMPRINT were recommended. When the ARL modelers analyzed the results across each mission segment, they used the Function Performance report. The Function Performance report provides analysts with detailed information on function duration, accuracy and frequency. This report is generated by looking at all the functions in the model that have started and finished during the model execution but does not report instances where functions are stopped or interrupted. The same is true for the Task Performance report that reports similar information but at the task level. Expanding these reports to include data about function or task stops and interrupts will provide more detailed results to the analyst.

Another recommended enhancement was to allow analyst to choose at what level they would like to define workload thresholds; at the function or mission segments level or at the overall mission

level. Currently, workload thresholds are set per operator over the length of the entire mission. There may be times where different segments of a mission may have different workload thresholds. Implementing this capability in IMPRINT would allow the analyst more flexibility in exploring new workload theories. These enhancements will be considered for implementation in the next IMPRINT development cycle.

5.0   REFERENCES

[1].   Chen, J. Y. and Joyner, C. T. (2009). *Concurrent Performance of Gunner's and Robotics Operator's Tasks in a Multitasking Environment.* Military Psychology, 21(1), 98-113.

[2].   Colle, H. A. & Reid, G. B. (2005). *Estimating a mental workload redline in a simulated air-to-ground combat mission.* The International Journal of Aviation Psychology, *15* (4), 303-319.

[3].   Directorate of Training, Doctrine, and Combat Development. (2007) *Tank Platoon (FM 3-20.15)*, 204 1st Cavalry Regiment Rd, Ste. 207, U.S. Army Armor Center, Fort Knox, KY 40121-5123, 2007.

[4].   Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & St. John, M. (2008). *The red-line of workload: theory, research, and design.* Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting, New York, pp. 1204-1208.

[5].   Just, M. A., Carpenter, P. A., Keller, T. A., Emery, L. Zajac. H. & Thulborn, K. R. (2001). *Interdependence of non-overlapping cortical systems in dual cognitive tasks.* NeuroImage, 14, 417–426.

[6].   Little, R. (1993). *Final report for the crew reduction in armored vehicles ergonomic (CRAVES) study (CDRL No. A006).* Human Engineering Laboratory, Aberdeen Proving Ground, MD.

[7].   McCracken, J. H. & Aldrich, T. B. (1984). *Analyses of Selected LHX Mission Functions: Implications for Operator Workload and System Automation Goals, (ARI479-024-84).* Army Research Institute, Fort Rucker, AL.

[8].   Mitchell, D. K. (in press). *Abrams V2 SEP Crew Workload Analysis: Impacts of Two Proposed Technologies.* U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

[9].   Mitchell, D. K. (2009). *Workload Analysis of the Crew of the Abrams V2 SEP: Phase I Baseline IMPRINT Model (ARL-TR-5028).* Army Research Laboratory, Aberdeen Proving Ground, MD.

[10].   Mitchell, D. K. (2005). *Soldier Workload Analysis of the MCS Platoon's Use of Unmanned Assets (ARL-TR-3476).* U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

[11].   Mitchell, D. K. (2000). *Mental Workload and ARL Workload Modeling Tools (ARL-TN-161).* U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

[12].   Mitchell, D. K. & Samms, C. L. (2009). *Workload Warriors: Lessons Learned from a Decade of Mental Workload Prediction Using Human Performance Modeling,* Human Factors and Ergonomics Society 53nd Annual Meeting Proceedings, 53(12), 819-823.

[13].   Mitchell, D. K., Samms, C. L., Henthorn, T., & Wojciechowski, J.   (September, 2003). *Trade Study: A Two- Versus Three-Soldier Crew for the Mounted Combat System (MCS) and Other Future Combat System Platforms* (ARL-TR-3026). U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

[14].   Reid, G. B. & Colle H. A. (1988). *Critical SWAT values for predicting operator overload.* Proceedings of the Human Factors Society – 32nd Annual Meeting. Santa Monica, CA: Human Factors Society, 1414-1418.

[15].   Rueb, J.D., Vidulich, M.A. & Hassoun, J.A. (1994). *Use of workload redlines: A KC-135 crew-reduction evaluation.* International Journal of Aviation Psychology, 4, 47-64.

[16].   Schlegel, B., Schlegel, R. E., & Gilliland, K. (1988). *Determining excessive mental workload with the subjective workload assessment technique.* Proceedings of the 10th Congress of the International Ergonomics Association (Vol. II), Sydney Australia, 475-477.

[17].   Wickens, C. (2008). *Multiple Resources and Mental Workload.* The Journal of the Human Factors and Ergonomics Society, 50(3), 449-455.

[18].   Yerkes, R. M. & Dodson, J. D. (1908). *The relation of strength of stimulus to rapidity of habit-formation.* Journal of Comparative Neurology and Psychology, 18, 459-482.

MODSIM WORLD
Conference & Expo

Predicting the Consequences of Workload Management Strategies with Human Performance Modeling

Charneta Samms
Diane Mitchell
U.S. Army Research Laboratory

October 13-15, 2010

MODSIM WORLD
Conference & Expo

Agenda

- Mental workload and human performance
- Human Performance Modeling (HPM)
- Improved Performance Research Integration Tool (IMPRINT)
- Analysis Approach
- Case Study
- Conclusions

Mental Workload and Human Performance

Inverted-U relationship between workload and performance

Modified from Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation, Journal of Comparative Neurology and Psychology, 18, 459-482.

Importance of Workload

- Indicator of problem areas within system design
- Peaks and valleys of workload indicate times when human performance may suffer, e.g.:
  - Sustained low workload (underload) leads to boredom, loss of situation awareness, and reduced alertness.
  - Sustained high workload (overload) leads to fatigue.
  - Workload peaks lead to dropped tasks, increased task time, cognitive tunneling, and increased errors.
- Reduces crew performance, system performance, and contribute to mission failure

  OBJECTIVE: Achieve evenly distributed, manageable workload. Avoid both overload and underload.

4

Why Human Performance Modeling (HPM)?

Concept System

Many Variables

Field Study Not Feasible

Too Dangerous

Model – Test – Model

System Performance $\cong$ f**(human performance)**

5

Improved Performance Research Integration Tool

IMPRINT
Improved Performance Research Integration Tool

334 users supporting Army, Navy, Air Force, Marines, NASA, Department of Homeland Security (DHS), Department of Transportation (DoT), Joint and other organizations across the country

http://www.arl.army.mil/IMPRINT

709

IMPRINT can be used to

- Set realistic system requirements
- Identify future manpower & personnel constraints
- Evaluate operator & crew workload
- Test alternate system-crew function allocations
- Assess required maintenance man-hours
- Assess performance during extreme conditions

- Examine performance as a function of personnel characteristics and training frequency & recency
- Identify areas to focus test and evaluation resources
- Quantify human system integration risks in mission performance terms to support milestone review
- Represent humans in federated simulations

IMPRINT is a trade-off analysis tool

Multiple Resource Theory (MRT) in IMPRINT

| Mission Tasks | → | Which Brain Resources Involved? | → | Degree of Resource Use? |

1. monitor alarms
2. decide response action
3. pull trigger
.
.
.
n. task n

Visual
Cognitive
Auditory
Motor
Speech

Speech
Visual
Auditory
Motor

Cognitive

0.0	No Cognitive Activity
1.0	Automatic (simple association)
1.2	Alternative Selection
3.7	Sign/Signal Recognition
4.6	Evaluation/Judgment (consider single aspect)
5.3	Encoding/Decoding, Recall
6.8	Evaluation/Judgment (consider several aspects)
7.0	Estimation, Calculation, Conversion

8

710

Analysis Approach

Quantify influence of human operator performance on system/mission performance

Soldier performance includes mission analysis   Mission relevant performance parameters

Conduct Domain Analysis → **Develop Question** → **Perform Function Analysis**

Literature Review & Knowledge Elicitation Sessions with SMEs

Include key performance parameters

NOT THE FIRST STEP!!!!

Refine Question → **Develop Experimental Design Matrix** → **Build Models**

Run Analysis or Experiment → **Summarize Results** → **Make Recommendations**

Executing the Approach

Build Models
- Create a scenario with segments representing events that change the goals of system operators
- Establish baseline and alternative system design
- Select unique workload threshold values for each operator

Run Analysis
- Compare workload results across conditions
- Higher workload than baseline = performance decrements

Case Study

Examine impact of two conceptual technologies on workload and system performance

BASELINE model	ALTERNATIVE model
Without technologies	*With technologies*

Movement to Contact
Seek, identify and eliminate potential threats
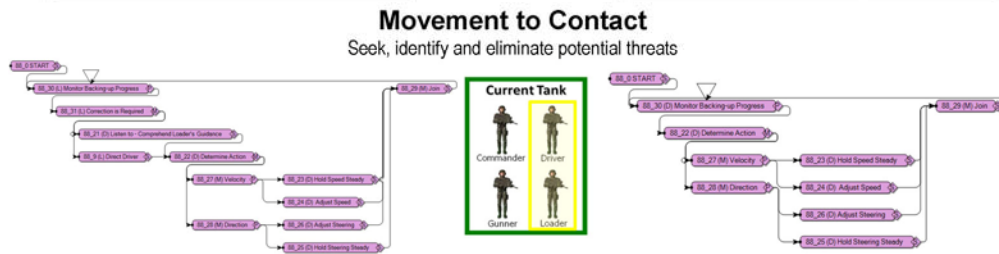
New technologies have potential to increase mission performance while reducing crew workload

Mitchell, D. K. (in press). Abrams V2 SEP Crew Workload Analysis: Impacts of Two Proposed Technologies. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

Summary

- Use analysis approach to setting workload thresholds in HPM or experimentation
  - Develop overarching scenario
  - Set up at least two conditions; e.g. baseline & alternative
  - Compare workload levels
  - Make recommendations based on workload comparisons
- Potential enhancements for IMPRINT
  - Expansion of function & task performance reports
  - Function level workload thresholds

5.12 Simulating Visual Attention Allocation of Pilots in an Advanced Cockpit Environment

Simulating Visual Attention Allocation of Pilots in an Advanced Cockpit Environment

F. Frische & J.-P. Osterloh & A. Lüdtke
OFFIS Institute for Information Technology
frische@offis.de osterloh@offis.de luedtke@offis.de

This paper describes the results of experiments conducted with human line pilots and a cognitive pilot model during interaction with a new 4D Flight Management System (FMS). The aim of these experiments was to gather human pilot behavior data in order to calibrate the behavior of the model. Human behavior is mainly triggered by visual perception. Thus, the main aspect was to setup a profile of human pilots' visual attention allocation in a cockpit environment containing the new FMS. We first performed statistical analyses of eye tracker data and then compared our results to common results of familiar analyses in standard cockpit environments. The comparison has shown a significant influence of the new system on the visual performance of human pilots. Further on, analyses of the pilot models' visual performance have been performed. A comparison to human pilots' visual performance revealed important improvement potentials.

1.0 INTRODUCTION

The European project HUMAN (EC's 7th Framework Programme) aims at developing virtual test pilots, in order to improve the human error analysis of future cockpit systems in early design phases, as a supplement of simulator tests with human pilots in later design phases. In HUMAN, a 4D Flight Management System (Advanced Flight Management System, AFMS) and its user interface (Airborne Human Machine Interface, AHMI), developed at the German Aerospace Center (DLR Braunschweig), have been selected as systems under investigation. The virtual test pilots are instances of a cognitive architecture named CASCaS (Cognitive Architecture for Safety Critical Task Simulation, see [12]). Cognitive architectures, such as ACT-R (see [1]), SOAR (see [10]), MIDAS (see [4]) and CASCaS implement cognitive plausible theories for human perception, memory operations and decision making. These theories are independent of specific human-machine interfaces. Thus, cognitive architectures are applicable not only in the aviation domain, but also in the automotive or maritime domain. Perception of system and environmental states – or of entities in the real world in general – is a key factor for situation awareness and for decision making [6]. The main channels for perception on human-machine interfaces are primarily eyes for visual perception and secondarily ears for auditory perception. A third upcoming channel is the skin for tactile interfaces, but this is – to our knowledge – currently not implemented in any of the cognitive architectures mentioned before. Due to the importance of visual perception for human-machine interaction, and for situation awareness and decision making, there is a need for an accurate simulation of visual performance in cognitive architectures.

Introduction of new user interfaces, e.g. into common cockpit setups, has influence on the visual attention allocation. Examples for this effect can be found in [7]. This could be explained by the following two points: On the one hand, new interfaces can trigger attention bottom-up, meaning that the interface presents information in a very dominant way which distracts visual attention from other interfaces. This is often referred to as selective attention, where eye movements and shifts of attention are triggered by the onset of a salient stimulus [16]. On the other hand, attention allocation can be affected top-down because the new interface provides new functionality or displays redundant information in a more accessible or usable way than other interfaces do. Top-down attention is caused primarily by underlying task models that comprise the allocation of visual attention.

Thus, a cognitive architecture that should simulate visual attention allocation humanlike requires both, valid cognitive theories for bottom-up attention and a valid task model embedding tasks on the new interface into the common task model for top-down attention.

In this paper we will present results of experiments with human line pilots and a pilot model interacting in a cockpit environment containing the AHMI. Although the datasets have also been used to validate the cognitive theories for visual attention allocation implemented in the model, the main focus of this paper is the validation of our task model for scanning activities in the new cockpit setup.

In the following section we describe top-down and bottom-up concepts for visual attention implemented in CASCaS (section 2). Then, the experiments conducted (section 3) and the results of these experiments are presented (section 4). The paper closes with a short discussion (section 5) and conclusions (section 6).

2.0 MODELING VISUAL ATTENTION
Visual attention allocation is a complex conglomerate of top-down (active) and bottom-up (reactive) processes triggering percept actions. Top-down and bottom-up attention compete against each other [3], e.g. a salient stimulus might distract pilots from tasks which they are focused on. This is often intended, e.g. in case of warnings. However, a salient stimulus might go undetected, because top-down attention causes the eyes to move to an area of interest where the stimulus is either out of the visual field or absorbed by a dynamic neighborhood. The cognitive architecture CASCaS implements both processes. In the following subsections we will describe how top-down and bottom-up processes have been implemented.

2.1 Top-Down Attention
The top-down attention is driven via three different levels of consciousness (see

Fig.1), which are based on Anderson's three layers of consciousness named autonomous layer, associative layer and cognitive layer [2]. This is also in line with Rasmussen, who defined three levels of behavior, called skill-based, rule-based and knowledge-based [13]. While nearly zero consciousness is needed on the autonomous layer, almost full consciousness is needed on the cognitive layer, where decision making, planning and problem solving are located.
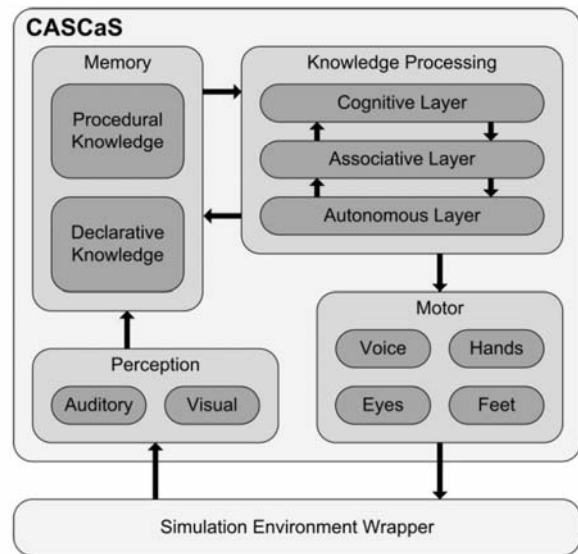
Fig. 1: The multi-layered architecture of CASCaS consists of components for perception, memory, knowledge processing and motor actions.

Top-down processes on the associative layer are the main driving factor for visual attention allocation of pilots, where they perform well-learned rules to achieve specific goals. These rules describe normative procedures – percept and motor actions that match correctly specific situations. With regard to visual attention of pilots we differentiate between two types of procedures: (1) scanning procedures and (2) interaction procedures. Scanning procedures only contain percept actions. Pilots regularly perform scanning of multiple aircraft and environment parameters in order to keep situation awareness for current and future aircraft states. These

scanning activities are the main driving factor for visual attention in our pilot model. Interaction procedures contain percept and motor actions. They are used to interact with interfaces in the aircraft, such as the AHMI. Percept actions are needed in order to assess current situations and because we assume that pilots look at buttons before they press them.

In CASCaS, normative procedures are described by formal rules. The rule format is a Goal-State-Means (GSM) format (see Fig. 2). All rules consist of a left-hand side (LHS) and a right hand side (RHS). The left-hand side contains a goal in the Goal-Part and a State-Part specifying boolean conditions on the current state of the environment in the memory. Apart from the condition the State-Part contains memory-read operators to specify that, in order to evaluate a condition, the associated values v_i of interaction elements i_j have to be retrieved from memory. The right-hand side consists of a Means-Part containing motor and percept operators (writing values and reading values in the simulated environment), memory-store operators as well as a set of partially ordered sub-goals.
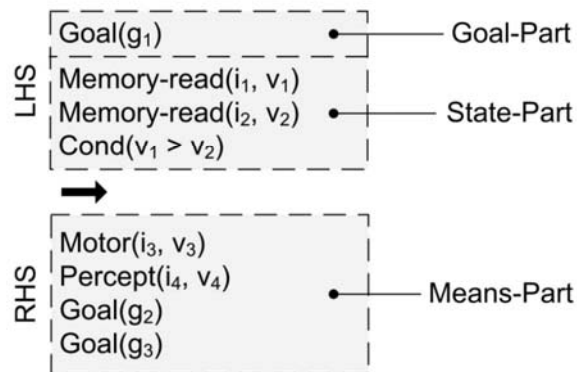
Fig. 2: Procedural knowledge is described is a specific rule format that consists a certain goal in a Goal-part, a State-part and a Means-part

During simulation the cognitive architecture selects rules based on their left-hand sides and executes the right-hand sides.

Rules are connected by a goal on the left-hand side and goals on the right-hand side. This allows us to break down complex procedures into a hierarchical ordered structure, similar to hierarchical task trees with a small difference: Our rule-based description of procedures permits transitions from lower levels to higher levels (see Fig. 3). Each procedure consists of $1 \dots n$ goals. Each goal is a unique entity that is allocated to $1 \dots m$ rules but each rule is allocated to exactly one goal.
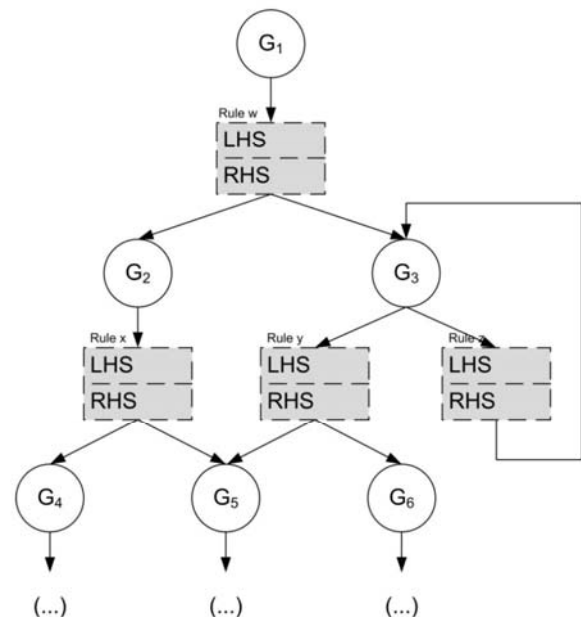
Fig. 3: Rules are connected by goals on the LHS and RHS

2.2 Bottom-Up Attention

Bottom-up processes are unconscious and triggered by the perceptual component of CASCaS. The main driving factor for bottom-up attention in CASCaS is a theory called selective attention. Selective attention is an effect where salient objects, e.g. flashing lights, moving objects, or high contrasts, cause an automatic shift of attention towards this object [16]. Attention shifts can also be triggered by acoustic and tactile stimuli, which are not investigated in this paper. In terms of visual stimuli, a salient stimulus means a discontinuity in space or time in the visual field. A

discontinuity in space represents a difference in a static property, like color, brightness, form or orientation. This could be for example a green dot in a set of red dots. In contrast to this, a discontinuity in time – or dynamic discontinuity – denotes a dynamic change, like abrupt onset, flashing or moving of an object. This effect may be restrained by the top-down process or by the saliency of other objects nearby, which suppress, with their own high saliency, other salient objects.

Bottom-up attention can trigger specific procedures on the associative layer, e.g. in case of a flashing emergency light the attention of pilots should be shifted to the flashing light which is followed by execution of a procedure to handle the emergency.

3.0 EXPERIMENTS
In order to validate the visual performance of the model, experiments have been conducted with human subject pilots and with CASCaS in a functionally equivalent simulation environment. In the following sections, we will describe how the experiments with the human pilots have been carried out.

3.1 Target System AHMI
The main objective of our analysis is the interaction between the pilot flying (PF) and the AFMS. The AHMI is a graphical user interface supporting interaction between the AFMS and pilots. Both, the AFMS and the AHMI have been developed by the German Aerospace Center (DLR, Braunschweig, Germany). The AHMI supports graphical information about the current positions of the ego-aircraft and other aircrafts, weather conditions and flight routes. It provides a horizontal view (as shown in Fig. 4) and a vertical view. It supports onboard management of flight trajectories and negotiation of trajectory changes with Air Traffic Control (ATC) via Data Link to reduce voice-communication. The AHMI is a powerful tool for pilots, improving predictability of conflicts between aircraft or

between planned routes and severe weather conditions.
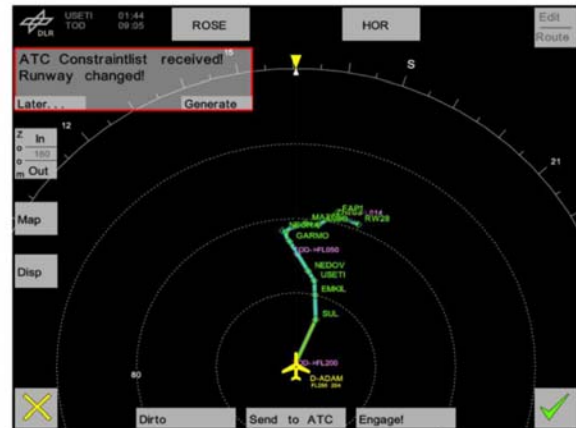
Fig. 4: The AHMI, a graphical user interface supporting interaction between AFMS and pilots

3.2 Flight Simulator Setup
The experiments have been conducted in the GECO (Generic Experimental Cockpit) simulator, which has been built and is maintained by the DLR in Braunschweig. The layout of the simulator has been derived from the Airbus A350 XWB aircraft. It is equipped with freely programmable wide-screen LCD displays and modern input devices like side sticks and a Keyboard Cursor Control Unit (KCCU), as used in the A380. The flight dynamics are derived from a VFW 614 (ATTAS), as used by the DLR as a test aircraft. The outside view is generated via three video projectors on a spherical screen with a diameter of 6 meters, providing highly realistic outside view. The GECO is a fixed-based flight simulator equipped with a visual head tracker (AR-tracking), and an iView-X eye-tracker system from SMI. Eye-tracker data has been matched on specific regions representing areas of interest (AOI) where visual attention allocation should be analyzed. These AOIs were the following:

- Airborne Human Machine Interface (AHMI)

- Primary Flight Display (PFD)

716

- Horizontal Situation Indicator (HSI)

- Engine Display (ENG)

- Flight Command Unit (FCU)

- Gears and Auto Break (GAB)

- Outside view (Windows)

In addition, pilot voices and all flight parameters have been recorded.

3.3 Scenarios
In order to analyze pilot behavior, 8 scenarios have been defined, containing different AHMI-related tasks. The scenario that we refer to in this paper contained 3 events that pilots had to handle. These events triggered pilots to perform re-planning of their current flight plan according to requirements sent by ATC. A flight plan is a list of waypoints the aircraft has to fly over or fly by. The scenario was divided into three phases: cruise, approach and landing. Communication between pilots and ATC has been restricted to non-auditory communication via the AHMI which allowed uplinks or downlinks of flight plans.

3.4 Participants
The experiments have been conducted with 13 male and 2 female German line pilots recruited from German airlines. None of the pilots has been experienced in the usage of the AHMI, and only some have been in the GECO before. All subjects participated as the pilot flying (PF). The crew was completed by a scripted pilot, who acted as a pilot monitoring (PM). Scripted PMs were a male DLR test pilot or a female first officer from Lufthansa. In addition to the normal duties of the PM, the scripted pilot was responsible for the training and supported the debriefing and analysis by taking notes during the flight.

3.5 Procedure
The experiments were distributed over two days. The first day started with a general briefing on the project. Afterwards training on the AHMI and the GECO has been

performed by the PM. After the pilots felt familiar with the tasks and the simulator, a talk-through was performed, in order to verify that the procedures where well-trained. After the talk-through was performed successfully, the subjects started to fly the first scenario. Typically, 2 scenarios were finished on the first day and 5 to 6 scenarios on the second day.

4.0 RESULTS
In this section we present results of analyses regarding top-down visual performance of human pilots and of our pilot model in a cockpit setup containing the AHMI. The analysis is based (1) on eye-tracker data, which have been recorded during the experiments with human pilots and (2) on log files for the pilot model. The output of both data sources has been pre-processed into a comparable format containing timestamps $t_{1...n}$ and AOIs $aoi_{1...m}$ describing where pilots have looked at a specific time. Each t_i in the datasets is associated with exactly one aoi_j. The experimental cockpit has been divided into 7 AOIs (see section 3.2) in order to analyze the gaze distribution. However, the results presented in this paper focus on 4 AOIs (AHMI, PFD, HSI, windows) which have been selected after a first review of the data for the following reasons: AHMI, PFD and HSI are the main displays for monitoring aircraft and environmental states in our scenarios during all flight phases. The windows are very important for perception of the outside world during the landing. We segmented the datasets according to 3 flight phases (cruise, approach, landing) and calculated the percent dwell times (PDT) for each phase, respectively. PDT is a format representing the dwell time spent on a specific AOI in relation to the sum of dwell time spent on all AOIs observed in (%). We analyzed the PDTs on two levels: First, we performed a separate comparison of the results of each phase for the human pilots and for the model. Second, a comparison of human data to model data has been performed.

4.1 Human Performance

The gaze distribution of pilots during flight can be seen as the main indicator of how important specific areas are for flying an aircraft – from a pilot's point of view. Huettig, Anders and Tautz [9] revealed the dominance of the PFD in modern glass cockpits with a value of around 40%. For the HSI a value of around 20% has been measured. This result is in line with results published by Mumaw, Sarter and Wickens (see [11] and [14]), who analyzed the monitoring behavior of pilots on an automated flight deck. They measured 35% on the PFD and 25% on the HSI. Futher on, eye movement analyses with a Boeing 747-400 desktop simulator have been conducted by Dietz et al. (see [5]).

In order to get an overview of our results, Table 1 depicts the average PDTs of our human subjects for each flight phase. Values do not sum up to 100 because dwells on other AOIs are still taken into account but are not displayed.

	Cruise	Approach	Landing
AHMI	60	42	21
PFD	15	28	40
HSI	7	11	12
Windows	6	7	17

Table 1: **Aggregated PDTs of human pilots during flight phases cruise, approach and landing**

In contrast to results mentioned above, our results reveal a dominance of the new introduced AHMI with a value of 60% during cruise phase. The PFD, with a value of 15%, is far behind the AHMI. This emphasizes the role of the AHMI in our scenarios. HSI, with a value of 7%, is behind the PFD, which is in line with results reported in literature. During cruise outside view is not important, thus, with a value of 6%, windows are behind the HSI. From cruise to approach PDT on PFD increases by 13%, while PDT on AHMI decreases by 18%. HSI is also increasing by 4% and windows by 1%. From approach to landing

PDT on PFD is increasing by 12% and PDT on AHMI is decreasing by 21%. PDT on HSI is increasing by 1% and PDT on windows is increasing by 10%. Thus, from approach to landing the rank orders of AHMI and PFD change as well as the orders of HSI and windows. We assume that changes in gaze distribution between different flight phases are caused by different task models for each flight phase. E.g. the high values on windows during landing phase are caused by the upcoming landing task which triggers the pilot to monitor the runway. Low values on the AHMI during landing phase are caused by degradation of the navigation task. These changes are caused by top-down attention as described in section 2.1.

4.2 Model Performance

Results of model performance show, with a value of 65%, a strong dominance of the AHMI during cruise phase. With a value of 31%, the PFD is behind the AHMI. From cruise to approach there is only a small change to 66% on the AHMI. PDT on PFD does not change. From approach to landing rank orders of AHMI and PFD change. PDT on AHMI decreases from 66% to 35% and PDT on PFD increases from 31% to 53%. HSI and windows are at a very low level between 0% and 2% during all phases. All results are presented in Table 2.

	Cruise	Approach	Landing
AHMI	65	66	35
PFD	31	31	53
HSI	0	2	2
Windows	1	1	1

Table 2: **Aggregated PDTs of pilot model during flight phases cruise, approach and landing**

4.3 Model Validation

Human performance data has been used to validate the visual performance of the pilot model based on two dimensions, trend and local fitness, that are often used in the domain of cognitive model validation.

718

4.3.1 Measure of Trend

A trend describes how a dependent variable v_d develops in relation to an independent variable v_i. We have measured the variable gaze distribution (= v_d) in relation to the flight phases (= v_i) for the human pilots and for the pilot model. An aspect of model validity is trend consistency, meaning that the relation between v_d and v_i is the same for the model and for the real world aspect observed. In the area of cognitive model validation, the use of Pearson's correlation coefficient (r and r^2) is a common measure of trend (see e.g. [7] and [15]). Having a look at the performance of the pilot model applying our scanning procedure, it can be seen that it fits the human visual performance rather well with $r^2 = 0.85$.

Figure 5 visualizes trends based on PDTs measured for the human pilots and for the pilot model during the flight phases cruise, approach and landing.
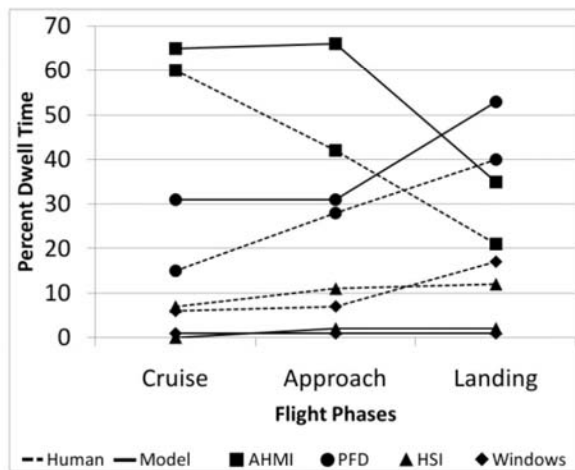

Fig. 5: Comparison of gaze distribution for human pilots and pilot model across flight phases

AHMI and PFD are the most dominant displays during all phases for human pilots and for the pilot model. AHMI and PFD change their ranks order from approach to landing phase. The human data trends for AHMI and PFD between cruise phase (AHMI = 65%; PFD = 31%) and approach phase (AHMI = 66%; PFD = 31%) have not been captured for the model. Indeed, the

trend for AHMI between these phases is slightly contrary to the human findings. The human data trend on HSI has been well captured for the model, where PDT is increasing from cruise (= 0%) to approach (= 2%) and then holding the level from approach to landing (= 2%). The model's PDTs for the windows are linear for all flight phases (= 1%). We had problems modeling this AOI, because dynamic AOIs, such as a runway "moving" on the windows, currently cannot be modeled within the architecture. Thus, we are not able to provide the model with information that is gathered by human pilots when they are looking out of the windows. Nevertheless, during our experiments we implemented some kind of "blind scanning" on the windows in order to simulate transitions between windows and displays. The intention was to model the effect of not looking at displays (for whatever reason) which has been identified as a cause for long reaction times because visual signals such as flashing buttons are not in the visual field (see section 2.2). This may also impact pilots' situation awareness.

4.3.2 Measure of Location

We analyzed the local fitness of gaze distribution by comparing the Root Mean Squared Successive Differences (RMSSD) values of human pilots and the pilot model as presented in [14]. Local fitness measures of model to human data are a bit problematic as trying to optimize local parameters bears the danger of overfitting the model. Instead of fitting the model to a static parameter value, it is more reasonable to fit the model into a range of parameter values. RMSSD can be used to gain insight into the differences of performance between an individual subject s_i and a group g_j of individual subjects $s_{1...n}$. We calculated RMSSD for each of the human subjects, pulling them one at a time, without replacement, from the group. In our case the group contained 10 subject datasets and we tested the fit of s_1 to $s_{2...10}$, then data from s_2 to s_1, $s_{3...10}$ and so on. The results of these measures were 10 values, one for each pilot, describing the deviation

from the performance of the group. This approach has been extensively described in [8]. We also calculated RMSSD for our pilot model by comparing the model dataset to the group of all human subjects $s_{1...10}$. Next, we will focus on results regarding the cruise phase as this is the most important phase for pilot interaction with the AHMI. Results are depicted in Fig. 6. RMSSD values for human subjects range from 5.52 for subject PF_03 to 24.11 for subject PF_09. The RMSSD for the pilot model is 19.21 which is within the range of human subject values. However, a comparison of this value to the median of human pilots' RMSSDs (= 7.63) shows that the model result is closer to the maximum than to median.
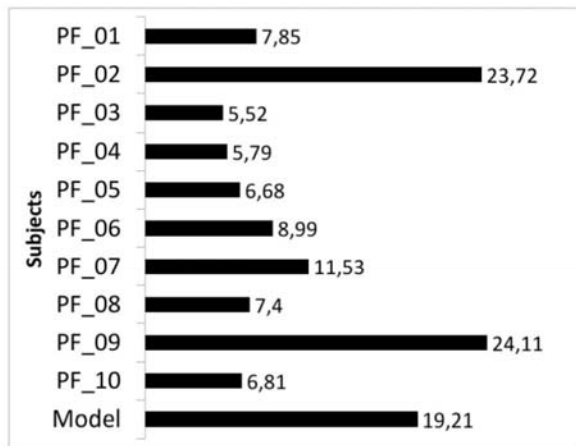

Fig. 6: Comparison of RMSSD values for PDTs of human pilots' and pilot model's gaze distribution in cruise phase

Except for subjects PF_02 and PF_09 all pilots are below a value of 12.0 which shows that these results are outliers in the sample. Analysis of outlier datasets showed that the deviations are caused by differences in PDTs on the AHMI. We have measured 60% mean PDT and 61% median PDT on the AHMI which is a hint on a well-balanced distribution. For PF_02 we have measured 80% PDT (= max) and for PF_09 we have measured 39% PDT (= min). PDTs of PF_02 on other AOIs were much lower, those of PF_09 much higher respectively. An explanation could be that PF_02 used redundant information shown on the AHMI

(such as speed, altitude) for monitoring. Thus, he has implemented the AHMI in his scanning procedure (top-down attention). On the other hand, PF_09 used the AHMI only if he had to react to ATC uplinks (bottom-up attention) instead of including the AHMI into his scanning procedure.

5.0   DISCUSSION

Analysis of visual attention is a useful means for assessment of situation awareness and derivation of task models for scanning activities in cockpit environments. We have modeled scanning procedures for an advanced cockpit environment and performed experiments with a pilot model applying this procedure and with human subject pilots. We used the visual performance data recorded for the human pilots and for the pilot model to validate the visual performance of the model. While Pearson's r and r^2 are useful trend measures, RMSSD can be used to measure the local match between model and human data. Good results for Pearson's r and r^2 are not sufficient to validate a model. A valid model must also perform within the natural range measured for the variable under observation of the human subjects. Comparing our result for the trend measure between human pilots' and model gaze distribution with the results of local fitness, we derive the following: As the trend measure between model and human performance revealed good fitness, we assume that we have a rather good assumption of how important specific AOIs are for the pilots relatively to the flight phases. As the gaze distribution is a good indicator for the correctness of the scanning tasks in the different flight phases, we also assume that we have a correct understanding of the importance of specific scanning tasks performed in these flight phases. However, RMSSD revealed that the performance of the model is at the upper bound of human subjects' performance. This can be improved by decreasing gaze on AHMI and PFD, and increasing gaze at least for the HSI, which has not been modeled sufficiently. Gaze on the windows

has not been modeled adequately. It has to be discussed if it is reasonable to put attention on an area, whose functionality cannot be simulated, only to provoke effects related to bottom-up attention. Alternatively, only flight in cruise phase could be modeled, which has shown to be the most relevant flight phase for AHMI interaction.

6.0 CONCLUSIONS

In this paper we have presented results concerning the visual attention allocation of human pilots and of a pilot model in an advanced cockpit environment. We have been able to show that the AHMI, a new interface for aircraft navigation, has a strong influence on the gaze distribution of pilots due to task models underlying the flight phases. Tasks (especially scanning activities) have been modeled in a rule based language. These rules have been applied by our pilot model as procedural knowledge during the flight phases. Analyses of human pilot performance and model performance in the dimensions of trend and local fitness revealed that there is still some potential left for improving the scanning behavior of the model. An open question is if it is useful to model "blind scanning" on AOIs whose functionality cannot be simulated in order to provoke effects related to bottom-up attention.

7.0 REFERENCES

1. **Anderson, J. R., et al.** *An Integrated Theory of Mind.* Carnegie Mellon University, Pittsburgh and Rice University, Houston : s.n., 2004.

2. **Anderson, J. R.** *Learning and Memory.* s.l. : John Wiley & Sons, Inc., 2000.

3. *Visual Attention: Bottom-Up Versus Top-Down.* **Connor, C. E., Egeth, H. E. and Yantis, S.** 2004, Current Biology, 14, pp. 850-852.

4. **Corker, K. M.** Cognitive Models and Control: Human and System Dynamics in Advanced Airspace Operations. *Cognitive Engineering in the Aviation Domain.* Mahwah, New Jersey, USA : N.B. Sarter, R. Amalberti, 2000.

5. **Diez, M., et al.** *Tracking Pilot Interactions with Flight Management Systems through Eye Movements.* Columbus, Ohio, USA : s.n., 2001.

6. **Endsley, M. R.** Measurement of Situation Awareness in Dynamic Systems. *Human Factors, 37(1).* 1995, pp. 65-84.

7. **Foyle, D. C., et al.** *Human Performance Models of Pilot Behavior.* 2005. Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting.

8. **Gluck, K. A., et al.** *A Computational Model of Basic Aircraft Maneuvering.* Bamberg : s.n., 2003. Proceedings of ICCM.

9. **Huettig, G., Anders, G. and Tautz, A.** *Mode Awareness in a Modern Glass Cockpit - Attention Allocation to Mode Information.* Columbus, Ohio State University, USA : s.n., 1999. Proceedings of the 10th International Symposium on Aviation Psychology.

10. **Lewis, R. L.** *International Encyclopedia of the Social and Behavioral Sciences.* Amsterdam : Pergamon, 2001.

11. **Mumaw, R. J., Sarter, N. B. and Wickens, C. D.** *Analysis of Pilots' Monitoring and Performance on an Automated Flight Deck.* Columbus, Ohio State University : s.n., 2001. International Symposium on Aviation Psychology.

12. **Osterloh, J. P. and Lüdtke, A.** *Analyzing the Ergonomics of Aircraft Cockpits Using Cognitive Models.* Las Vegas, Nevada, USA : Proceedings of Advanced Human Factors and Ergonomics International, 2008.

13. **Rasmussen, J.** *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering.* New York, NY, USA : Elsevier Science, 1986.

14. **Sarter, N. B., Mumaw, R. J. and Wickens, C. D.** Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data. *Human Factors: The Journal of Human Factors and Ergonomics Society.* 2007.

15. **Schunn, C. and Wallach, D.** *Evaluating Goodness-of-Fit in Comparison of Model to Data.* 2004.

16. **Yantis, S. and Jonides, J.** Abrupt Visual Onsets and Selective Attention: Voluntary Versus Automatic Allocation. *Journal of Experimental Psychology: Human Perception and Performance.* 1990, pp. 121-134.

8.0 ACKNOWLEDGEMENTS

Simulating Visual Attention Allocation of Pilots in an Advanced Cockpit Environment

F. Frische, J.-P. Osterloh, A. Lüdtke

OFFIS – Human Centered Design Group

OFFIS
INSTITUTE FOR
INFORMATION TECHNOLOGY

October 14, 2010

Motivation
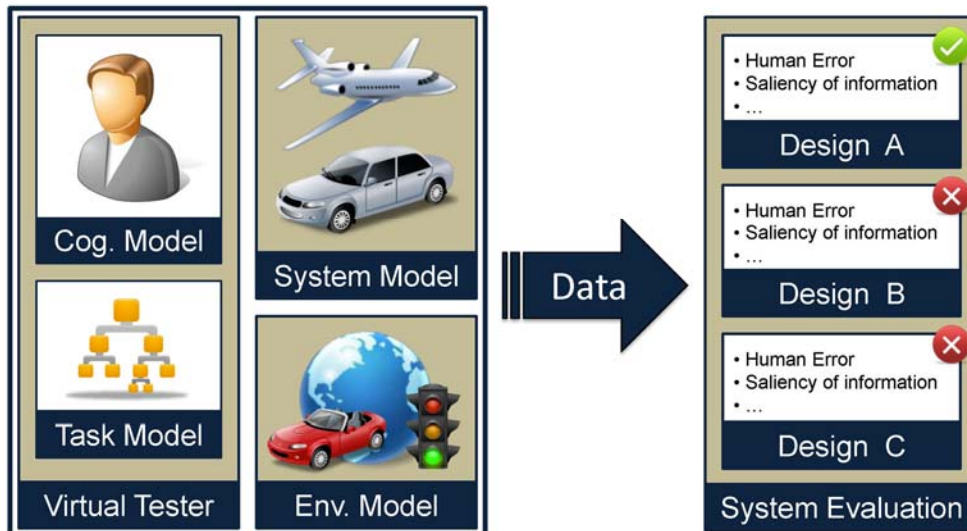
Support evaluation of system designs in early phases of system development process

The Role of Visual Perception for Modeling Pilot Behavior

- Select Strategy (Rules)
- Evaluate situation (Parameters)
- Decide what to do next (Goals)

The Role of Visual Perception for Modeling Pilot Behavior
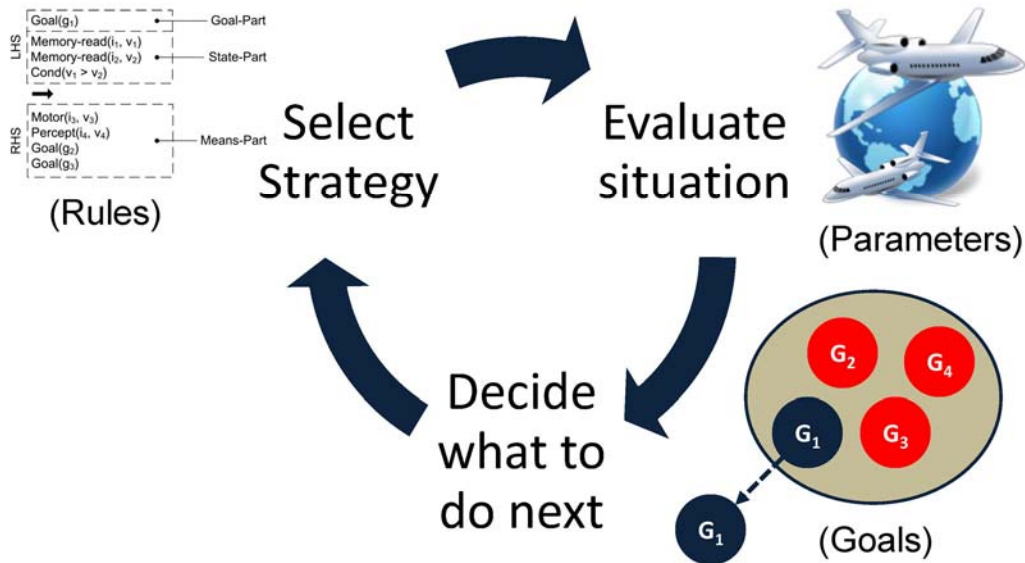
1. Evaluation of situation depends on parameter values
2. Parameters have to be percepted from the environment
3. Pilots percept environment mainly via visual channel
4. No valid model of visual perception → No valid model of pilot behavior
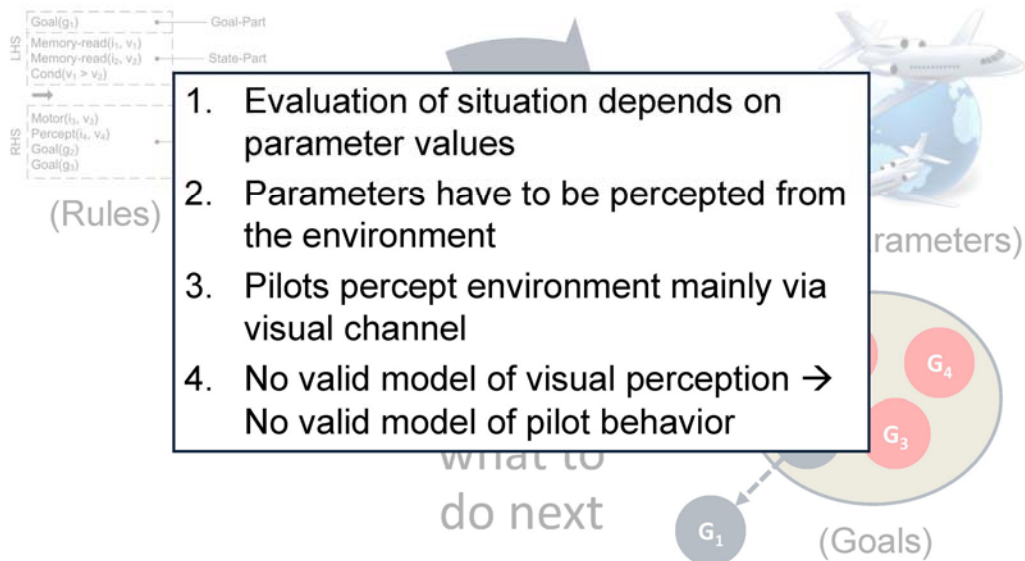
Visual Attention: Top-Down

- Active scanning behavior
- Depends on context of situation
- Abstraction of SEEV Model (Saliency, Expectancy, Effort, Value)
- Probability value for each AOI

Visual Attention: Top-Down
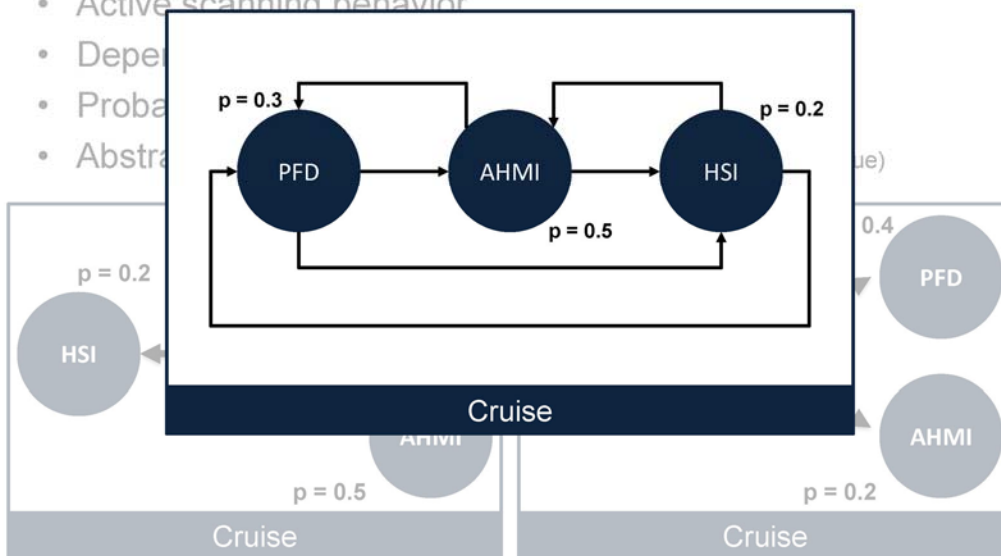
Improved Top-Down Attention

- Pilots tend to optimize scanning behavior
- Probabilities on transitions between AOIs
- Different probability values for each transition
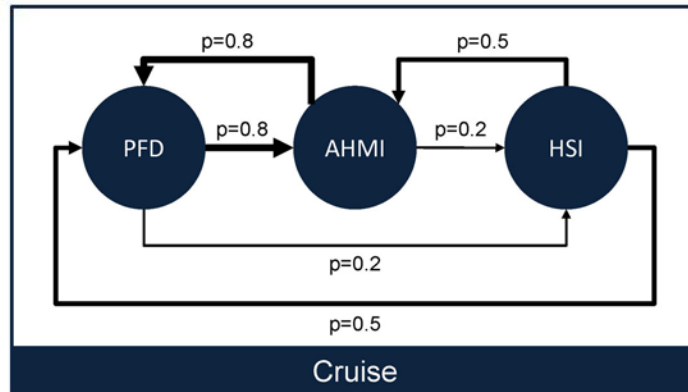
Visual Attention: Bottom-Up

- Reactive scanning behavior
- Depends on saliency of objects in visual field
- SEEV Model → Saliency
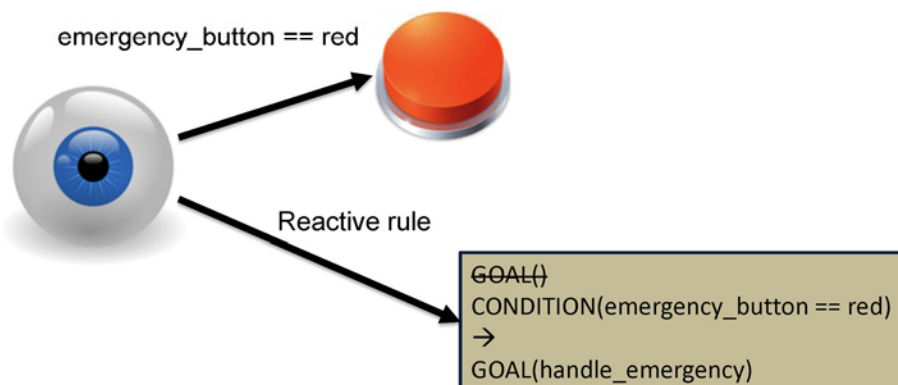
Visual Attention Model

- Consists of
  - Top-down attention (active)
  - Bottom-up attention (reactive)
- Visual attention is mainly influenced by top-down attention
  - Considers context of different situations
  - Supports modeling of human optimization strategies

Experimental Setup

- Scenario duration: ~35 minutes
- 15 Airline pilots
  - 13 male, 2 female
  - Average Age: 34.0 (SD 5.9)
- Events triggering re-planning on AHMI
- Three flight phases:

Cruise   Approach   Landing

Target System: AHMI

- **A**irborne **H**uman **M**achine **I**nterface
- Data link communication between Crew and ATC
- Negotiation of 4D flight plans and trajectories
- View on ego-aircraft
  - Horizontal
  - Vertical

Simulator Layout and AOIs
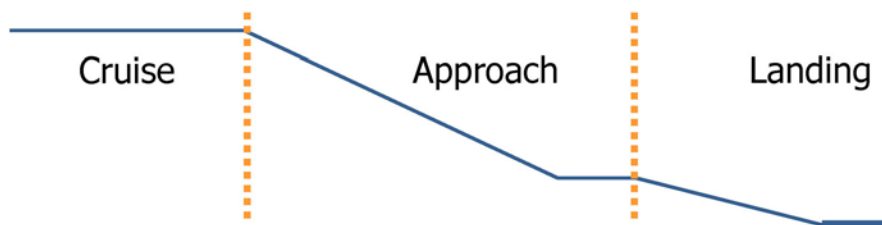
Results

- Focus of pilots' attention is mainly on AHMI
- AHMI has no influence on rank order of standard displays

Comparison of mean PDTs of human pilots to literature (aggregated for all flight phases)

Measure of Trend

728
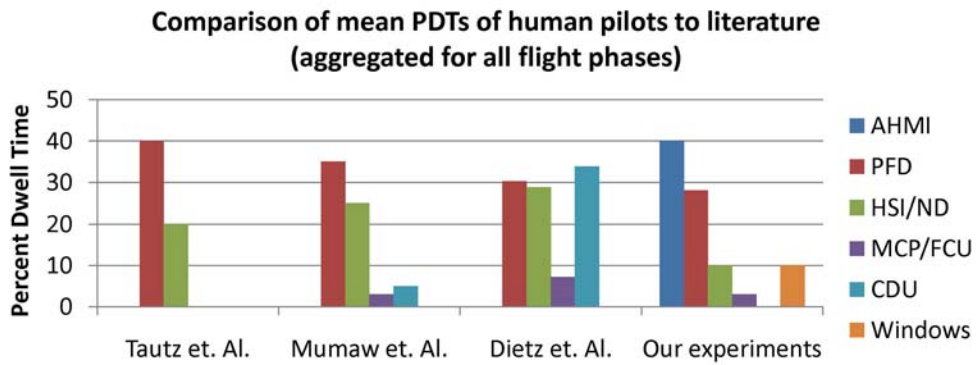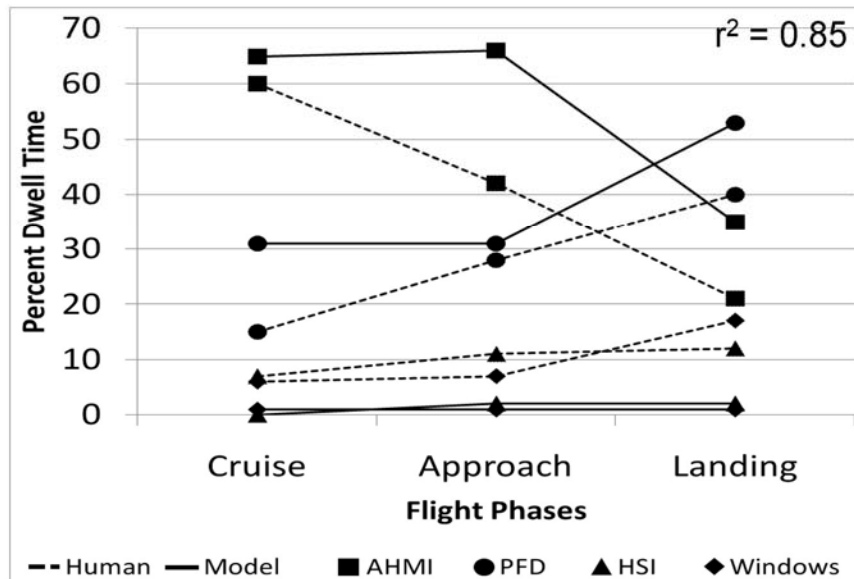
Measure of Location: RMSSD*

1. Calculate individual difference of each human pilot to the group

2. Calculate difference of virtual pilot to the group

3. See if model is in range of human performance
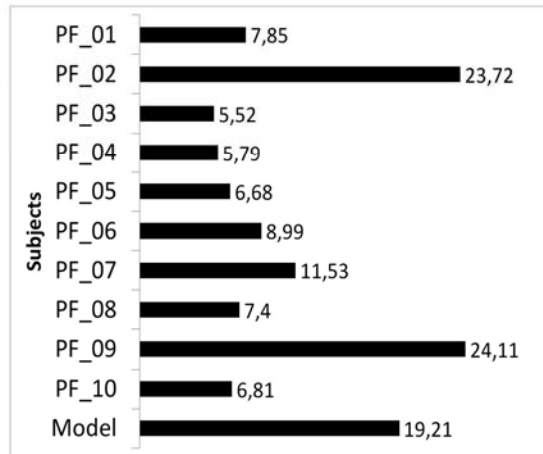
*in cruise phase

Questions/Comments

Thank you for listening!

Any questions?

5.13 Modeling Being "Lost": Imperfect Situation Awareness

Modeling Being "Lost": Imperfect Situation Awareness

Victor E. Middleton

V. E. Middleton, Enterprises, LLC.

2356 Whitlock Place

Kettering OH 45420-1360

(937) 253-1257

middletv@woh.rr.com

Abstract: Being "lost" is an exemplar of imperfect Situation Awareness/Situation Understanding (SA/SU)– information/knowledge that is uncertain, incomplete, and/or just wrong. Being "lost" may be a geo-spatial condition - not knowing/being wrong about where to go or how to get there.More broadly, being "lost" can serve as a metaphor for uncertainty and/or inaccuracy - not knowing/being wrong about how one fits into a larger world view, what one wants to do, or how to do it.

This paper discusses using agent based modeling (ABM) to explore imperfect SA/SU, simulating geo-spatially "lost" intelligent agents trying to navigate in a virtual world. Each agent has a unique "mental map" - its idiosyncratic view of its geo-spatial environment. Its decisions are based on this idiosyncratic view, but behavior outcomes are based on ground truth. Consequently, the rate and degree to which an agent's expectations diverge from ground truth provide measures of that agent's SA/SU.

1.0 INTRODUCTION

A current emphasis in the development of information systems technologies is improving situation awareness/situation understanding (SA/SU)[1] for military and civilian applications. Such improvement requires understanding what is, or may be, wrong with current capabilities.

Modeling and simulation (M&S) can play a significant role in exploring problems with current capabilities, as well as in assessing the efficacy of new or proposed informationtechnologies and determining how best to employ them.Unfortunately current M&S tools face major limitations with respect tothe representation of imperfect SA/SU.

These tools do a reasonable job in representing **incomplete** SA/SU, supporting decision-making and risk assessment with respect to missing data. They fall short in their ability toinvestigate and assess the consequences of**incorrect** and **inconsistent** SA/SU, which requires exploring how to recognize and correct SA/SU based on information that is just plain wrong.

While the focus on incomplete SA/SU probably reflects the current emphasis on providing more information to war fighters through improved information technologies, it discounts equally pertinent issues with respect to the capabilities and fallibilities of the human operator. Although it is hard to argue against giving decision makers more data, it is true that humans can (and frequently do)function well with information that is incomplete or imprecise. Incorrect or flawed informationmaybe even more problematic for SA/SU and associated decisions than missing data. For one thing, plans based on known data gaps and uncertainties are generally more robust to account for unknown factors. Plans based on wrong information may rely too heavily on fallacious assumptions to optimize outcomes, with potentially catastrophic results. In

---

[1] Rather than engage in a discussion as to the differences between SA and SU, I choose to blur them together to a single over-arching concept following the pragmatic definition of [Adam 1993] "knowing what is going on so I can figure out what to do." For more on SA/SU the reader is directed to [Middleton 2010] and the references therein

addition, an incorrect understanding of an operational situation may bias subsequent information processing, and lead to flawed decision-making based on persistent problems with SA/SU.

1.1 Objective and Approach

This paper examines the nature of imperfect SA/SU, how individual decision-makers might recognize problems in their SA/SU, how they might seek to correct those problems, and/or strategies they might employ to mitigate the negative effects of imperfect SA/SU. The paper is based on an easily appreciated exemplar of imperfect SA/SU, the concept of being "lost". The "being lost" exemplar is attractive for a number of reasons:

- First, in both civilian life and military operations being "lost" is a metaphor for uncertainty as to how one fits into a larger context or world view, not knowing exactly what to do, or worse, not knowing where one wants to go and what one wants to accomplish.

- Second, the phenomena of being lost in the non-metaphorical sense, i.e., geo-spatially "lost", provides context for decision-making in which imperfect SA/SU can be expressed in terms of concrete, measurable characteristics of the environment, describing natural and man-made geographical features and expressed in mathematically rigorous geometrical and topological relationships.

- Third, since many M&S tools already incorporate extensive, technologically mature, representations of terrain and geo-spatial relationships, modeling the phenomena of being geo-spatially "lost" provides an accessible and easily understandable test bed for exploration of imperfect SA/SU.

- Finally, there is a well-documented body of research dealing with human way finding, route planning, and navigation, all of which are characteristic of general human abilities to function with imperfect SA/SU. See for example:[Rabul2001]; [Timpf 2002];

[Timpf& Kuhn 2002]; [Richter &Klippel 2005]; [Klippel& Winter 2005]; [Reece, Kraus &Dumanoir 2000]

1.2 Background: Agent Based Modeling

This paper proposes the use of agent based modeling (ABM: see for example: [Ilachinski 1997,2004]; [Borshchev&Filippov 2004];[Macal& North 2005]; [Middleton 2008]; [Easton & Barlow 2002]) to simulate being geo-spatially "lost" in a simulation world.

In ABM, agents (simulated entities)make decisions according to their own individual (and probably imperfect) SA/SU. Each entity will have a "perceived truth" knowledge base – an idiosyncratic view of the operational situation, as seen by that individual and obscured by the agent's local "fog of war".

This paper agues that monitoring the divergence between this idiosyncratic view and simulation "ground truth" can provide a measure, in quantitative terms, of the degree to which each agent's SA/SU may be imperfect. Such a measure, based on allowing each agent to act on an imperfect worldview, supports evaluation of the operational costs of uncertain, incomplete and/or incorrect information. It also supports explicit modeling of leader decision-making processes based on such data, of imperfect command and control, and/or imperfect subordinate receipt of and subsequent execution of orders. This kind of modeling is critical if we are to estimate the benefits of proposed new or modified systems, and/or adjustments to tactics, techniques and procedures.

1.3 Terrain Representation and Movement

The SA/SU measures discussed above are dependent on both the way in which a simulation represents terrain and

731

movement over that terrain. Movement generally has several, possibly overlapping components, which will be referred to herein as:

- *way finding* - the process of learning one's environment to avoid obstacles and find features and points of interest (as is typical of robot "navigation"), which may also incorporate following a general search pattern or algorithm until one's objective is reached;
- *route planning* - the use of algorithms and heuristics to plot a path and/or define a list of instructions describing how to get from one point to another,
- *route following* - the process of actually moving along that path or in accordance with those instructions.

Current models describe terrain (see for example:[Reece 2003]; [Heib et al. 2006]; [Donlon&Forbus 1999] [Glinton et al. 2004]) in either metrical/Euclidean or topological terms, or in some combination of both. Euclidean schemes focus on straight-line distances between features of interest, while topological schemes describe spatial relationships (e.g., adjacency, connectivity, and containment)between such features. In both cases terrain is often overlaid with covering polygons, which can be regular tessellating polygonal tiles (triangles, squares or hexagons), or irregular polygon covering schemes such as Voronoi diagrams.

In strictly Euclidean schemes node-to-node "distance" metrics are based on regular grid coordinates, while more generic topological approaches can reflect a myriad of relational factors, such as trafficability, the availability of cover/concealment, and/or influence ambits based on the proximity of geo-political configurations, static and/or dynamic adversary threats, and the like.

One of the most popular approaches to route finding uses arc-node graphs and shortest path algorithms, e.g., A* or Dijkstra's algorithm. Nodes specify

waypoints along a path, with arcs describing the connections between these nodes. In the case of Euclidean tessellation approaches, nodes typically coincide with the polygons or tiles covering the space, with arcs for each shared boundary line. In a strict topological view the nodes are only defined for points of interest, with the arcs representing possible connections. In either case, arc costs from one node to another can reflect any and all of the "distance" metrics described above, and can be used in "shortest" path algorithms to determine the optimal path through the arc-node structure.

Under any of these schemes, the key questions become first, what does it mean to be lost, second how does an agent find itself in such a state or states, and finally can the agent recognize the problem (i.e., "know" it has bad spatial SA/SU) and correct it?

2.0 METHODOLOGY

In truth, of course, in virtually any real world operational situation, the SA/SU of any individual or organization involved in that operation is going to be less than perfect, with imperfections that range from negligible to catastrophic. Fortunately, as mentioned above, human decision-making and course of action (COA) selection tend to be robust with respect to even many significant imperfections, and, in fact, "good" decision-making considers such imperfections explicitly. For example, military plans strive to make provision for inadequate/poor intelligence and associated unexpected events; "no plan survives first contact with the enemy".

Humans can find their way from one point to another with very rudimentary and/or inaccurate maps. They can frequently function satisfactorily with ambiguous and unclear directions. Of course, in such cases some degree of vituperation may be directed at the

providers of these "direction" aids; a reaction that itself further speaks to the nature of decision-making under stress and uncertain information. An effective simulation of getting or being "lost" should incorporate both this human resilience and the effects of such stresses as uncertainty and time pressure, as important parts of the costs and effects of imperfect SA/SU.

In addition to incorporation of resilience and the effects of stress on decision-making, other requirements for an effective simulation of being lost include:

- The capability for an entity's view of where it is and where it is going to be different from ground truth.
- An error taxonomy that reflects both types of being lost and degrees of "lostness";
- The mechanisms by which an individual achieves different states of being lost; and
- The mechanisms by which an individual recognizes and attempts to correct being lost.

2.1 Mental Maps and Ground Truth

The approach taken herein to meet these requirements begins with assuming a specific formulation for each simulated entity's idiosyncratic view of the world, a "mental map" that represents its own particular, probably distorted, view of ground truth geography.

The mental maps proposed herein are based on an arc/node graph representation. Such a structure be generally accommodated by any of the terrain representations discussed above, and can be used for both route planning and route following.

Each node in the graph will have one or more generic "color" attributes, characteristics that describe features of the node that may be recognizable by

an agent. Such attributes might reflect terrain trafficability, population density, type of buildings or other structures, and so forth. Color attributes can also be used to suggest regional affiliations for nodes, for example geo-political associations, threat areas, broad geographical relationships and the like.

Each node will also have a set of node neighbors listing the color attributes of those nodes with the additional information that defines relationships of the parent node to its neighbors, principally direction and distance.

The ground truth descriptions of node attributes will be numerical or crisp set attributes, the mental map descriptions will be generally be fuzzy set membership attributes. For example, ground truth population density will be in people per square mile, the mental map representation may be some degree of urban, suburban, rural. Ground truth distance will be in meters or kilometers, mental map distances will be close, not to far, remote. Ground truth directions will be in degree from true north, mental map directions will be north, north east, east, and so forth.

The mental maps of each agent in the simulation will allow those entity's to misrepresent ground truth at both the perceptual level (failing to correctly observe ground truth data) and the cognitive level (failing to understand or discern ground truth from the data available to it). The use of fuzzy set relationships in the mental map, however, allows the entity to make decisions based on fuzzy inference rules, i.e., using a best guess or best fit approximation between an uncertain mental map and a crisp ground truth.

The fundamental decision component for each entity is the "next node" selection operation. An agent plans its movement from its mental map, but actual movement takes place in ground truth terrain.

733

Each entity will have a route plan based on the mental map arc/node graph and actual entity movement will be to the ground truth node that best corresponds to the "next node" in that route plan. In the case of multiple candidates for the ground truth "next node" a Monte Carlo selection will be made based on the degrees of fuzzy correspondence to ground truth exhibited by the mental map nodes.

In addition, the mental maps may have incorrect data, arcs between nodes that are not actually connected, and vice versa missing arcs between nodes that actually are, or similarly have nodes that do not actually exist or fail to have nodes that do.

Finally, mental maps will be dynamic, with data being continually filled in, confirmed, or refuted by observation, while ground truth values are typically static, unless they pertain to presence/absence of entities.

2.2 Being "Lost"

A taxonomy of being "lost" then begins with one or more of the following general conditions:

- having a mental map that coincides with ground truth, but with a different registration point – *the agent thinks its current node is different from its real ground truth node*;

- having a mental map that corresponds with ground truth, but with an uncertain or unknown registration point – *the agent is unsure or doesn't know what its current node is*;

- having a mental map that fails to correctly correspond to ground truth witha bad arc/node network connections - *the agent thinks roads or paths lead to places they don't*;and/or

- having a mental map that fails to correctly correspond to ground truth with incorrect characterizations of nodes and arcs - objects the agent isinterested in are in incorrect positions with misleading fuzzy set attributes

and/or deceptive directional/distance relationships as represented on the agent's mental map.

Clearly the "mental map" can be a complex data structure, incorporating for example,a hierarchal structure with different levels of terrain representation based on scale of movement.[Richter and Klippel 2005] , for example, discuss the concept of routes "as a sequence of decision point / action pairs"; which may be combined through *spatial chunking*, grouping several decision point / action pairs into a single route segment, which they refer to as*higher order route direction elements* (HORDE).

2.3 Getting "Lost"

At its core a simulated agent's mental map must address the fundamental question at each stage in an agent's movement: "where to go next?" The mental map needs to answer this question at the level of resolution appropriate to the simulation, which, without loss of generalization, will be taken to be the "next node" selection whether that node represents a "nearest neighbor" point on a grid, or the degree of advancement along a specific route segment or path.

In such a simulation, how does an entity "get lost"?

- by suffering from incorrect initial registration, i.e.,actually starting movement at node or grid coordinates in ground truth network that do not correctly correspond with the mental map;

- by first order "next node" decision point errors - failing to correctly choose the correct ground truth "next node" in a route plan based on ambiguities in the mental map;i.e., misinterpreting the ground truth features that correspond to that map,as for example in failing to recognize the correct intersection to make a turn and/or making a turn at an incorrect intersection;

- by second order "next node" decision point errors - failing to recognize errors in theroute plan itself based on fundamentalmental map errors – trying to move along mental map networks that have extra/missing nodes, and/or bad arcs.

2.4 Recognition of Being "Lost"

Separate from being lost is recognizing that condition; an entitycan be totally wrong about where it is and/or where it's going, but until it recognizes that fact it will continue act in accordance with what it believes its mental map to be. An entity can recognizeit is "lost" in several ways, each of which alsocorrespondsmore broadly tothe recognition of poor SA/SU in non geo-spatial domains:

- insufficient mental map data - the entity's mental map (or more broadly its SA/SU of the current operational situation) does not provide enough information to make a reasonedjudgment as to the correct nextmove or other action, literally not knowing which way to turn. Having absolutely no data is relatively rare, but having missing and/or uncertain data is fairly commonplace. In such cases the "next node" selection would be basically a random draw from available ground truth nodes;

- cumulative mental map discrepancies - the entity'sgeneral accumulation of evidence throughout several "next node" selections, i.e., as the entity moves it observes critical differences between ground truth and environmental features expected in accordance with the mental map, to the point where the entity lack of belief in its mental map renders it as above without sufficient information as to the correct next move

- an abrupt discontinuity between the mental map and ground truth, for example running into a ground truth dead end.

Comparing expectations to actual observed ground truth phenomena can be likened in some ways to the use of a "dead reckoning" function in navigation. Given the uncertainty and possible inaccuracies of an agent's mental map, the agent needs some way to determine its degree of being "lost", which will be defined by thresholds for increasingly aggressive measures to correct mental maps and plans of movement.

2.5 Correcting for Being "Lost"

Given that an entity does recognize it is lost, what measures can it take to correct this situation?The answer is dependent on the way in which the entity become lost and what kind of "lost " it perceives itself to be:

- if the entity suffers from insufficient mental map data, it can either attempt to gain more data, to "scout out the environment" through exploration, or it can pick a robust localway findingstrategy. For example, if lost in a city one can frequently head in a fixed direction with some confidence of eventually striking some linear feature or boundary landmark that will allow reorientation or re-registration of one's mental map;

- as long as the entity appears to be making reasonable progress towards its objective, it can adapt its mental map to remove incongruities between that map and observed ground truth. Such incongruities are likely to be metrical in nature, such as inter-node direction and distance values that may be somewhat off kilter;

- on the other hand, the perception of topological errors, such as missing and/or extra nodes/arcswill probably result in the need to make fundamental changes in the mental map, requiring the acquisition of additional ground truth data through exploration, or the provision of intelligence from sources external to the entity in question.

- on recognition of accumulated route following errors or a discontinuity, the

735

entity may choose the option of retracing the route to some earlier "next node" decision point where there may have been a significant possibility of error, as for example when the choice between two ground truth "next nodes" was in someway difficult, either because of not enough information or because of a choice between two or more very similar nodes.

3.0 DISCUSSION

Of course, the methodology proposed in section 2 is only useful if one can demonstrate a correspondence between the actions of simulated entities and real world behaviors, and more importantly, if the simulation can provide insight into those behaviors that supports improvement in SA/SU for real world operations.

Such demonstration begins with the conduct of simulation experiments that explore:

- relating simulation outcomes to the quality of SA/SU asmeasured by mental map/ground truth incongruities - by the divergence between the expected result of the entity's actions and the observed results;

- appropriate incongruity threshold values for different degrees of corrective actions;

- possible "dead reckoning" functions to support movement towards an objective in the face of imperfect SA/SU;

- the use of landmarks – unmistakable ground truth features, to solve registration problems with mental maps;

- incorporation of mental map uncertainty or belief values in the calculation of route "distance", thus allowing consideration of route robustness with respect to risk of errors in a mental map; and

- the use of various information technologies to support mental map corrections and updates.

4.0 SUMMARY AND CONCLUSIONS

The goal of this paper is not to develop a new theoretical understanding of SA/SU and decision-making. Rather itis to propose an engineering solution to the practical problems faced by decision-makers who must devise information system requirements and evaluate the technological approaches that may be proposed to meet those requirements.

The bottom line for that solution is simulation of the actions of an entity taken in accordance with that agent's unique SA/SU and in expectation of fulfilling one or more goals. By implementing an appropriateset of data structures and inference procedures, an entity should be able to compare expectations to observable aspects of the environment. Entity behaviors are then seen as a cycle of updating/correcting SA/SU, followed by modification of behaviors as that new SA/SU suggests, until goals are achieved or a recognized failure point occurs.

The hope is that focusing on the simulation of "being lost" in a geo-spatial sense canalso provide a template for dealing with being "lost" in more generic imperfect SA/SU contexts. The uncertainty and errors that may be present in geo-spatial information certainly provide a potentially rich source of imperfect SA/SU for simulation experiments and studies.

5.0 REFERENCES

Adam, E. C. (1993). Fighter cockpits of the future. The 1993 IEEE/AIAA Digital Avionics Systems Conference. **Proceedings of the 12th DASC:** (pp. 318-323).

Borshchev, A. and A. Filippov (2004). From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools. The 22nd International System Dynamics Conference, Oxford, 2004.

Donlon, C. and K. Forbus (1999). Using a Geographic Information System for Qualitative Spatial Reasoning about Trafficability. QR99 Thirteenth International Workshop on Qualitative Reasoning, Loch Awe, Scotland.

Easton, A. and M. Barlow (2002). CROCADILE: An Agent-based Distillation System Incorporating Aspects of Constructive Simulation. SimTecT 2002, Melbourne Convention Centre, Melbourne, Australia.

Glinton, R., S. Owens, et al. (2004). Terrain based information fusion and inference. Seventh International Conference on Information Fusion, Stockholm, Sweden, International Society of Information Fusion.

Hansen, S., K. Richter, et al. (2006). Landmarks in OpenLS - A Data Structure for Cognitive Ergonomic Route Directions. Geographic, Information Science. S. B. Heidelberg: 128-144.

Hieb, M., M. Powers, et al. (2006). A geospatial battle management language (geoBML) for terrain reasoning. 11 th International Command and Control Research and Technology Symposium, DE VERE UNIVERSITY ARMS, CAMBRIDGE, UNITED KINGDOM.

Ilachinski, A. (1997). Irreducible Semi-Autonomous Adaptive Combat (ISAAC): An Artificial-Life Approach to Land Warfare (U), Center for Naval Analyses.

Ilachinski, A. (2004). Artificial war: Multiagent-based simulation of combat, World Scientific Pub Co Inc.

Klippel, A. and S. Winter (2005). Structural Salience of Landmarks for Route Directions. Spatial Information Theory, Springer Berlin / Heidelberg. 3693/2005.

Macal, C. and M. North (2005). Tutorial on agent-based modeling and simulation, Winter Simulation Conference.

Middleton, V. (2010). Imperfect Situation Awareness: Representing Error and Uncertainty in Modeling, Simulation & Analysis of Small Unit Military Operations. 19th Annual Conference of Behavioral Representation In Modeling and Simulation (BRIMS), Charleston SC.

Middleton, V. E. (2008). ""Emergent Validation": Defensible and Authoritative Modeling, Simulation, and Analysis for Complex Adaptive Systems." Phalanx, the Bulletin of Military Operations Research 41 No.1 (March 2008).

Raubal, M. (2001). Agent-based simulation of human wayfinding: A perceptual model for unfamiliar buildings. Institute for Geoinformation. Vienna Vienna University of Technology. Doctor of Technical Sciences: 159.

Reece, D. A. (2003). "Movement Behavior for Soldier Agents on a Virtual Battlefield." Presence Teleoperators 12(4): 387-410.

Richter, K. and A. Klippel (2005). A Model for Context-Specific Route Directions. Spatial Cognition IV. Reasoning, Action, and Interaction, Springer Berlin / Heidelberg.

Timpf, S. (2002). "Ontologies of Wayfinding: a traveler's perspective." Networks and Spatial Economics 2(1): 9.

737

5.14   Investigating Intrinsic and Extrinsic Variables During Simulated Internet Search

Investigating Intrinsic and Extrinsic Variables During Simulated Internet Search

Molly M. Liechty
Old Dominion University
Mcris005@odu.edu
Poornima Madhavan
Old Dominion University
pmadhava@odu.edu

Abstract. Using an eye tracker we examined decision-making processes during an internet search task. Twenty experienced homebuyers and twenty-five undergraduates from Old Dominion University viewed homes on a simulated real estate website. Several of the homes included physical properties that had the potential to negatively impact individual perceptions. These negative externalities were either easy to change (Level 1) or impossible to change (Level 2). Eye movements were analyzed to examine the relationship between participants' "stated preferences"[verbalized preferences], "revealed preferences"[actual decisions], and experience. Dwell times, fixation durations/counts, and saccade counts/amplitudes were analyzed.   Results revealed that experienced homebuyers demonstrated a more refined search pattern than novice searchers. Experienced homebuyers were also less impacted by negative externalities. Furthermore, stated preferences were discrepant from revealed preferences; although participants initially stated they liked/disliked a graphic, their eye movement patterns did not reflect this trend. These results have important implications for design of user-friendly web interfaces.

1.0 INTRODUCTION

Everyday a large number of people are utilizing the internet for everything from email to grocery shopping. This use places a greater emphasis on the quality and quantity of information being presented, thus making the design and layout of web pages a crucial component to decision making and user satisfaction. The internet affords people the opportunity to make decisions and purchase goods online with the simple click of a button. Whether the decision involves the purchase of a computer, a car, or even a home, a significant proportion of preliminary purchase decisions (or, "homework") can be accomplished without ever having to leave the comfort of one's home. The information on specific aspects of these designs and their impact on a consumer becomes a very important consideration in this environment.

1.1 Role of Experience

Experience often affects how individuals interact with their environment and the internet is no exception. The amount of expertise an individual possesses has been shown to guide visual search [1]. With experts having a much more refined and effective visual search pattern. The study performed by Reference [1], demonstrated that experts tended to have longer fixations on items of importance to their search and their gaze remained central to the visual scene. Novices in comparison tended to scan the entire scene, with no true direction or long fixations on anything of particular importance to the search.

Experts and novices not only differ in the manner that they scan a visual scene but also in the approach taken to analyzing and inferring information from it. Reference [2] found that when it came to induction and reasoning experts were more flexible than novices in their ability to reason and induce information from a visual scene. Overall, it has been found that experts use past experience and previous knowledge to not only guide visual search, but to compensate for any declining task-specific abilities [3]. Experts use contextual cues and location cues to guide many of their visual searches. This also allows them to become much faster at refining visual searches, with

reaction times shortening with age and expertise. Experience with certain visual cues can also have an effect on visual search of a scene, with knowledge of former individual cues influencing an individual in either a positive or a negative way.

1.2 Negative Externalities

We know from past research [4] that the visual display of a website can have a large impact on an individual's task performance and in general their primary search and satisfaction. Individuals place a premium on their time; when they use the internet, they expect to find the most relevant information to their problem quickly. Most of the research generated on visual layout is studied from the perspective of the effectiveness of a graphic. This study differs from previous research on graphics in that we are looking at how the unpleasantness of a graphic, or a negative object (referred to as a "negative externality") can impact the user. Not only in the way it impacts their visual search, but also their preference for a particular visual scene.

1.3 Stated versus Revealed Preferences

The question of interest is whether it is possible to design an effective website using the stated preferences of individuals. Do internet users really know what it is they are searching for and if so, are they able to convey it verbally? Do verbally stated preferences match with preferences that are revealed during actual internet search?

Organizations of all sizes and interests spend large amounts of money every year on gathering a consumer's stated preference or SPs' and revealed preferences or RPs'[5]. They use this information to do what they called "Consumer Forecasting," or predicting what consumers would want in the future. They were given access to large databases filled with survey and interview information (SP) as well as purchase histories (RP) and were then asked to predict what consumers would do based on all of the data. The

predictions they made were conflicting depending on the type of information they primarily used (SP or RP). This would seem to demonstrate that there is a potential discrepancy between SP (stated preference) and RP (revealed preference).

1.4 Eye Tracking

Eye movements are the most frequent of all human movements and a reliable physiological measure of a psychological state. Eye tracking methodology is based on Reference [6] "eye-mind" hypothesis: the location of a person's gaze directly corresponds to the most immediate thought in their mind. Monitoring an individual's eye fixations (where the eye stops for a moment), their saccades (the rapid movements of the eye), and scan paths allows us to gain insight into certain aspects of an individual's cognitive processes at a particular moment in time. This is due to the eye movements close tie to attentional mechanisms.

Previous eye tracking studies have been used to specifically study how individuals read and scan websites on the internet [7]. When people encounter cognitively complex material, the rate at which they read tends to slow down, as can be indicated from increases in fixations and decreases in saccade durations [8]. In our domain of interest, eye tracking can be used as an unobtrusive way to gain access and insight into what a potential homebuyer is interested in as they view homes on the internet.

1.5 Purpose of the present study

This study was designed to assess the intrinsic factor of experience and its relationship to extrinsic negative externalities (pink paint and power lines). SP and RP were evaluated in order to determine if a discrepancy existed. RP was assessed through length and number of fixations, which is the point at which the eye stops moving for a moment. Also, number as well as amplitude of saccade. From this we are able to measure how difficult and

important the information being viewed is [9], due to the fact that we know intense cognitive processing occurs during a fixation [10]. Thus, we hypothesized that if a person views something important to them they should have a greater number of fixations and longer durations for each fixation. It was also hypothesized that the greater experience an individual possessed for the search task the more refined their visual search pattern would be. In assessing these variables, valuable information was gathered regarding the optimal design of these websites. This information will allow web designers to present the most salient and important information to potential homebuyers quickly and effectively.

2.0 METHOD

2.1 Participants

Twenty-five undergraduates from Old Dominion University and twenty experienced homebuyers from the community were recruited to participate in this study. There were no age requirements for participants, who all had normal or corrected vision (some participants wore contacts but no participants wore glasses) none of the participants were colorblind. Undergraduate participants who finished the experiment were compensated 2 extra credit points at the end of the experimental session and experienced homebuyers were given a $50 gas card if they completed the study.

All participants viewed ten homes; the same ten homes were shown to all participants albeit in a different order. Four homes were digitally altered such that they possessed two levels of what we designated as "negative externalities." A home with a Level 1 externality had a living room with a bright pink wall; this was considered a Level 1 negative externality due to the fact that a homebuyer could easily change pink paint. A home with a Level 2 externality present included a power line in the curb appeal picture (the first picture of the home a participant saw). Power lines were labeled a

Level 2 externality due to the fact that the homebuyer could not change them. The homes as well as the individual rooms within each home were viewed in random order except for the curb appeal picture that always appeared first; a separate computer program generated this random order. Only four homes were altered to include the negative externalities; each participant viewed a home with a Level 1 and Level 2 externality during the experiment.

2.2 Materials and procedure

We used an Eye link 1000 eye tracker, which is a desk mounted eye tracking system offering 1000 Hz pupil and CR (corneal reflection) eye tracking (takes 1000 measurements per second). Participants were asked to rest their head on a chin rest during the experiment, ensuring reliability of the eye link camera. All participants viewed 10 homes presented in random order, very similar to a typical real estate website; the experimenter kept track of the sequence of the homes for data collection purposes later. Of these, each participant viewed two 'substandard' homes - one home with a bright pink painted wall (Level 1 negative externality) and one home with power lines in the curb appeal photograph (Level 2 negative externality). Photographs were selected by the real estate agency.

In order to counterbalance these homes, the first half of the participants observed the house with pink paint as house #4 and the house with power lines as house #7. The second half of the participants viewed the pink paint on house #5 and the power lines on house #9. The homes were presented in random order.

After viewing a room, participants would rate them on a scale from 1 (worst version of that room) to 9 (best version of that room). This rating for each room was treated as the measure of "SP" in the analyses below. Once the rating had been provided, the experimenter would move on to the next picture. All participants received

a short 5 minute break after viewing the first 5 homes.

The dependent variables of interest were fixation duration/count, saccade count, and saccade amplitude.

3.0 RESULTS

To evaluate revealed preferences, the eye tracking variables of fixation duration/count, and saccade count/amplitude were analyzed for each of the homes containing a negative externality (house #3 & 4, pink paint/ house #6 & 8, power line) using 6 (rooms) X 2 (gender) X 2 (negative externality) repeated measure ANOVAs. This allowed for evaluation of the relationship between intrinsic and extrinsic factors and their effects on revealed preferences.

3.1 Role of Experience

A 6 X 2 X 2 ANOVA of fixation duration revealed a significant interaction of Homebuyer X Level 1 negative externality, $F(5,185) = 4.91$, $p = .03$, *partial* $\eta^2 = .117$. Experienced homebuyers had a longer fixation duration when a Level 1 negative externality was present and the novice students had a distinctly opposite reaction with fixation duration declining with the presence of pink paint (Level 1 negative externality). The 6 X 2 X 2 ANOVA also revealed a significant interaction of Homebuyer X Level 2 negative externality, $F(5,185) = 12.09$, $p < .001$, *partial* $\eta^2 = .246$. Experienced homebuyers again had a longer fixation duration when the house contained a Level 2 negative externality (power line), but novice students, as before, had a decreased fixation duration in the presence of a Level 2 negative externality (see figure 1 & 2).

The 6 X 2 X 2 ANOVA of fixation counts revealed a significant 3-way interaction of room X gender X experience, $F(5,185) = 2.41$, $p < .04$, *partial* $\eta^2 = .061$. Experienced male homebuyers had a significantly smaller number of fixations, specifically for the curb appeal photograph (M = 57.6, SD =
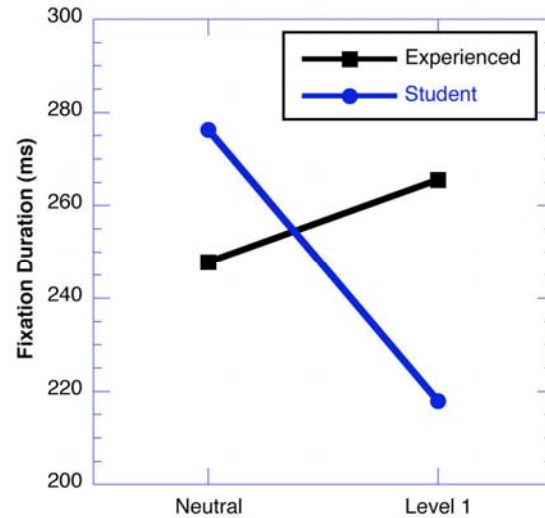

Figure 1. Fixation duration comparison between neutral and Level 1 negative externality.
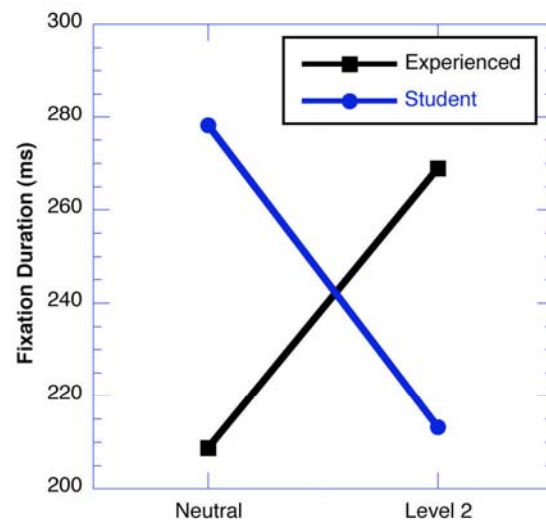

Figure 2. Fixation duration comparison between neutral and Level 2 negative externality.

9.10) compared to novice male undergraduates (M = 75.4, SD = 10.24). In contrast, experienced female homebuyers had a greater number of fixations (M = 66.5, SD = 9.10) compared to the novice female undergraduates (M = 51.58, SD = 7.10) in
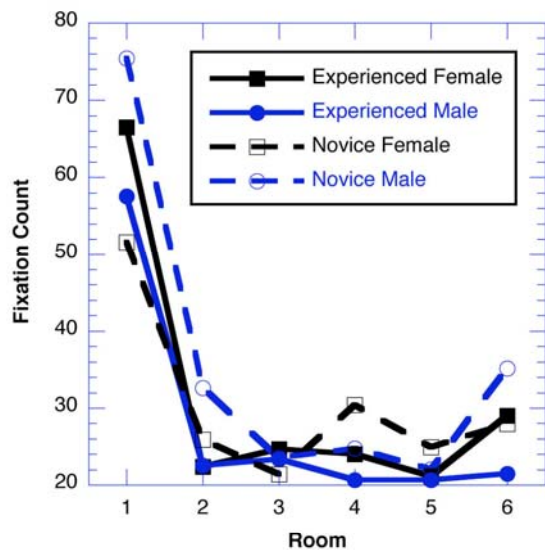
741

Figure 3. Fixation counts of experienced homebuyers vs. novice students.

the presence of a Level 1 externality (see Figure 3).

A significant 3-way interaction of room X gender X experience was also found for saccade count, $F(5,185) = 2.37$, $p < .04$, partial $\eta^2 = .060$. Experienced male homebuyers had a significantly smaller number of saccades, specifically for the curb appeal photograph (M = 57.5, SD = 9.10) compared to novice male undergraduates (M = 75.3, SD = 10.31). Experienced female homebuyers, on the other hand had a greater number of saccades (M = 66.2, SD = 9.10) compared to the novice female undergraduates (M = 51.47, SD = 7.10) in the presence of a Level 1 externality.

Lastly, results indicated that in the presence of both a Level 1 and Level 2 negative externality experienced homebuyers had greater saccade amplitudes than their student counterparts, $F(5,185) = 4.53$, $p < .04$, partial $\eta^2 = .100/ F(5,185) = 3.39$, $p < .07$, partial $\eta^2 = .076$.

3.2 Stated Preferences
Participants were given a General home survey that asked them to rate on a scale of 1 to 9 how important a room would be to them in a home search. The 6 (rooms) X 2 (experience) ANOVA revealed a main effect of room, $F(5, 105) = 3.95$, $p < .01$, partial $\eta^2 = .084$. The curb appeal photograph was consistently rated low, in terms of perceived importance by the novice students (M = 6.34, SD = .36) compared to the experienced homebuyers who gave it a much higher rating in terms of importance (M = 7.45, SD = .35).

Evaluating the Home specific surveys a 3-way interaction of room X Level 1 externality X experience was found, $F(5, 105) = 4.94$, $p < .03$, partial $\eta^2 = .108$. Of interest was the rating given for the living room; when it contained pink paint experienced homebuyers rated it lower (M = 4.0, SD = .57) than novice students (M = 5.69, SD = .99). When it was neutral, experienced homebuyers rated it higher (M = 6.5, SD = .56) than novice students (M = 5.8, SD = .52).

3.3 Scan paths
Observing the scan paths of experienced homebuyers compared to novice students a difference was observed in the number of saccades and the amplitude of saccades, this also appeared to be tempered by gender. Results and scan paths demonstrated that novice male students had a greater number of saccades; their eyes traveled around the photographs more often and their saccade amplitudes were shorter such that their movements were small bursts across the visual scene. This when compared to experienced male homebuyers reveals that the latter had a smaller number of saccades; their eyes moved around the photograph less often, and because their saccade amplitudes were longer with fixations closer together, it appears experienced male homebuyers had a predetermined idea of where in the visual scene they wanted to look.

Similar to the pattern for male participants, novice female participants also had fewer saccades but just like their novice male
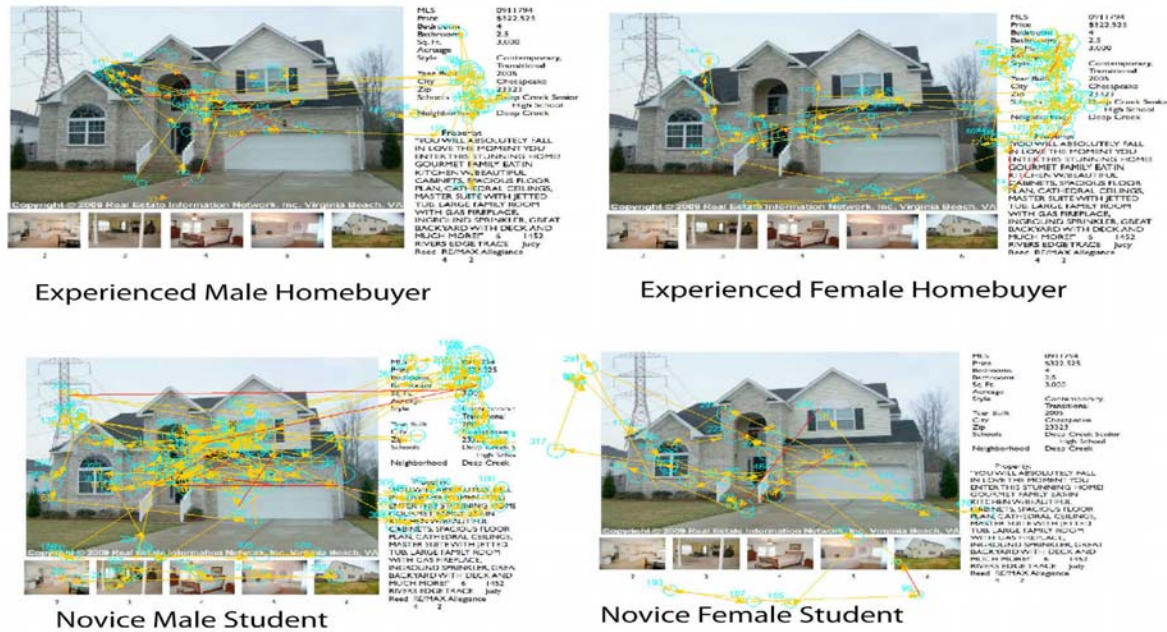
742

Figure 4. Scan paths with Level 2 negative externality present.

counterparts with shorter amplitudes. Their eye movements were also short, quick movements around the visual scene. Experienced female homebuyers, in contrast to novice female participants, had a greater number of saccades with longer saccade amplitudes. They again appeared to have specific points on the screen that they wished to analyze as indicated by the longer saccade amplitudes, similar to experienced male homebuyers (see figure 4 for sample scanpaths).

4.0 DISCUSSION

The purpose of this study was to investigate the intrinsic factor of experience and the extrinsic factor of negative externalities on stated and revealed preferences in an internet search task.

4.1 Role of experience

Results revealed experience with internet home search made a difference in the way in which the search task was performed. Gender also appeared to have an effect on experience. For males experience was expressed through fewer fixations, with

longer durations, fewer saccades, and longer saccade amplitudes indicating that they were focused on specific aspects of photographs and had preconceived ideas about what they wished to investigate. Experienced females in contrast to their novice female counterparts had a greater number of fixations, with longer durations, and fewer saccades, but as with experienced male homebuyers they also expressed longer saccade amplitudes again, indicating a clear idea for the direction of their eye gaze in the visual scene. Experienced male participates had the fewest number of fixation counts and saccade counts which seems to be indicative of low interest in the photographs altogether (reference [10]).

Saccade amplitudes and fixation placement were interesting in this study. The saccade amplitudes followed distinctly opposite patterns for experienced versus novice participants. Experienced homebuyers demonstrated a longer array of visual movements than novice students across the webpage; which may indicate a

743

preconceived idea of where in the visual scene they would find relevant information to their search. Evaluating scan paths it is clear that experienced homebuyers tended to fixate on the central portions of the visual scene and spent less time on the perimeters which supports previous research (reference [1]) that experts tend to focus on the central aspects of a visual scene.

The present study is unique in that it takes into account differences in experience using physiological measures of an individual during a search task focusing on eye tracking indices of dwell time, fixation duration/count, and saccade count/ and amplitude. These physiological data suggest that experienced homebuyers might be better at acquiring target specific information than novice students since they seem to localize their area of interest quickly (revealed by longer saccade amplitudes and a fewer number of saccades).

4.2   Role of negative externalities

Negative externalities in this study were operationalized in two levels depending on the ease to which they could be modified by the user – Level 1 (pink paint) and Level 2 (power lines). From previous studies we know that a greater number of fixations and the longer the duration indicate that viewers are focusing intense cognitive resources on the object being viewed [10]. In this study the effect of the negative externality varied by the experience of the participant and the room that was being viewed. When viewing the living room photograph with pink paint (Level 1 externality) experienced homebuyers found it to be less of a detractor than novice students. The presence of the pink paint did not stop the experienced homebuyers from investigating the home more fully in contrast to the novice students.

The same was true when a power line or Level 2 negative externality was present. Again, the fixation duration and fixation count, were affected. The experienced

homebuyer when presented with the Level 2 externality would spend more time looking at the home. Their eye would stop more often and for a longer duration, compared to the novice student. The novice student would spend less time looking at the home photographs, and would fixate less often for a shorter duration of time, compared to the experienced homebuyers. The results indicate that the experienced homebuyers were less distracted by the presence of a Level 2 negative externality and had more interest in the home photographs as a consequence.

It appears that the Level 1 and Level 2 externalities led to fewer physiological influences on visual search for experienced homebuyers; instead their presence perhaps gave experienced homebuyers additional reason to scrutinize the photographs carefully possibly to find positive aspects to compensate for the presence of the Level 1 and Level 2 negative externalities. In either case the presence of a negative externality appeared to affect how participants viewed the entire home. This is interesting news for designers; it is evident that one "bad apple" could have the potential to spoil the entire barrel.

4.3   Stated versus Revealed Preferences

In the General survey novice students rated that the curb appeal photograph would be of little importance to their visual search which is in contrast to the experienced homebuyer who gave it a much higher rating of importance for the home search. A discrepancy was observed when these ratings were compared to RP determined through eye tracking variables for the novice student. For fixation count and saccade count there was a main effect of room caused by the curb appeal photograph, regardless of experience. In other words, the curb appeal photograph generated the maximum interest during visual search, the experienced homebuyers realized this and there was no discrepancy between their SP

and RP, but the novice students did not. It is interesting that this finding was consistent even for participants that took the General survey after viewing all of the homes.

The Home specific survey also demonstrated a discrepancy between SP and RP, but only for the Level 1 negative externality. When pink paint was present the ratings given (SP) were opposite of the eye tracking measures recorded (RP). Regardless of experience participants rated the living room photographs low, but when we look at their eye tracking variables, they spent a considerable amount of time viewing those same photographs, illustrating a discrepancy. This discrepancy was larger for the novice student than the experienced homebuyers. This discrepancy was not found for the Level 2 negative externality.

Overall, our results indicate that there is a difference between a person's stated preferences and revealed preferences; although not consistent across all variables a discrepancy was found and that experience may temper how large a discrepancy exists or if one will exist at all.

4.4 Implications
In the present study, a trend toward discrepancies between SP and RP were found dependent on experience, supporting previous research [11] that preference may be something that is formed in many different stages of a decision and that experience may solidify that preference [2]. Furthermore, differences were found in the search patterns that were used by the experienced and novice participants as a function of what they were looking at on the webpage. These results have significant implications for web design for the population in general. The scan paths revealed that the graphic portions of the web pages were indeed where participants spent the greatest amount of their time looking, reinstating the idea that visual aspects of a web page are the most important. Knowing your audience and the amount of experience they possess as they view a webpage carries important considerations for design in the future.

5.0 REFERENCES
[1] Willaims, A., Ward, P., Knowles, M., Smeeton, N., 2002. Anticipation Skill In a Real-World Task: Measurement, Training, and Transfer in Tennis. *Journal of Applied Experimental Psychology,* Vol. 8, No. 4, 259-270.

[2] Shafto, P., Coley, J., 2003. Development of Categorization and Reasoning in the Natural World: Novices to Experts, Naïve Similarity to Ecological Knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Vol 25, No. 4. 641-649.

[3] Hoyer, W. & Ingolfsdottir, D., 2003. Age, Skill, and Contextual Cueing in Target Detection. *Psychology and Aging.* Vol. 18, No. 2. 210-218.

[4] Galitz, W. O. (1996). The essential guide to user interface design. *Holboken, NJ: Wiley.*

[5] Verhoef, P., & Franses, P. (2003). Combining revealed and stated preferences to forecast customer behaviour: Three case studies. *International Journal of Market* Research, 45(4), 467-474.

[6] Just, M.A. & Carpenter, P.A. (1976). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.

[7] Grant, E., & Spivey, M. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science,* 14(5), 462-466.

[8] Liversedge, S.P., Paterson, K.B., Pickering, M.J. (1998). Eye movements and measures of reading times. *Eye Guidance in Reading and Scene Perception.* G. Underwood (Ed). Amsterdam, Netherlands: Elsevier, pp. 55-75.

[9] Fitts, P. M., Jones, R. E. & Milton, J. L. 1950. Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautiical Engineering Review*, 9 (2), 24-29.

[10] Rayner, K. 1998. Eye movement and information processing: 20 Years of Research. *Psychological Bulletin, 124*(3), 372-422.

[11] Simonson, I. (1999). The effect of product assortment on buyer preferences. *Journal of Retailing,* 75(3), 347-370.

Investigating Intrinsic and Extrinsic Variables During Simulated Internet Search

Molly Liechty, B.S.
Poornima Madhavan, Ph. D.
Old Dominion University

Purpose of the Study:

- Assess the intrinsic factor of experience and its relationship to extrinsic negative externalities (pink paint and power lines)

- Stated vs. revealed preferences

Previous Research

- Role of Experience
  - Expertise has been shown to guide visual search (Willaims, Ward, Knowles, & Smeeton, 2002)
    - Experts – more refined/ Novices – scanned entire scene
    - Experts – more flexible/ Novices – rigid ideas (Shafto & Coley, 2003)
- Preference
  - Discrepancies have been studied, but not through the use of physiological measures/ (Horskey, Nelson, & Posavac, 2004; Simonson, 1999; Zajonc, 1980)

Research Questions:

- Will experience alter the way in which homes are viewed?
- How will negative externalities impact the overall visual search?
- Will stated preferences differ from revealed preferences?
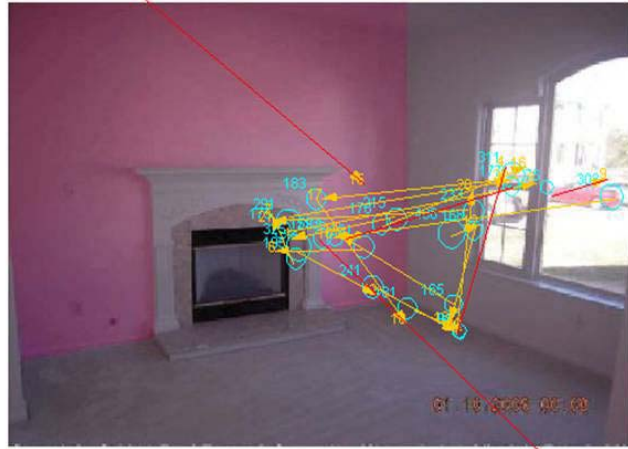
Ocular Tracking

Just & Carpenter (1976) "Eye-mind" hypothesis

Eye Tracking Variables

- Dwell Time
- Fixation Duration
- Fixation Count
- Saccade Count
- Saccade Amplitude

  - (Loftus and Mackworth, 1978)
  - (Rayner 1998)

Scan Path

MLS	0902171
Price	$320,000
Bedrooms	3
Bathrooms	2
Sq. Ft.	2,240
Acreage	0.35
Style	Ranch
Year Built	1997
City	Chesapeake
Zip	23321
Schools	Western Branch Senior High School
Neighborhood	Waterstone

Property:
"SIMPLY EXTRAORDINARY! PRISTINE 3BR, 2 BATH HOME IN HIGHLY SOUGHT AFTER NEIGHBORHOOD. FRESHLY PAINTED. FEATURES LARGE MASTER SUITE W/FP, FORMAL LR, DR, FR AND MORE. A MUST SEE! CORPORATE OWNED AND SOLD AS-IS, WHERE-IS. NO REPORTS OR REPAIRS BY SELLER. A GREAT HOME."
3   2000 WATERSTONE COVE   Geoffrey Cardwell   "Long & Foster Real Estate, Inc." 4   5

Website Display

MLS	0902171
Price	$320,000
Bedrooms	3
Bathrooms	2
Sq. Ft.	2,240
Acreage	0.35
Style	Ranch
Year Built	1997
City	Chesapeake
Zip	23321
Schools	Western Branch Senior High School
Neighborhood	Waterstone

Property:
"SIMPLY EXTRAORDINARY! PRISTINE 3BR, 2 BATH HOME IN HIGHLY SOUGHT AFTER NEIGHBORHOOD. FRESHLY PAINTED. FEATURES LARGE MASTER SUITE W/FP, FORMAL LR, DR, FR AND MORE. A MUST SEE! CORPORATE OWNED AND SOLD AS-IS, WHERE-IS. NO REPORTS OR REPAIRS BY SELLER. A GREAT HOME."
3   2000 WATERSTONE COVE   Geoffrey Cardwell   "Long & Foster Real Estate, Inc." 4   5

Negative Externalities

Pink Paint   Power line

Experimental Design:

- 25 ODU undergraduates
  - Received class credit
- 20 Homebuyers from the community
  - Received $50 gas card
- No time limit to view photographs
- 10 homes, each home has 6 photographs (curb appeal, kitchen, living room, master bedroom, master bathroom, and back yard)

Surveys

- Demographic
- General
  - Curb Appeal

    1   2   3   4   5   6   7   8   9

- Home Specific
  - 1   2   3   4   5   6   7   8   9

    Worst   Average   Best
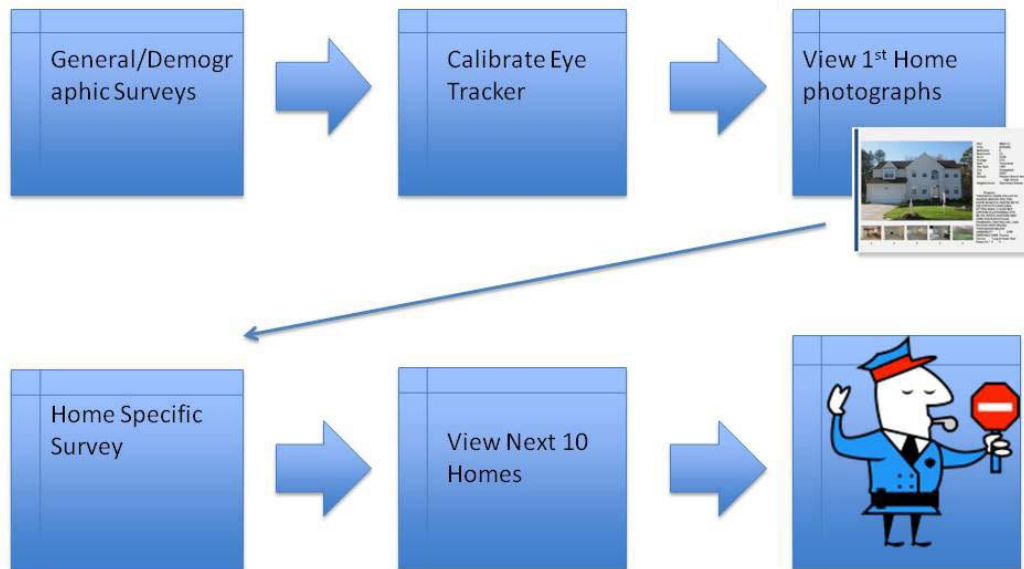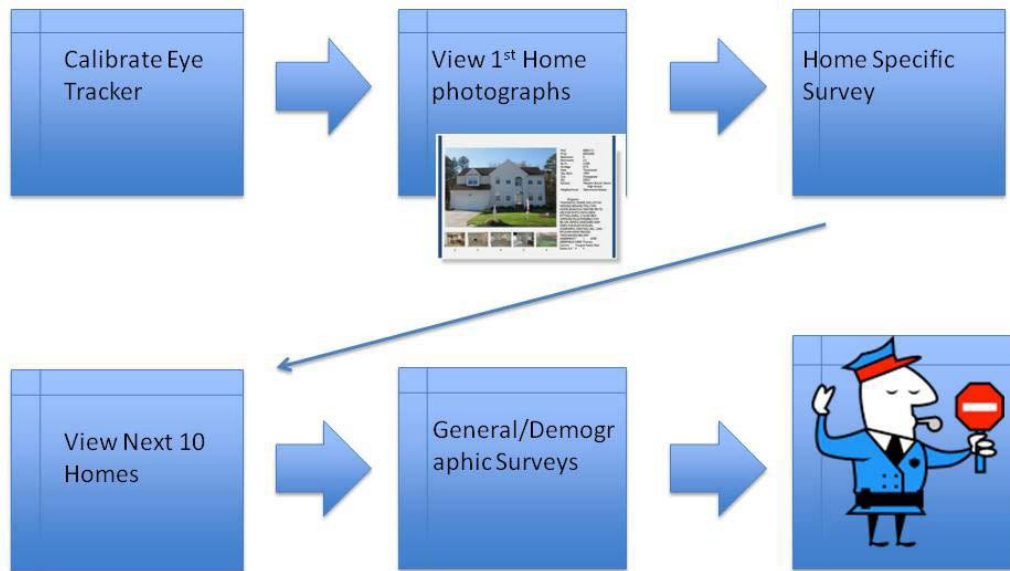
Hypotheses:

- **Interest in a scene =**
  - longer dwell times
  - longer fixation durations
  - greater number of fixations

- **Greater experience =**
  - more refined search

- **The presence of negative externalities =**
  - shorter dwell times
 - shorter fixation durations/counts

Analysis

- ✓ 6(rooms) X 2(expert/novice) X 2(Negative Externality)
- ✓ 6(Home survey) X 2(expert/novice) X 2(Negative Externality)
- ✓ 6(General survey) X 2(expert/novice)
- ✓ Scan paths

THIS WAS DONE FOR THE FOUR HOMES THAT CONTAINED OUR NEGATIVE EXTERNALITIES (HOUSE #3 & #4, PINK PAINT/ HOUSE #6 & #8, POWER LINE)

Experience

- Fixation Duration: Homebuyer X Negative externality
  - $F(5,185) = 4.91$, $p = .03$, $partial$ $\eta^2 = .117$ (Level 1)
  - $F(5,185) = 12.09$, $p < .001$, $partial$ $\eta^2 = .246$ (Level 2)
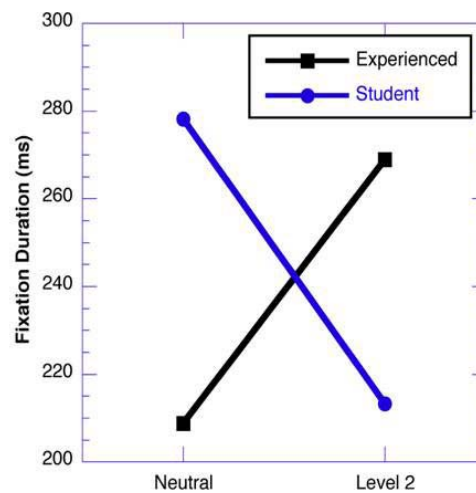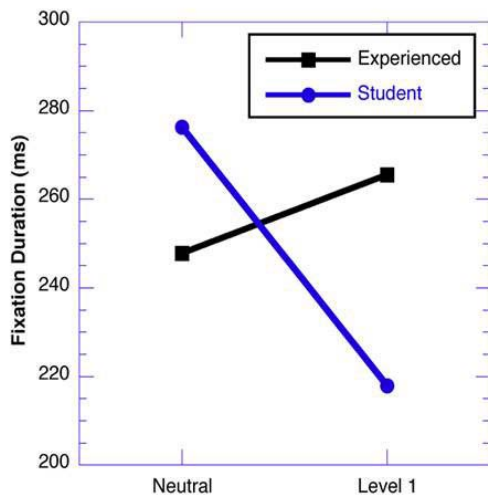
Experience

- Fixation Count: room X gender X experience
  - $F(5,185) = 2.41$, $p < .04$, *partial* $\eta^2 = .061$.
- Saccade Count: room X gender X experience
  - $F(5,185) = 2.37$, $p < .04$, *partial* $\eta^2 = .060$
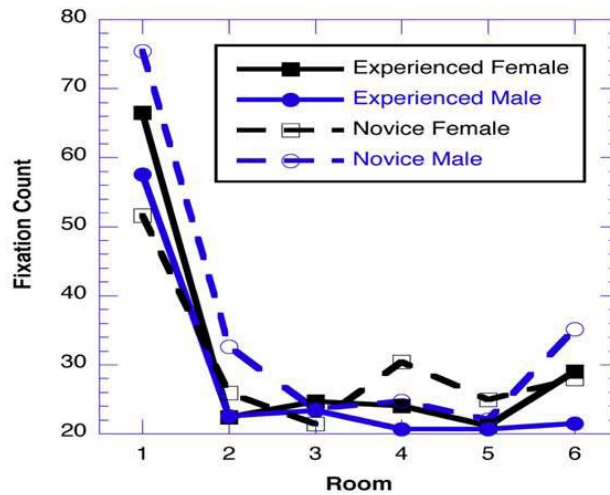
Experience

- Saccade Amplitudes
- 6 (room) X 2 (experience) X 2 (negative externality)
  - Main effect of experience (both Level 1 and 2 neg. externality)

Experienced Male Homebuyer

Experienced Female Homebuyer
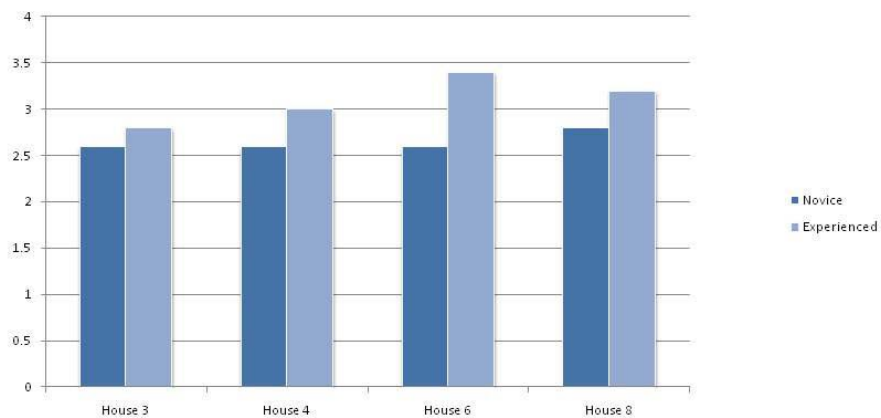
Novice Male Student

Novice Female Student
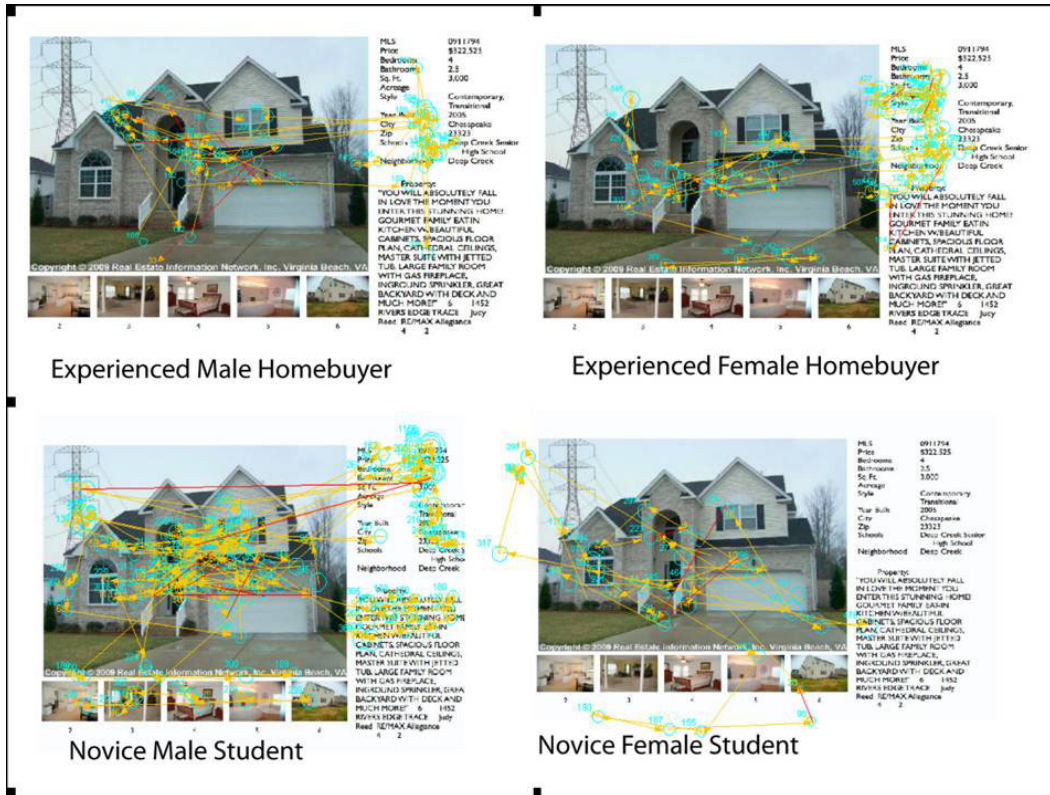
Interest Areas

Main Effect of Interest Area:

House 3: $F(5,185) = 22.07$, $p < .01$, partial $\eta^2 = .408$

House 4: $F(5,185) = 27.07$, $p < .01$, partial $\eta^2 = .451$

House 6: $F(5,185) = 10.62$, $p < .01$, partial $\eta^2 = .244$

House 8: $F(5,185) = 13.27$, $p < .01$, partial $\eta^2 = .287$

Interest Areas

Stated vs Revealed Preferences

General Survey Rating
    main effect of room, $F(5, 105) = 3.95$, $p < .01$,
    partial $\eta^2 = .084$
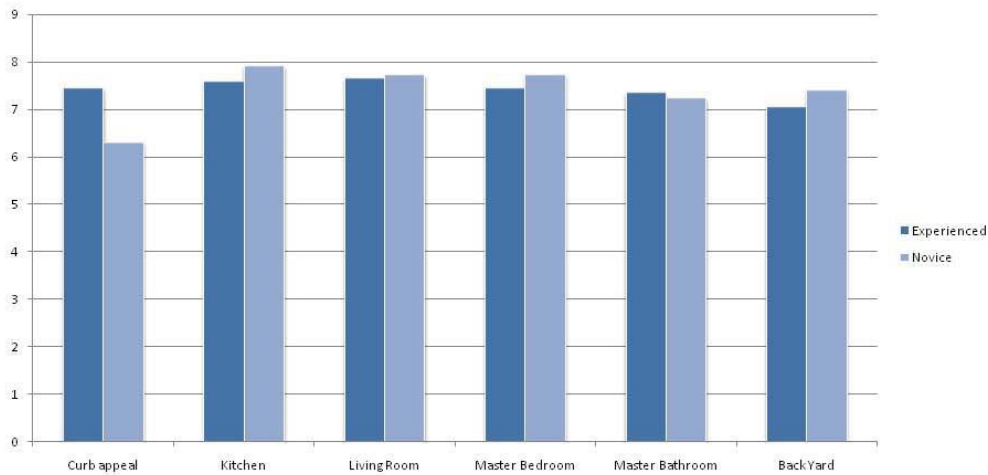
Stated vs Revealed Preferences

- Home Survey: room X Level 1 externality X experience was found
  - $F(5, 105) = 4.94$, $p < .03$, partial $\eta^2 = .108$
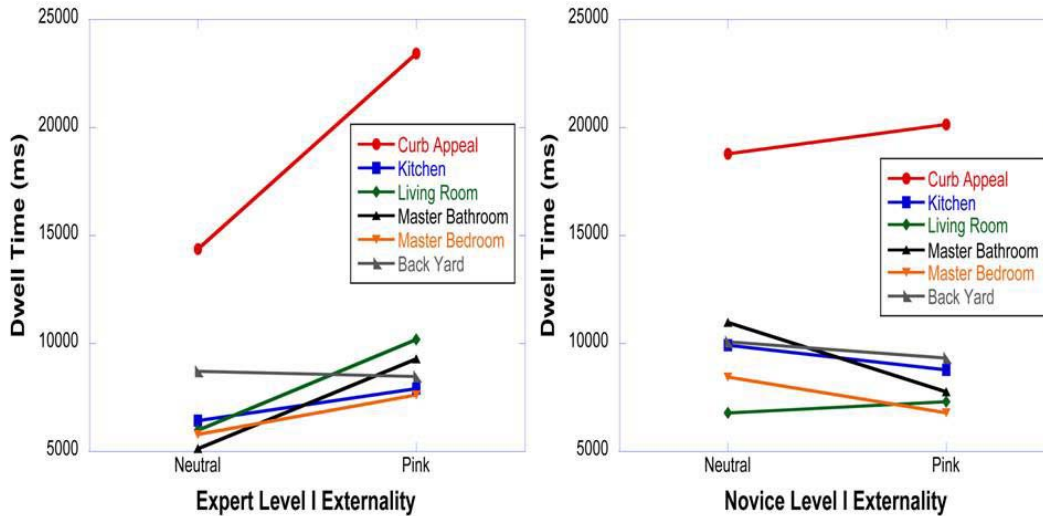
Discussion

- Experience:
- Experience with internet home search made a difference in the way in which the search task was performed.
- Saccade amplitudes were of a distinctly opposite patterns for experienced versus novice participants. Experienced homebuyers demonstrated a longer array of visual movements than novice students across the webpage; which may indicate a preconceived idea of where in the visual scene they would find relevant information to their search

Discussion

- Experience:
- These physiological data suggest that experienced homebuyers might be better at acquiring target specific information than novice students since they seem to localize their area of interest quickly (revealed by longer saccade amplitudes with a fewer number of saccades and fixations).

Discussion

- Negative Externalities:
  - Varied with the experience of the participant
  - Level 1 and Level 2 externalities led to fewer physiological influences on visual search for experienced homebuyers compared to novice homebuyers
- Stated vs Revealed Preferences
  - There was a discrepancy present
    - Greater for novice homebuyers

Practical Implications

- **Research Implications**
  - Discrepancies between stated and revealed preferences can occur and researchers need to be conscious of this. Level of experience seems to interact with this discrepancy. An individual very familiar with a situation may have more concrete preferences.

- **Web Design Implications**
  - Graphics can be very influential
  - Negative aspects of an image may or may not be detrimental depending on the individual's experience level

5.15 Following Human Footsteps: Proposal of a Decision Theory Based on Human Behavior

Following Human Footsteps: Proposal of a Decision Theory Based on Human Behavior

Faisal Mahmud, M.Sc.
Ph.D. Student and Research Assistant
Department of Civil and Environmental Engineering
Old Dominion University
fmahm001@odu.edu

Abstract. Human behavior is a complex nature which depends on circumstances and decisions varying from time to time as well as place to place. The way a decision is made either directly or indirectly related to the availability of the options. These options though appear at random nature, have a solid directional way for decision making. In this paper, a decision theory is proposed which is based on human behavior. The theory is structured with model sets that will show the all possible combinations for making a decision. A virtual and simulated environment is considered to show the results of the proposed decision theory

1.0 INTRODUCTION

When we eat, move or work, it is natural that we are driven by some will power of our own. Sometimes this will power is guided or supported by available options. These options though appearing as random nature, have a solid directional way for decision making. No matter what we choose to do, our action is based on sets of decisions that influence our will power. It is also true that human behavior to make a decision for the same goal may vary from time to time. There have been done various types of work related to human behavior and decision making. Stewart Robinson has been investigating the use of artificial intelligence methods as a means for representing human decision-making in simulations since mid-1990s. One motivation for his work for modeling human decision-making was to add extra complexity to a model in order to improve its accuracy. The goal of this paper is quite different than the previous works. The aim of this paper is to establish a theory which is based on human behavior for decision making. The theory itself is developed by observing human acts in real-life and field surveys. It should be mentioned that the results presented here are not yet supported from a psychological point of view by the experts in that field; rather, an engineering analysis with field level observation is put together to support the study.

2.0 METHODOLOGY

2.1 Approach for Analysis

Consider a simple scenario like this - a person wants to go from one place to another, suppose from home to work (home based work trip). He/she has the choice from three modes of alternatives - (1) by walk, (2) by car or (3) by public transport like bus. Now let's look at when the person will choose a specific mode of transportation to reach the destination.

(1) The person will walk from origin to destination when -

- Other two options are unavailable
- Those options will take more time than walking when he/she is in a hurry
- Relative cost
- He/she is prescribed by a doctor to walk for this particular trip
- On his/her way to the work, it requires to do another work which is easy if walking option is chosen.

760

(2) The person will choose car from origin to destination when -

- This option is available to him/her
- It will take less time than walk or by taking bus
- Relative cost
- Weather condition
- On his/her way to the work, it requires to do another work which is easy if car option is chosen
- If he/she can drive or someone will drive for him/her
- It is comfortable as well as safe and secured.

(3) The person will choose bus from origin to destination when -

- This option is available to him/her
- It will take less time than walk or by taking car
- Relative cost
- Weather condition
- Taking bus than car is a better option for this particular trip
- It is comfortable as well as safe and secured
- He/she is supposed to take the bus rather than car because there is a special bus service for the office where he/she works.

Now, if we go for an analysis for this example, we can identify some common matches within these three particular options -

- Availability of the options
- Relative cost
- Time required for the trip
- Safety, security and comfort.

Again, if the person has equal availability to all three alternative options, then the way

he/she can decide to go for a particular option depends on -

- Better available option within the options or
- Random choice.

This random choice is also a factor which is related to other specific taste factors like style, job position and protocols, or simply a person's own desire at that very time and day.

This example reveals the fact that human decision for a particular thing (in this case home based work trip) may vary depending on the availability of the options as well as by random choice. To create a structured decision tree in this regard may give a solid base to explain human behavior in making decision theory.

3.0 DECISION TREE

In order to put all available options together for a particular decision making, the following decision tree is shown:
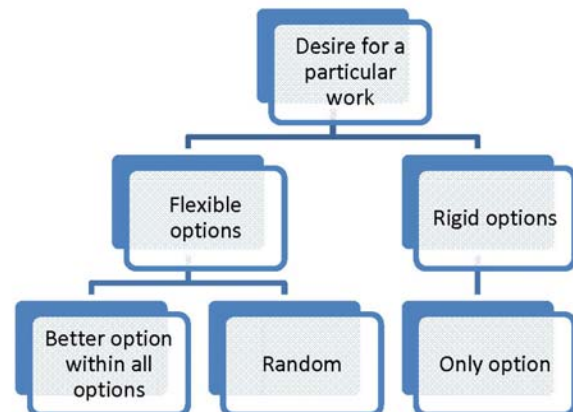

Figure 1: General Decision Tree for Making a Particular Decision.

In Fig. 1, a decision tree is shown for making a particular decision. If we correlate this decision tree with the given example we can find out how human behavior captures the options in making a decision. Each box

in the figure corresponds to different types of decision sets.

4.0 DATA SETS FOR ANALYSIS

The data for this analysis is collected from Dhaka, the capital of Bangladesh. Two months worth of data from different particular locations within Dhaka is used for the analysis. All the field surveys and interviews of public as well as observations are put together for the proposal of the decision theory. For this study, the surveys and observations are conducted for person's trip from one place to another considering different available options. It was checked from public opinions how they react and get used to with the changes of transportation alternatives. Three categories of people were chosen based on their income - (1) High income, (2) Middle income and (3) Low income people. The reason to choose the income category is because it is one of the most influential factors that may control the choice of modes in transportation alternatives.

4.1 Simulated Analysis

In addition to real-life field surveys and observations, a simulated environment was also created to explain the human behavior. Because of the extensive simulation runs with all the decision sets, corresponding results are not yet achieved.

5.0 RESULTS

The results from this analysis can be explained from observed data as well as field surveys. It has been found that flexible option is only available to the people of higher income and in some cases to the middle income people. People with low income have none but the rigid option. Therefore, if we want to investigate the movement analysis of low income people based on the use of transportation alternatives, we do not need to go for further

details - they are accessible to a place by either on feet or by public transportation. Now, let's take into account the high and middle income people. In this case, it is quite uncertain which alternative mode they will consider for travel - personal car, or bus, or by walking? They have the flexible option and when a category of people are within this flexible option, in depth analysis is quite necessary to explain their movement. The analysis result showed that people will choose a better option within all available options if and only if we can control the random choice - which is nothing but a choice based on personal satisfaction. A question may arise why decision theory is proposed based on human behavior or what is the importance of this type of research? First of all, if we can analyze human behavior for a particular action, for an example, in this research choice of alternative modes is the primary area of analysis, we may go ahead to control that behavior and thus do further analysis for controlling specific kind of behavior when we need to do so. Suppose, in the field of transportation engineering, sometimes we need to introduce special types of traffic management policy to a certain route or area for either a special event or to control and avoid traffic jam. A city without proper planning and lack of sufficient roads may suffer a gridlock situation and therefore need to control human behavior in choosing alternative modes of transportation. The study area, Dhaka, is now facing extreme traffic jam resulting in economic loss to the country. An option to solve the system is to control human behavior in selecting alternative modes for travelling. If we can control human behavior to make a decision for choosing a mode in this city area, we may overcome the problem to a certain extent. The way to control this is to control the flexible option of the people. Though this is not a fair way to guide human decision to a particular track, but to some extent it is better. The reason is that, if people become dependants of using private cars by ignoring public transportation and for using private cars by 20% of the people in the city center

causes a problem to the rest of the 80% people, it is better to cut that available flexible option of the 20% people in using their private cars. There may be argument on this matter, but the goal of this paper at this point is not to go for these arguments, rather than to suggest a theory on decision making.

6.0 CONCLUSION

Human behavior is a complex nature to explain with theoretical or simulated analysis. But, real-life observation with practical field survey can help to explain this complex nature in simulated environment. "The motivation is to model human decision-making so it is better understood and it can be improved. This should help to improve the performance of the systems in which the humans are interacting. The concentration is no longer on making models more accurate, but on using the models to assess the effects of human interaction and to look for ways of changing the human interaction in order to improve system performance. Model accuracy plays a secondary role to generating insight and understanding. This is the motivation behind the knowledge based improvement methodology," (Stewart Robinson, Modelling Human Decision-Making, para. 27). The aim of this paper is to do this with supportive evidence. The only limitation that may exist is the psychological analysis of the particular study. Otherwise, this work is a unique approach to propose a decision theory.

7.0 REFERENCE

[1]. Robinson, S. (2003). Modelling Human Decision-Making. *Proceedings of the 17th European Simulation Multiconference (ESM 2003)* (D. Al-Dabass, ed.). SCS, Delft, pp. 448-455.

8.0 ACKNOWLEDGMENT

5.16 Use of Inverse Reinforcement Learning for Identity Prediction

Use of Inverse Reinforcement Learning for Identity Prediction

Roy Hayes, Jonathan Bao, Peter Beling, Barry Horowitz
Department of Systems Engineering
University of Virginia
Charlottesville, VA 22904
{rlh8t, jhb6f, pb3a,bh8e}@virginia.edu

Abstract. We adopt Markov Decision Processes (MDP) to model sequential decision problems, which have the characteristic that the current decision made by a human decision maker has an uncertain impact on future opportunity. We hypothesize that the individuality of decision makers can be modeled as differences in the reward function under a common MDP model. A machine learning technique, Inverse Reinforcement Learning (IRL), was used to learn an individual's reward function based on limited observation of his or her decision choices. This work serves as an initial investigation for using IRL to analyze decision making, conducted through a human experiment in a cyber shopping environment. Specifically, the ability to determine the demographic identity of users is conducted through prediction analysis and supervised learning. The results show that IRL can be used to correctly identify participants, at a rate of 68% for gender and 66% for one of three college major categories.

1 INTRODUCTION

There has been significant work in the field of machine learning to understand human decision making. Inverse Reinforcement Learning (IRL) is a method for computers to learn to perform complex tasks by watching human operators [2]. IRL is built upon Markov Decision Processes (MDPs), which examine sequential decision making over time. Decision makers are modeled to choose actions based upon maximizing reward, which is captured by a reward function that assigns preferences to being in certain states. Decisions made in the present directly impact future decisions and opportunities, often stochastically, so short-term gain must be balanced against future goals. Decisions are complex because an individual may have many actions to choose between and may have to assimilate various pieces of information and trade-offs between conflicting goals. These types of decisions are commonplace in daily life, from choosing which lane to drive in on the interstate to choosing when to buy or sell stocks.

Our thesis is that IRL techniques can be used to understand human decision making by creating a mathematical model of the human's decision strategy. We do not claim that people solve complex mathematical formulae mentally while making difficult decisions; however, a projection of their preferences can be captured through machine learning. Specifically, we can begin to understand under which conditions an individual would take a certain action and therefore find if people adopt different strategies to the same problem. There is reason for optimism that IRL can model decision making. Researchers have run controlled experiments where a participant is instructed to exhibit certain preferences and have shown heuristically that a computer is able to mimic the behavior by solving a mathematical version of the problem [2]. We feel that IRL does indeed capture aspects of an individual's true decision rules, but the previous work has not tried to verify this important requirement for many applications through rigorous analysis.

1.1 Expected Contribution

This work identifies a bridge between those who develop solutions to sequential decision problems and those who have methods to test and quantify human behavior. In broad terms, the two fields can be defined as machine learning and cognitive science. Machine learning encompasses artificial intelligence and reinforcement learning as researchers who train computers to solve decision problems that may be too difficult for humans to solve. Cognitive science studies how the human brain uses information, and cognitive scientists run controlled experiments to investigate the impact of some changing condition on human performance. The two fields join when researchers use machine learning algorithms to understand human decision making. This work lies in this middle area, as we investigate the potential of IRL to analyze decision strategies through human experimentation.

In the machine learning literature found predominantly in the engineering field, researchers have not validated that IRL captures human decision making through robust experimentation. The literature is focused on improving algorithms in terms of speed and accuracy [6], or adapting work to apply to a larger class of problems [1,3]. The algorithms are heuristically validated by instructing human experts to follow different strategies that map well onto the qualities the computer was trained to learn. The machine learning literature lacks hypothesis testing that would demonstrate that IRL can find differences in decision making between groups of people, and we therefore look to the cognitive science field to find studies analyzing human sequential decision making.

Cognitive science is devoted to understanding how humans make use of information in the brain and is therefore closely related to characterizing decision making. Researchers in cognitive science make use of human experiments to perform hypothesis testing; often to compare two groups of people to one another. There have been studies where IRL and MDPs could be used to analyze the data gathered from human experiments, but researchers lost power by only using results-based analysis. For instance, [4] performed a sophisticated experiment with a motorcycle simulator and asked the riders to identify potential hazards and collected eye-gaze data. The researchers could have sought to understand where the user was looking as a function of the objects on the screen, but instead were relegated to analyzing the higher-level metric of general size of viewing area.

There has been a great deal of work in the economics field to investigate the ability of mathematical models to describe real human behavior. Ref. [5] completed a survey of research in predominantly the economic field that analyzed human decision making with respect to MDPs. They found that humans perform near-optimal behavior in discrete decision problems, but the opposite was true for continuous decision problems. As a case study, they highlighted work by RAND where the decision of Air Force pilots to remain in service or retire to the civilian sector was analyzed. Among other practical conclusions, the work showed that prediction is a valid method for testing MDPs as a decision framework.

2 MATHEMATICAL FORMULATION

IRL refers to any method where a reward function is learned to mimic expert behavior through observation [2]. The foundational premise is that a rational actor may choose between several actions and may conduct analysis to determine the best course of

action. Decisions are captured in a mathematical model that can be analyzed and optimized to find the best action. The theory is applied to sequential decision making where the actor will have to make a series of time-ordered decisions. This raises a difficult problem that requires analysis to solve because current actions impact future decisions and opportunities.

2.1 Markov Decision Processes

IRL uses the well-understood framework of Markov Decision Processes (MDP). MDPs are built upon the idea that all of the information one needs to make a decision is characterized by the state of the system. Markov chains become powerful when applied to decision making because the probability of transitioning to a certain state is dependent on the current state and the decision maker's action. The decision maker chooses actions at every time point with the updated knowledge of his or her situation. Decisions can be chosen greedily to maximize short-term gain, but it is clear that since decisions made in the present directly affect future opportunities that a farsighted strategy is needed to make the best possible decisions.

We use the notation from [2] to formulate Markov Decision Processes. An MDP is fully described by the tuple (S, A, T, γ, D, R), where:

- S is the set of all possible states, and the state at time t is given by s_t.
- A is the set of all possible actions, and the particular action chosen at time t is a_t
- T is the function of state transition probabilities.
- $\gamma \in [0,1)$ is a discount factor
- D is the initial-state distribution

- R is the transition reward gained from taking action a_t at s_t while transitioning to s_{t+1}

Once the MDP has been completely formulated, the goal is to solve the problem by developing an optimal policy π that maps an optimal action to every state. Due to the stochastic nature of MDPs, the objective is to choose actions that maximize total expected reward. The goal of the decision maker is to find π that maximizes V_π and therefore know which action to choose at $t = 0$. Once the system transitions to the next state at $t = 1$, then the actor has the information necessary to take the best action, i.e. the actor does not determine at $t = 0$ how he or she will act in the future. Once the problem has been formulated as such, the optimal policy may be derived through dynamic programming or reinforcement learning.

2.2 Discretized-Reward Search Method for IRL

As discussed above, the computer learns to mimic a human by learning the problem that the expert is attempting to solve. [2] places constraints on the problem definition so that IRL uses a linear reward function in order to apply standard optimization techniques to perform policy evaluation. If we relax these constraints, then we void the developed algorithm and must perform IRL in another manner. We have developed an exhaustive search algorithm by discretizing the space of reward functions to a finite set in order to attribute reward functions to actions which, although it has its limitations, works for a broader class of problems.

The process of mapping a reward function to an observed action path x is as follows:

1. Start with initial weight w^0, which is the starting point for weight iteration. There must be some method to iterate through all of the feasible weights. For example, if we choose $|w|_1 = 1$, then the first weight could be $w^0 = (1, 0, \dots)$, and the next weight would be $w^1 = (0.9, 0.1, \dots)$. Set $i = 0$.
2. Solve or approximate the optimal policy to the MDP where $R = w^i \cdot \varphi(s)$. Simulate an action path using the optimal policy and set as x^i. Use the size of observed actions x as stopping criteria if necessary.
3. Use a reward distance function to find the difference in the rewards generated by x and x^i with respect to w^i and set as d^i.
4. If w^i is not the last weight, then find w^{i+1} and set $i = i + 1$. Return to Step 2.
5. Find the minimum value for d^i and create a set of all the w^i with that value. These are all of the weights and corresponding reward functions that match the observed actions.

There are several design choices in the problem definition that are necessary to implement this method. The set of all weight vectors must be discretized into a finite countable set, and there must be a method for iterating through the set. The MDP must either be able to be solved through dynamic programming or an optimal solution must be approximated with reinforcement learning. Finally, a distance function must be developed to compare the expert's policy and optimal policies generated for candidate reward functions.

3 METHODOLOGY

We conducted human experiments to investigate the capability of using inverse learning methods to perform identity prediction. A task that meets the criteria of a sequential decision problem is online shopping. Shoppers navigate an online environment searching for items, and their actions can be readily extracted from looking at browsing history data. By recording their browsing history, we have a noninvasive sequential view of their actions and can determine how the user assimilated information to make decisions. Inverse learning calculates the user's policy in all situations and will describe the user's objective function. We will be able to characterize how a particular user performs the task of shopping for an item.

We developed an experiment to test how participants perform the task of purchasing a gift using an online shopping website. Each participant underwent a 30 minute experiment, during which they performed 4 trials. At the start of each trial, the participant is given a profile of a person to buy a gift for, which includes personal characteristics and possible suggestions of what that person may like or dislike. The user was given 5 minutes and a budget of $100 to perform the task, during which time he or she browsed the item selection provided by the website and selected one or more gifts to purchase. Participants were not given any instruction except for the profile of the participant and to remain on the shopping site and not view another site. After some pretesting, we determined there were 10 predominant types of pages available at Walmart.com (e.g. store department page, item list page, and checkout page).

3.1 Setup of the MDP and Corresponding IRL Method

We set the state vector to represent the number of pages of each type the user has viewed. State transitions are deterministic, as the user fully decides which page type to view next. With a standard reward function, the optimal policy would simply choose to view the page type with the highest reward

over and over again. A reward function that causes users to switch pages, as opposed to choosing the same one over and over again, would be one that took into account the law of diminishing returns. A user may prefer to view one type over another, but as they view that page multiple times they receive decreasing reward. If we let M be the maximum number of pages a user wishes to view of a certain type, we could scale the reward gained from choosing a page by a factor that is inversely proportional to the number of times the page was viewed up to M visits. In Eq. (3.1), a is the action corresponding to the page the user wants to view next, s is the complete state, s_a is the current number of pages of type a that the user has viewed, and w_a is the weight corresponding to that page type.

$$R(s, a) = w_a * \left(M - (s_a + 1) \right) \quad (3.1)$$

This reward function is nonlinear; it is not a linear combination of the state variable because only the part of the state regarding the action taken contributes to the transition reward. We therefore use the Discretized-Reward Search Method. There are many different ways to discretize the space. We chose to have each weight be nonnegative, and the sum of the weights was equal to 1, so that the possible value for each weight w_i was [0, 1]. We also set the granularity of each weight, such that a value of 10 meant we divided the range of [0, 1] into 10 equal parts, i.e., $w_i = 0.0, 0.1, ..., 0.9, 1.0$. The analysis reported here was performed using finer granularity of 20.

We developed a method to find the distance between two policies under a single reward function. Instead of simply counting how many times the user policy and optimal policy differed, we used the amount of reward each policy generated as a differencing metric. The Incremental

Reward Difference method (IRD) compares two action paths by sequentially examining each time period and finding the difference in the total accumulated reward up to that point. For example, consider a simple reward function of $R = 0.4s_1 + 0.6s_2$, and we had one policy of (1,1,2,2) and another policy of (2,2,1,1). The total reward accumulated by both policies is 2.0, so it is important to have a metric that takes into account sequence order. In our method, the difference of the total reward accumulated after the first period is 0.2 (0.6-0.4), after the second it is 0.4 (1.2-0.8), after the third it is 0.2(1.6-1.4), and after the fourth it is 0 (2.0-2.0). Therefore, the difference between the policies is 0.8, which takes into account sequence and end result.

For each experiment observation, we store all of the reward functions that were closest to the expert and use a measure of central tendency as the point estimate of the true reward function. The standard method to measure distance between two n-tuple vectors is Euclidean distance. Standard cluster analysis uses the centroid as the averaging measure for a group of points, but this most likely will lead to an impossible reward function. Instead, we find the medoid (found in k-medoid cluster analysis), which is the element in the cluster that has the shortest average distance to every other point in the cluster.

3.2 Weights of Evidence Prediction Models

Rating the quality of generated rewards by IRL is directly dependent on the application. We have chosen to examine identity prediction in the sense that we could find someone's reward function and correlate identifying information by comparing against known data. We therefore desire the reward functions to group people into clusters based upon demographic similarities. In this

section we discuss how we rate whether meaningful clusters are formed by analyzing experimental data.

Scoring models can be used to identify separation in the data and provide a means for prediction. Weights of Evidence (WOE) are used to convert data from an individual into a single score, and it is desired that scores are able to differentiate people. Scoring models predict a binary outcome, such as good (G) or bad (B), according to a vector of features. Given a feature vector x, the quatities of interest are $P(G|x)$ and $P(B|x)$. The score s is the log odds score, which can be broken into a population score s_{pop} and an information odds score s_{inf} by using Bayes Rule and the properties of logarithms, as shown in Eq. (4.7).

$$
\begin{aligned}
s(x) &= ln\frac{P(G|x)}{P(B|x)} \\
&= ln\frac{p(G)}{p(B)} + ln\frac{f(x|G)}{f(x|B)} \\
&= s_{pop} + s_{inf}
\end{aligned}
\tag{4.7}
$$

The information odds score can be calculated from the data as the distribution that the feature vector takes a value given the person is good or bad. If each variable in the feature vector is conditionally indendent given the individual is good or bad, then the information score is given by Eq. (4.8).

$$
s_{inf} = ln\frac{f(x_1|G)}{f(x_1|B)} + \cdots + ln\frac{f(x_n|G)}{f(x_n|B)}
\tag{4.8}
$$

Each log odd in the information score is the WOE indicating G for that particular variable. The WOE is the log odds that the feature x_i takes on a particular value given the person is good, and can be directly calculated from the data. For instance, the value $f(x_1 = 0.1|G)$ is the proportion of the number of good people where $x_1 = 0.1$ over the total number of good people. This method requires a descritization of each variable x into multiple bins.

4 IMPLEMENTATION AND RESULTS

We discuss our findings with the caveat that the analysis was exploratory, and there was no previous work or principles that people grouped according to the tested demographic factors are expected to perform the task differently (e.g., there is no definitive theory that males utilize a different shopping strategy than females). However, IRL methods that find more correlation between demographic group and strategy are preferable, and this metric can be used in model selection when choosing between several predictive methods.

4.1 Results from WOE Scoring Models

For each IRL model, we developed credit scoring models for the gender and major variables. For the binary variable gender we calculated for male and not male, while for major we had to make three models for arts and not arts, engineering and not engineering, and commerce and not commerce. Each model was built using the 10 weights from the reward function as predictive features. The features were separated into bins based upon taking values of 0 through 0.3 and an additional bin for being greater than 0.3. Once the weights of evidence were calculated by determining the log odds that a feature took a particular value, scores were assigned to each trial based upon the reward function. Frequency plots showed the distribution of scores according to the group the individual belonged to. The frequency plots for the model are shown in Fig. 4.1.
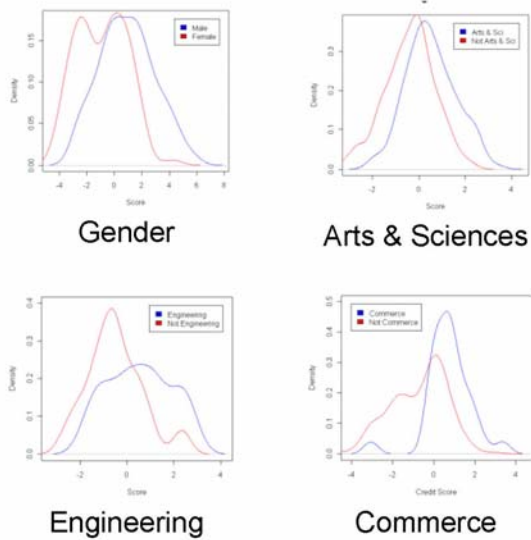
Gender | Arts & Sciences

Engineering | Commerce

Fig. 4.1. Results from WOE scoring according to gender and school

The scoring models based upon WOE had the potential to perform the two tasks of identifying separation and predictive power. A benefit of the scoring analysis was the ability to visualize the data. If there was separation between the two demographic curves, we could have determined an optimal score threshold and tested for accuracy with training and testing data as in the regression analysis.

To investigate the predictive power of the scoring models, Receiver Operating Characteristic (ROC) curves were built to show the tradeoff between sensitivity and specificity with choosing a particular cutoff point. For instance, one may choose a cutoff such that a high percentage of males were correctly labeled as males, but in general there is a tradeoff associated with having an increased number of females that are incorrectly labeled as males. The former is the true positive rate, while the latter is the false positive rate, and a sample ROC curve is shown in Fig. 4.2 for gender.
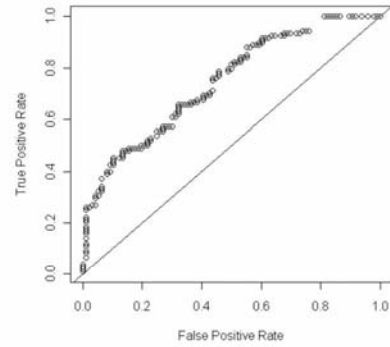

Fig. 4.2. ROC Curve for gender

An ideal ROC curve would be one that included the point (0,1) indicating it was possible to achieve a 100% true positive rate with a 0% false positive rate. Using this logic, curves are measured by the area under the curve (AUROC) where a value of 1 is considered the best while 0.5 is the worst. We show the AUROC score for the model in Table 4.1.

We developed decision rules to identify each participant and record the number of correct identifications. As an example, we found that classifying those with an engineering score above 1.05 as engineers and below as non-engineers yielded a 78% success rate. To further discriminate, we separated the non-engineers based upon the commerce score threshold of 0.92, and subsequently had a total success rate based on major of 66%.

Table 4.1. Performance metrics to predict user identity

	AUROC	% Correct
Gender	0.745	67.6%
Arts & Sci	0.718	68.6%
Engineering	0.716	77.9%
Commerce	0.810	86.2%
Total Major	N/A	66.2%

5 CONCLUSION

Inverse reinforcement learning has the capability to quantify human decision making through observation. This machine

770

learning method can be used in many applications, including attribution. However, the literature does not verify that IRL captures real decision making. IRL has been tested to heuristically demonstrate its merit through controlled experimentation. In this work, IRL was used to analyze human behavior in experiments where the participants were not given any instruction regarding strategy. The most difficult aspect of performing IRL is developing an MDP that can capture the different strategies real participants use when performing a task. We provided a methodology that allows researchers to statistically test the ability of various IRL models to map reward functions to actions with respect to some application, in this case attribution. Models were compared based upon group significance testing and predictive power. These statistical methods can be used with any IRL scheme to test their usefulness with respect to attribution.

Without IRL, it is very difficult to understand the strategy that each participant used to perform shopping. At the most, the other study could only analyze the relative frequencies of the number of times each page type was visited, and would lose any information on the *order* that the participant viewed pages. People choose the next page as a direct result of the page they are currently viewing and overall preferences of the final goal and the required steps to achieve satisfaction. Most work on analyzing differences in humans choose to test the change in an observable variable, and it is rare to see analysis on the mathematical formulation of strategy.

The next step in assessing IRL as it pertains to capturing decision making is to analyze individual consistency. This work focused on analyzing differences between groups, whereas consistency analysis would investigate similarities of an individual over time. The primary goal of consistency analysis would be to show that an individual has an underlying strategy to perform tasks, and although actions may appear to be different across trials where the individual is placed in new situations, the strategy captured by the reward function would remain constant. This would serve to demonstrate that the user has a reward function and that IRL could recover the correct one. Users would need to be observed performing the same task multiple times, which would require additional testing than the data gathered for this experiment.

REFERENCES

[1] Abbeel, P.; Dolgov, D.; Ng, A.Y.; Thrun, S., "Apprenticeship learning for motion planning with application to parking lot navigation," Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, pp.1083-1090, 22-26 Sept. 2008

[2] Abbeel, P.; Ng, A., "Apprenticeship learning via inverse reinforcement learning," Proceedings of the twenty-first international conference on Machine learning, pp.1, 04-08 July 2004

[3] Choi, J. and Kim, K. 2009. Inverse reinforcement learning in partially observable environments. In Proceedings of the 21st international Jont Conference on Artifical intelligence (Pasadena, California, USA, July 11 - 17, 2009). H. Kitano, Ed. International Joint Conference On Artificial Intelligence. Morgan Kaufmann Publishers, San Francisco, CA, 1028-1033.

[4] Hosking, S., Liu, C., Bayly, M. The visual search patterns and hazard responses of experienced and inexperienced motorcycle riders, Accident Analysis & Prevention, Volume 42, Issue 1, January 2010, Pages 196-202

[5] Rust, J. Do people behave according to Bellman's principle of optimality? Stanford - Hoover Institution, 1992

[6]   Syed, U.; Schapire, R. E., "A game-theoretic approach to apprenticeship learning," Advances in Neural Information Processing Systems, 2008

Use of Inverse Reinforcement Learning for Identity Prediction

Authors

Roy Hayes, Jonathan Bao,
Peter Beling, & Barry Horowitz

Presentation Outline

➤ **Introduction**
➤ Methodology
➤ Experiment
➤ Conclusions
➤ Next Steps
➤ Question

Markov Decision Processes

- S is set of all possible states at time t, given by S_t
- A is the set of all possible actions, given your in state S_t
- T is the Transition Probabilities, given your in S_t and chose a_t
- $\gamma \in [0,1)$ is a discount factor
- D is the initial-state distribution
- R is the Transition reward gained from taking action a_t at S_t

Inverse Reinforcement Learning (IRL)
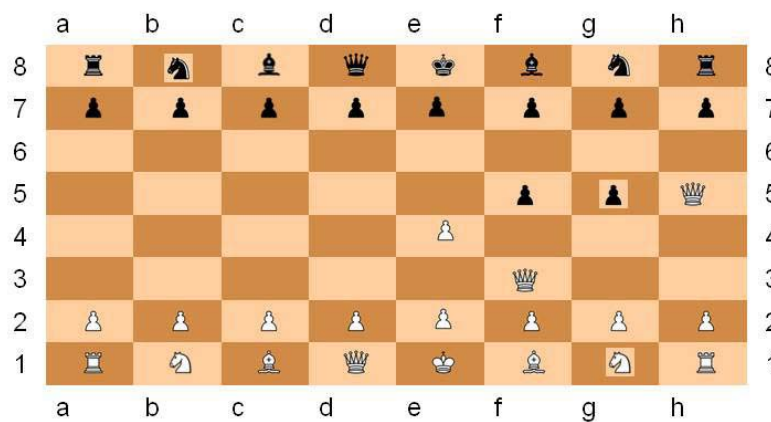
1.e2 e4
2.d1 f3
3.f3 h5

4

Defining Reward function

What move to make?
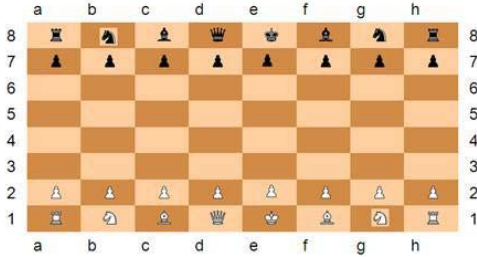
$$R(s,a) = \sum W_i * \phi_i$$

What to eat?

$$R(s,a) = W_i*(M-(S_a+1))$$

Selecting Optimal Reward Function

Linear Programming
- Only works on linear reward functions
- Computationally efficient
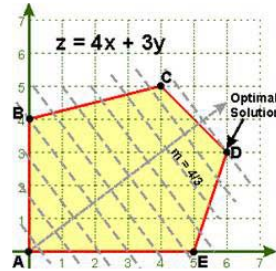
Discretizing the space & perform exhaustive search
- Conceptually easy
- More robust
- Computationally expensive

$$R(s,a)_1 = 0.9* \phi_1 + 0.1* \phi_2$$
$$R(s,a)_2 = 0.8* \phi_1 + 0.2* \phi_2$$
$$R(s,a)_3 = 0.7* \phi_1 + 0.3* \phi_2$$

Presentation Outline

➢Introduction

➢**Methodology**

➢Experiment

➢Conclusions

➢Next Steps

➢Question

Comparing Different Reward Functions

Medoid - Which is the element in the cluster that has the shortest average distance to ever other point in the cluster
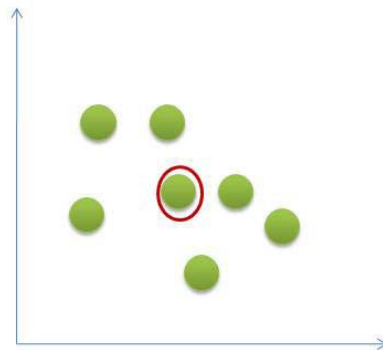
Predicting Identity

$$S(X) = ln\left(\frac{P(G|x)}{P(B|x)}\right)$$

$$= ln\frac{P(G)}{P(B)} + ln\frac{f(x|G)}{f(x|B)}$$

$$= S_{pop} + S_{inf}$$
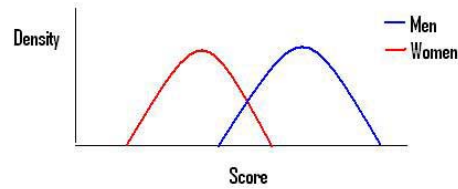
$$S_{inf} = ln\left(\frac{f(x_1|G)}{f(x_1|B)} + \cdots + ln\left(\frac{f(x_n|G)}{f(x_n|B)}\right)\right)$$

Presentation Outline

- Introduction
- Methodology
- **Experiment**
- Conclusions
- Next Steps
- Question

777

Experimental Setup

Goal : To determine attributes of an individual shopping on walmart.com

Procedure: Participant is given 4 portfolio of people to shop for and 5 mins per person to complete the shopping

Data Collected: The sequence of different types of pages viewed

Number of Participants :30

Walmart.com

Experiment's Markov Decision Process

State – The current type of page you are viewing (e.g. store department page, item list, and checkout page)

Action – The next type of page selected

Transition – The transition probability is 100%

Reward Function – $R(s,a) = W_i*(M-(S_a+1))$

Experimental Results
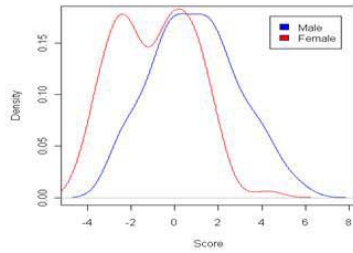
Gender
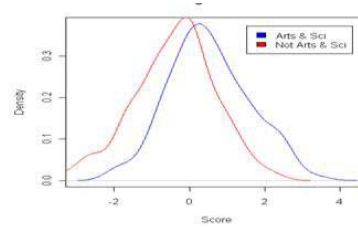
Arts & Science
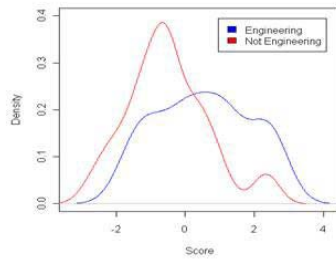
Engineering
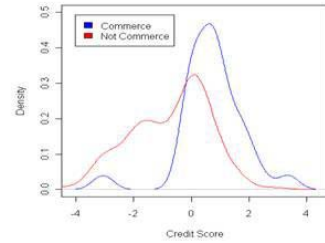
Commerce

Experimental Results

Performance metrics to predict user identity
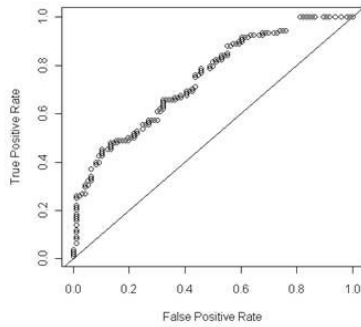
ROC Curve for gender

	AUROC	% Correct
Gender	0.745	67.6%
Arts & Sci	0.718	68.6%
Engineering	0.716	77.9%
Commerce	0.810	86.2%
Total Major	N/A	66.2%

779

Conclusions

Inverse Reinforcement Learning can be set up in many different ways

Machine Learning methods can be applied to attribution

The statistical techniques presented are a good way to harness the predictive power of Inverse Reinforcement Learning

Next Steps

To determine consistency of an individual's reward function

To examine Inverse Reinforcement Learning for training purposes

Any Questions?

5.17 Social-Cognitive Biases In Simulated Airline Luggage Screening

SOCIAL-COGNITIVE BIASES IN SIMULATED AIRLINE LUGGAGE SCREENING

Jeremy R. Brown & Poornima Madhavan
Old Dominion University
jrbrown@odu.edu   pmadhava@odu.edu

Abstract. This study illustrated how social-cognitive biases affect the decision making process of airline luggage screeners. Participants (n = 96) performed a computer simulated task to detect hidden weapons in 200 x-ray images of passenger luggage. Participants saw each image for two (high time pressure) or six seconds (low time pressure). Participants observed pictures of the "passenger" who owns the luggage. The "pre-anchor group" answered questions about the passenger *before* the luggage image appeared, the "post-anchor" group answered questions *after* the luggage appeared, and the "no-anchor group" answered *no* questions. Participants either stopped or did not stop the bag, and rated their confidence in their decision. Participants under high time pressure had lower hit rates and higher false alarms. Significant differences between the pre-, no-, and post-anchor groups were based on the gender and race of the passengers. Participants had higher false alarm rates in response to male than female passengers.

1.0 Visual Search Tasks

The primary goal of visual search tasks is to effectively differentiate critical signal stimuli from irrelevant non-signals (known as *distractors*). There have been various studies looking into different aspects of visual search tasks. Many of the visual search studies focus on visual clutter and its effects on the search task [1, 2, 3,& 4]. Another factor that affects visual search is age [1, 2]. Visual clutter is typically caused by "distractors". Studies by Grahame Laberge, and Scialfa (2004) [1] and McPhee, Scialfa, Dennis, Ho, and Caird (2004) [2] found that as clutter is increased, the time it takes to detect the target also increased. They found that the task increased in perceived difficulty as a consequence of increased clutter. This is because it is harder to recognize an object as the clutter increases [5]. As there are more objects to search through to find a target, the search will take longer and will be less efficient. In some instances, however, detection time can *decrease* with clutter, especially when the clutter causing objects are of a larger size than the target [6]. This is due to attention being drawn to the "empty" space between the clutter causing objects.

In addition to the amount of clutter, search efficiency is affected by what the clutter consists of and its physical similarity or dissimilarity to the target. The more similar the distractor is to the target, in terms of color, brightness, and orientation, the more difficult it is to find the target [3]. Target objects that have multiple colors or textures are harder to detect in a cluttered environment, especially when the clutter is of a similar color or texture to that of the target [7].

The reason visual search tasks are the focus of several researchers is that there are several jobs in the real world that use visual search as the main component of the work such as airport luggage screening. The primary task for airline luggage screening requires the screener to search through an x-ray image and detect a particular dangerous target from the clutter of non-lethal objects. On one level, luggage screening is a simple signal detection task where the screener must differentiate critical signals (or, threat objects) from background noise. However, the detection task is complicated by the fact that on several occasions, the threat object must be detected within an initial glimpse of the x-ray image, spanning just a few seconds.

Airport luggage screening is further complicated by the number and diversity

782

of threat objects that might potentially be embedded in a piece of luggage. All the studies described above have effectively addressed the cognitive aspects of visual search in luggage screening at the level of the individual. However, no study so far has attempted to address extraneous issues (social, cultural, environmental) that might potentially influence screening efficiency over and beyond those that extend beyond simple visual search processes.

The purpose of this study was to examine what effect, if any, variables such as race, age and gender of the passenger have on the screener's decisions to stop the passenger's luggage or not. Computer simulation was used instead of observing actual luggage screeners so that the study could be more controlled than would be possible in the real environment. Simulation also allowed the study to be run using the same luggage images for several student "screeners" allowing comparison between different screeners and luggage images.

1.1 Social –Cognitive Biases
1.1.1 Age
Age bias is a social bias related to a person's age that can have an effect on decision making. Older people often tend to be discriminated against for jobs. Specifically, the belief is that older individuals are not as flexible in their thinking as younger individuals. Therefore a job that requires flexibility would not be a good fit for an older worker, [8] whereas younger people are believed to have more potential for development than the older people [9]. Based on this, younger people may be more likely to be employed as airport luggage screeners, as their thinking must be very flexible to figure out what constitutes a target.

1.1.2 Gender
When one gender is given preferential treatment over the other, it is typically referred to as "gender bias" [10]. Gender bias is pervasive especially in the workplace. When men and women are evaluated for the same type of work male workers are often found to get better rewards for good evaluations compared to female workers; on the flip side, males also receive harsher punishments than females in response to poor evaluations [11]. Research has revealed that performance ratings are more strongly related to promotions for female workers compared to male workers, which suggests that females are held to higher standards than males [12]. For example, in one study wherein men and women were fired from similar jobs, men received more compensation than women [13].
Clearly, gender-related biases play a major role when decisions to hire, promote or fire are made in several job contexts.

1.1.3 Race
Though we would like to think differently, racial bias is still prevalent throughout the world. There have been numerous studies looking at racial bias among police and their decision to shoot or not shoot [14]. In the Correll et al. (2007) [14] study, comparing police to civilians in the same district, civilians were found to be more likely to shoot when shown a minority suspect compared with the police. Both police and civilian participants took longer to react when the White suspect had a gun, and the minority suspect did not have a gun. The researcher concluded that seeing a white person with a gun violated people's expectations leading them to take longer to react; the opposite was true when observing a person of minority race who was perceived as dangerous even without a weapon [14]. The police officers and

783

civilians were White, Black, Native American, and Hispanic so there was a mix of races.

2.0 The Luggage Screening Study
This study was designed to examine whether the social-cognitive biases described above would influence the decision making process in an airport security screening context. What makes this study unique is the focus on social-cognitive biases which differs from existing studies that have focused on either the luggage screening process [15, & 16] or on the decision making made by the luggage screener [17, 18, 19]. This study was designed to examine whether these biases will influence active decision making during the luggage screening process. As mentioned earlier, we implemented a laboratory-based experimental task along with a luggage screening simulation to study this.

3.0 Method
3.1 Participants
Participants were 96 Old Dominion University undergraduates completing the study for class credit. The study took approximately 1 hour to complete, for 1 hour of research credit.

3.2 Materials
Gateway computers were used, which were running Microsoft XP with service pack 2. These computers were used to run a computer simulation of airline luggage screening created by E-prime 2.0.

3.3 Procedure
Participants were randomly assigned to a control group, (n=24), and three experimental groups (n= 72) in a 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, no-anchor, post-anchor) design. Participants filled out an entrance questionnaire prior to running the study. The task was for participants to detect

the presence of dangerous objects in x-ray images of passenger luggage. Participants scanned 200 images distributed into two blocks of 100 images each. At the beginning of each block, participants were shown the five targets that they needed to look for in the 100 bags that were to follow. For the experimental groups, the appearance of the luggage image on each trial was preceded by the picture of a random passenger (drawn from a new set of 100, that includes the following races: White, Black, Asian, Middle Eastern, and Hispanic) to whom the bag supposedly "belongs". For each of the experimental groups, half the participants performed the task under high time pressure (2 seconds for each luggage image exposure) and the other half performed under low time pressure (6 seconds for each luggage image exposure). After deciding whether to pass the bag or not, participants rated their confidence in their decision on a five point scale.

Participants in the *pre-anchor group* (n = 24) were first required to answer two statements about the passenger *before* the x-ray image appears. After answering the statements, they clicked "next" and the x-ray image was brought up onto the screen, after which, they rated their confidence on their decision of whether or not to pass the bag. The two statements that were used were statement #1:*"I think this person is attractive"* and statement #2 *"I will most likely stop this person's luggage."* These two statements appear to be the most powerful indicators of the existence of such cognitive biases.

For participants in the *no-anchor group* (n = 24), after 4 seconds of the passenger appearing the x-ray image of a bag appeared beside the passenger. These participants were not required to answer any questions about the passengers, but they still rated their

confidence on their decision to pass or not pass the luggage.

For the *post-anchor group* (n = 24), the program ran the same experimental procedure as for the no-anchor group. However, participants were required to answer the two statements answered by the pre-anchor group about each passenger *after* the participant has chosen whether or not to pass the bag and rated their confidence in that decision. Once they have answered the questions and clicked "next", the next picture of a passenger appeared. This procedure continued until the end of the trial block.

A control group (n = 24) performed the screening task alone without observing the pictures of passengers. Of these 24 participants, 12 participants performed under high time pressure and the other 12 performed under low time pressure. This group served as a baseline for performance under each level of time pressure without the additional anchoring information provided by the passengers' pictures.

The base rate for the targets was 50% for all groups. Participants were not informed about the base rate. At the end of each trial, participants received feedback in the form of a text message, telling them whether they made a correct decision or not. Also they received a cumulative percent correct score shown after each decision to pass or not pass the bag. At the end of the experiment, participants filled out a final "task knowledge" questionnaire. The participant with the highest score for their experiment session received a piece of candy as a prize.

4.0 Data Analysis
The data was analyzed for normality. If normality is violated, box plots were used to examine which sections of the data were outliers, and the outliers were

brought to 2 standard deviations away from the mean. A 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA was ran for each dependent variable. For the interactions that were significant, a mixed measures ANOVA was run, followed by paired t-tests with a bonferroni correction.

The dependent variables of interest were:
- Hit rate - the probability of correctly detecting a target.
- False alarm rate - probability of an incorrect detection when there was no target
- Sensitivity (d') – the perceptual ability to differentiate between a target and non-target.
- Response criterion setting (c) – the propensity to generate "yes" or "no" responses.

The data analytic strategy was based on a two-pronged approach. We used hit rate and false alarm rate as pure performance measures which directly measure a participant's performance on the task. In addition, we used the signal detection variables of sensitivity and response criterion setting to understand the decision making processes that drive performance (resulting in hit and false alarms).

5.0 Results
Due to the complexity of the experimental design, the study was broken up into two different sets of variables. Hit rate and false alarm rate are grouped under "performance analysis", and sensitivity and response criterion setting are grouped under "signal detection analysis".

5.1 Performance Analysis
5.1.1 Hit Rate
All "p" values below .05 are statistically significant. The hit rate data was normally distributed with no outliers, therefore no data cleaning was necessary.

A 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA was used to analyze the hit-rate data. The mixed measures ANOVA revealed that there was a significant main effect for time pressure ($F(1,66) = 56.18$, $p \leq .001$, $\eta^2 = .46$). Participants under low time pressure had higher hit rates ($M = .82$, $SE = .01$) than the participants under high time pressure ($M = .69$, $SE = .01$). All other main effects and interactions were statistically non-significant.

5.1.2 False Alarm Rate
The data set was not normally distributed, and the box plots revealed 12 outliers, which were brought in to within 2 standard deviations of the mean. This made the data set normally distributed. A 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA, similar to that used for the Hit Rate analysis, was used to analyze the False Alarm Rate data. The results of the ANOVA revealed that there was a significant main effect for passenger gender ($F(1, 66) = 7.81$, $p = .007$, $\eta^2 = .11$), and time pressure ($F(1, 66) = 10.80$, $p = .002$, $\eta^2 = .14$). Participants had a significantly higher false alarm rate for male passengers ($M = .16$, $SE = .01$) than they did for the female passengers ($M = .13$, $SE = .01$).

Participants under high time pressure ($M = .19$, $SE = .02$) had significantly more false alarms than did the participants under low time pressure ($M = .11$, $SE = .02$). All other main effects and interactions were statistically non-significant.

5.2 Signal Detection Analysis
5.2.1 Sensitivity: d'
Sensitivity, also known as discriminability index, is a measure of how far apart the signal and noise curves are for an individual (Heeger, D., 1997). In other words, this implies that the more the signal (or, target) stands out from back ground clutter, the easier it will be for the human to locate the target. So, in this experiment, higher sensitivity implies that it was easier for the participant to distinguish the target from non-targets. Specifically, the higher the sensitivity, the better was the detection performance.

The sensitivity data was normally distributed with no outliers, therefore no data cleaning was necessary. A 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA was used to examine the data obtained for sensitivity. The main effect of time pressure ($F(1, 66) = 47.34$, $p \leq .001$, $\eta^2 = .418$) and the interaction between passenger gender, passenger race, and anchor ($F(8, 264) = 3.34$, $p = .001$, $\eta^2 = .092$) were both significant. Under low time pressure ($M = 2.23$, $SE = .07$) participants had higher sensitivity than did participants under the high time pressure ($M = 1.54$, $SE = .07$).

To further analyze the relationship between passenger gender and passenger race within each anchor group, a 2 (gender) X 5 (race) mixed measures ANOVA was run within each

786

of the anchor groups and is described in the following sections.

5.2.1.1 Pre-anchor
All of the main effects were non-significant which include the following: passenger gender and passenger race. The interaction between passenger gender and passenger race was significant ($F(4, 92) = 2.863$, $p = .028$, $\eta^2 = .102$). Sphericity was violated, and by using the Greenhouse-Giesser ($p=.063$), Huynh-Feldt ($p=.057$), and the Lower Bound ($p=.104$) correction the interaction became statistically non-significant. All of the other interactions were found to be non-significant which include the following: passenger gender by time pressure, passenger race by time pressure, and passenger gender by passenger race by time pressure.

5.2.1.2 No-anchor
All of the main effects were non-significant which include the following: passenger gender and passenger race. The only interaction that was found to be significant was the interaction between passenger gender and passenger race ($F(4, 92) = 2.621$, $p = .04$, $\eta^2 = .102$). Sphericity was violated and by using the Greenhouse-Giesser ($p=.048$), and Huynh-Feldt ($p=.04$) correction the interaction was still statistically significant. However, using the Lower bound ($p=.119$) correction rendered the interaction statistically non-significant. All of the other interactions were found to be non-significant including the following: passenger gender by time pressure, passenger race by time pressure, and passenger gender by passenger race by time pressure.

The only statistically significant difference between male and female passengers was between the White passengers; participants had higher sensitivity for detecting targets when the passengers were male compared to female (male: $M = 1.87$, $SE = .17$; female: $M = 1.45$, $SE = .16$; $t = 2.786$, $p = .011$).

5.2.1.3 Post-anchor
All of the main effects and interactions were non-significant.

5.2.2 Response Criterion Setting: c
Response Criterion Setting is the propensity to generate "yes" or "no" responses. This means that the human sets an arbitrary threshold or "cutoff point" for responding; when the signal to noise ratio is perceived as being above this level, the participant will indicate a target is present. Likewise, if the ratio is perceived as being below this threshold, they will indicate that a target is not present (Heeger, D., 1997). Typically, if the participant sets his/her response criterion high such that the criterion setting is high or positive, responding is said to be conservative. This means than the participant has a propensity to say "no" more often than "yes". The opposite occurs when a participant sets his/her response criterion low. In such cases, responding is said to be more liberal; this will result in low or negative criterion settings and a general tendency to say "yes" more frequently than "no".

The data set was not normally distributed, and the box plots revealed 12 outliers, which were brought in to within 2 standard deviations of the mean. This made the data set normally distributed. A 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA was used to analyze the response criterion setting data. The ANOVA indicated a significant main

787

effect of passenger race ($F(4, 264) = 8.48$, $p \leq .001$, $\eta^2 = .114$) and an interaction between passenger gender, time pressure, and anchor ($F(2, 66) = 3.50$, $p = .036$, $\eta^2 = .096$). Participants had significantly more conservative response criteria for passengers of Hispanic race ($M = 1.19$, $SE = .09$) compared to all the other races (White $M = .85$, $SE = .05$, $t = 3.97$, $p \leq .001$; Black $M = .83$, $SE = .05$, $t = 4.33$, $p \leq .001$; Asian $M = .82$, $SE = .05$, $t = 4.35$, $p \leq .001$; Middle Eastern $M = .92$, $SE = .06$, $t = 3.14$, $p = .002$). This means that participants were less likely to say there was a target present when confronted with a Hispanic passenger relative to passengers of other races.

To further examine criterion settings within anchor groups, a 2 (gender) X 2 (time pressure) mixed measures ANOVA was run within each of the anchor groups described below.

5.2.2.1 Pre-anchor and Post-anchor
All of the following main effects were non-significant: passenger gender and time pressure. The interaction between passenger gender and time pressure was found to be non-significant as well.

5.2.2.2 No-anchor
For this group all the main effects were non-significant which include passenger gender and time pressure. The interaction between passenger gender and time pressure ($F(1, 22) = 8.391$, $p = .008$, $\eta^2 = .276$) was found to be statistically significant. One tailed t tests were used for post hoc analysis of the interaction. The t-tests revealed that there was a non-significant difference for participants' response criterion setting for the male versus female passengers under low time pressure. However, under high time pressure criterion setting for male passengers ($M = 1.14$, $SE = .71$) was significantly

higher than for female passengers ($M = .84$, $SE = .70$, $t = 2.18$, $p = .036$).

6.0 Discussion
Most luggage screening studies to date have focused on either mechanics of the luggage screening process [15, 16] or on the decision making of luggage screeners [17, 18, & 19]. What has seldom been addressed in these studies, in particular the decision making studies, is a consideration of extraneous factors, namely social-cognitive variables, that can affect the decision making process. One of these factors is the passenger himself/herself, and any biases the screener may have towards the passengers. The purpose of this study was to examine whether such social-cognitive biases as gender bias, and racial bias would influence decision making during the luggage screening process. We were also interested in examining the role of time pressure, and if the screening process would be affected by decision heuristics such as anchoring.

6.1 Role of anchoring
While time pressure played a significant role in the results, we found that anchoring also played a significant role in impacting decision making. Anchoring is the tendency for decision makers to focus on one particular piece of information and use that to base subsequent decisions [20]. The anchoring heuristic works by giving people a reference point to help them make a decision. For example, in an early experiment on anchoring, when asked a question, "is the percentage of African countries in the United Nations greater than or less than a 25 percent?" [20] Participants generally used the "25 percent" to base their judgment of exactly what percentage of African countries is in the United Nations. This worked even when the percentage was randomly selected in front of the

participant. In general, if an anchor is present, the anchor can influence the decision making process of a participant, and therefore influence overall performance.

In this study, the "anchor" was a series of questions drawing attention to the passenger to whom the luggage belonged. Results revealed that when participants had the anchor, either before (pre-anchor) or after (post-anchor) they saw the luggage image, it appeared to suppress rather than enhance the social-cognitive biases, relative to the participants in the no-anchor group. The results also revealed significant interactions between cognitive anchoring and race and gender of passengers on performance. Contrary to our initial expectations and hypothesis, this anchoring effect was particularly salient when time pressure was low and participants had more time to 'attend to' the passengers.

The results suggest that participants used their personal biases as 'anchors' to help in the decision making process, particularly when they had time to pay more attention to passengers. Research has revealed that minority races, such as Hispanics, have been associated with negative behavioral connotations. For instance studies of police officers and their decisions to shoot or not shoot [14], have demonstrated that police were more likely to shoot suspects of minority races even when they did not have a gun. The higher hit rate associated with the Hispanic male passengers in our study could possibly be due to the interaction of these social-cognitive biases. Based on the surmise that the participant already had a negative association with male members of minority races, it is possible that they were more suspicious of the two passenger categories (men and minority races) during the luggage screening process. Therefore, when

searching through the x-ray image, they perhaps used gender and race as decision heuristics, paid more attention to the items in bags that were accompanied by male passengers of Hispanic race, and detected the targets more accurately when they were indeed present. This actually suggests a potential benefit of social-cognitive biases in this instance! However, it must be noted that this effect was only observed under conditions of low time pressure when there was ample time to attend to the bags.

The existence of social-cognitive biases in detection behavior is supported, albeit in a slightly different manner, by the false alarm analysis as well. Similar to the effects found in the hit rate data, male Hispanic passengers had a higher false alarm rate associated with them than female Hispanic passengers. Interestingly, the false alarm effect was found under conditions of high time pressure rather than low time pressure. This indicates the negative effects of social-cognitive biases. Although target detection was benefited to an extent due to anchoring under low time pressure, high time pressure led to negative effects in the form of higher false alarms.

Similar effects for racial bias were found in participants' criterion settings wherein participants had a more conservative response criterion setting for certain passenger races. This means that participants were more conservative and had to have a higher subjective evidence of a target being present before they would indicate that one was present. This is very interesting since we have already seen in the false alarm rate data that participants also had a higher false alarm rate for the male Hispanic passengers compared to the other races of passengers. At first glance the conservative criterion setting for Hispanic passengers appears to

789

contradict the finding that participants stopped luggage more (i.e., said "target present" more) in response to these passengers. Is it possible that participants' lower response criterion for the female Hispanic passengers relative to male Hispanic passengers has raised the criterion setting for the Hispanic passengers overall, although this is not evident in a statistically significant difference between the male and female Hispanic passengers per se. As hypothesized, the participants had higher false alarm rates for minority passengers than they did for the White passengers.

As hypothesized, participants had a higher false alarm rate when the passenger was male which would lead them to being stopped more. Also the interaction between passenger gender and time pressure for the no-anchor group was an interesting indication of how not providing an anchor significantly impacted performance more than providing anchors in this study. When time pressure was low, participants had a more liberal response to the male passengers thereby stopping the luggage belonging to male passenger more often. Conversely, participants had a more conservative response towards the female passengers, thereby stopping their luggage with lower frequency than for male passengers. Surprisingly, the opposite became true under high time pressure; participants had a higher, more conservative response to the male passengers, while they had a more liberal response to the female passengers. It is possible that when participants had time to think about the passenger and the luggage, as in the case of the low time pressure group, their biases against male passengers were mitigated to an extent leading them to become more conservative. The opposite might be true for female passengers wherein the index of

suspicion possibly increased with the availability of more time to scan the image.

7.0 Conclusions

The results of this research have demonstrated how social-cognitive biases affect people in the real world and how they can subsequently impact the luggage screening process and eventually national security. Through the use of computer simulation we have shown that social-cognitive biases actually do have an effect on the detection of anomalies during luggage screening wherein decision makers use these inherent biases as decision heuristics, particularly under conditions of time pressure. Clearly, such biases would be difficult to detect through mere observation of screening processes at airports. Hence, the use of behavioral experimental and computer simulation is invaluable in such sensitive contexts.

Most importantly, our results revealed a clear relationship between decision making process and performance. Through the use of both signal detection variables and performance variables in our analyses, we are able to draw conclusions not just about the impact of social-cognitive variables on performance, but also the processes that led to the observed behaviors. This is especially important in the current security conscious world we live in and for training of personnel for optimal decision making that is free of biases and prejudices. An associated goal of this research is to the design community for improving the design of imaging equipment and luggage screening stations.

8.0 References
[1] Grahame, M., Laberge, J., & Scialfa, C. T. (2004). Age differences in search of web pages: The effects of link size, link number, and clutter. *Human Factors, 46,* 385-398.
[2] McPhee, L. C., Scialfa, C. T., Dennis, W. M., Ho, G., & Caird, J. K. (2004). Age differences

in visual search for traffic signals during a simulated conversation. *Human Factors, 46,* 674-685 [10] Baker, K., Craddock, A., &Orwig, A. (2002) *Gender Bias*. An educator's guide to access issues. Retrieved July 19, 2008. http://www.ed.uiuc.edu/wp/access-2002/gender_bias.htm.

[3] Verghese, P., & McKee, S. P. (2004). Visual search in clutter. *Vision Research, 44,* 1217-1225.

[4] Nagy, A. L., Neriani, K. E., & Young, T. L. (2005). Effects of target and distractor heterogeneity on search for a color target. *Vision Research, 45,* 1885-1899.

[5] Bravo, M. J.,& Farid, H. (2006). Object recognition in dense clutter. *Perception & Psychophysics, 68,* 911-918.

[6] Bravo, M. J., & Farid, H. (2004). Search for a category target in clutter. *Perception, 33,* 643-652.

[7] Neider, M. B. (2007). Set-size effects during visual search in realistic scenes. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 68,* 655. Abstract retrieved June 15, 2008 from PsychINFO Database.

[8] Craik, F. I. M. (2002). Human memory and aging. In Psychology at the turn of the millennium, Vol. 1: Cognitive, biological, and health perspectives (pp. 261-280). Bäckman, Lars; von Hofsten, Claes; Hove, England: Psychology Press/Taylor & Francis.

[9] Diekman, A. B., & Hirnisey, L. (2007). The effect of context on the silver ceiling: A role congruity perspective on prejudiced responses. *Personality and Social Psychology Bulletin, 33,* 1353-1366.

[10] Baker, K., Craddock, A., &Orwig, A. (2002) *Gender Bias*. An educator's guide to access issues. Retrieved July 19, 2008. http://www.ed.uiuc.edu/wp/access-2002/gender_bias.htm.

[11] McKay, S., & Tate, U. (2001). Student attitudes regarding gender bias in performance evaluations of salespeople. *Journal of Business and Psychology, 16,* 249-258.

[12] Lyness, K. S., & Heilmann, M. E. (2006). When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology, 4,* 777-785.

[13] Rollings-Magnusson, S. (2004). Gender implications of wrongful dismissal judgments in Canada. *Canadian Review of Sociology and Anthropology, 41,* 27-45.

[14] Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology, 92,* 1006-1-23.

[15] Hilscher, M. B. (2005). Performance implications of alternative color-codes in airport X-ray baggage screening. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66,* 3449. Abstract retrieved June 20, 2008 from PsychINFO Database.

[16] McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. C., & Boot, W. R. (2004). Visual skills in airport-security screening. *Psychological Science, 15,* 302-306.

[17] Brown, J. R., & Madhavan, P. (2008). Effects of target base rates on visual search performance: A comparison of two scenarios. *Proceedings of the Human Factors and Ergonomics Society 52nd annual meeting,* 333-337.

[18] Madhavan, P., Gonzalez, C., & Lacson F. C. (2007). Differential base rate training influences detection of novel targets in a complex visual inspection task. *Proceedings of the Human Factors and Ergonomics Society 51st annual meeting,* 392-396.

[19] Parasuraman, R., Warm, J. S., & Dember, W. N. (1987). Vigilance: Taxonomy and utility. In L.S., Mark, J.S., Warm, & R. L. Huston (Eds.), *Ergonomics and human factors* (pp. 11-32).

[20] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185,* 1124-1131.

791

SOCIAL-COGNITIVE BIASES IN SIMULATED AIRLINE LUGGAGE SCREENING

Jeremy Brown
Advisor:
Poornima Madhavan

Introduction

- ◉ **Previous studies**
  - › X-ray color[1]
  - › Search Patterns[2]
  - › Textual/Spatial Cuing[3]

- ◉ **Influence of Social-Cognitive Biases**
  - › Implicit Attitudes Test[4].

1. Ghylin, Schwaninger, Drury, Redford, Lin, and Batta, 2008
2. McCarley, Kramer, Wickens, Vidoni, & Boot, 2004
3. Wiegmann, McCarley, Kramer, and Wickens, 2006
4. Greenwald, McGhee, and Schwartz, 1998

Visual Search Tasks

- goal of visual search tasks

- Effects of Visual Clutter
  - Detection time and perceived difficulty[5]
  - Object recognition[6]

5. McPhee, Scialfa, Dennis, Ho, & Caird, 2004; Grahame, Laberge, & Scialfa, 2004
6. Bravo & Farid, 2006

Time Pressure

- Change in Response Pattern

  - Chess[7]

  - Word Recognition[8]

  - Simulation: Trucking game[9]

7. Harreveld, Wagenmakers, & van der Maas, 2007
8. Light, Chung, Pendergrass, & Van Ocker, 2006
9. Betsch, Fielder, and Brinkman, 1998

Social-Cognitive Biases

- Age
  - Flexibility[10]
  - Potential[11]
  - Facial Recognition[12]
- Gender
  - Rewards and punishments[13]
  - Performance rating[14]

10. Craik, 2002
11. Diekman, & Hirnsey, 2007
12. McKay, & Tate, 2001
13. Anastasi, & Rhodes, 2006
14. Lyness, & Heilmann, 2006; Rollings-Magnusson, 2004

Social-Cognitive Biases

- Race
  - Shoot/Don't shoot scenarios[15]

- Impact of Social-Cognitive Biases on Luggage Screening
  - Young Screeners vs. Old Screeners
  - Males vs. Females
  - Screeners race vs. Passengers race

15. Correll, Park, Judd, Wittenbrink, Sadler, & Keesee, 2007; Plant, & Peruche, 2005

The Luggage Screening Study

- Purpose
  - To examine if Social-Cognitive Biases would influence active decision making

Hypothesis

- Time Pressure and Anchor
 - High time pressure vs. low
  - Pre-, No-, and Post-anchor
- Passenger Gender and Race
  - Male passengers will be stopped more
  - Minority races stopped more

Methods

- Participants
  - 96 Old Dominion University Students
- Materials
  - Gateway computers Running Windows XP SP2
  - Simulated Airline Luggage Screening
- Procedure
  - Control Group (n=24)
  - 3 Experimental Groups (n=72)

Control Group Procedure
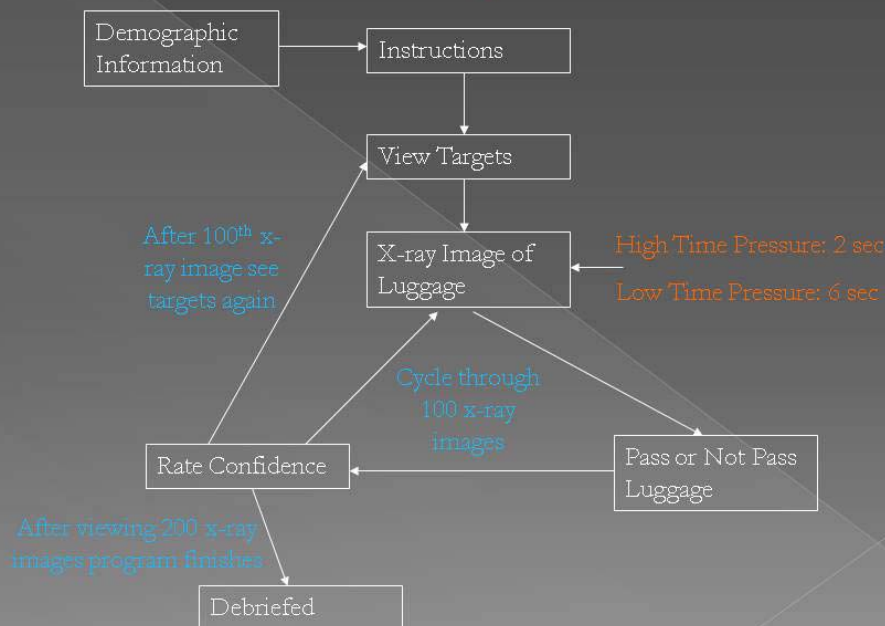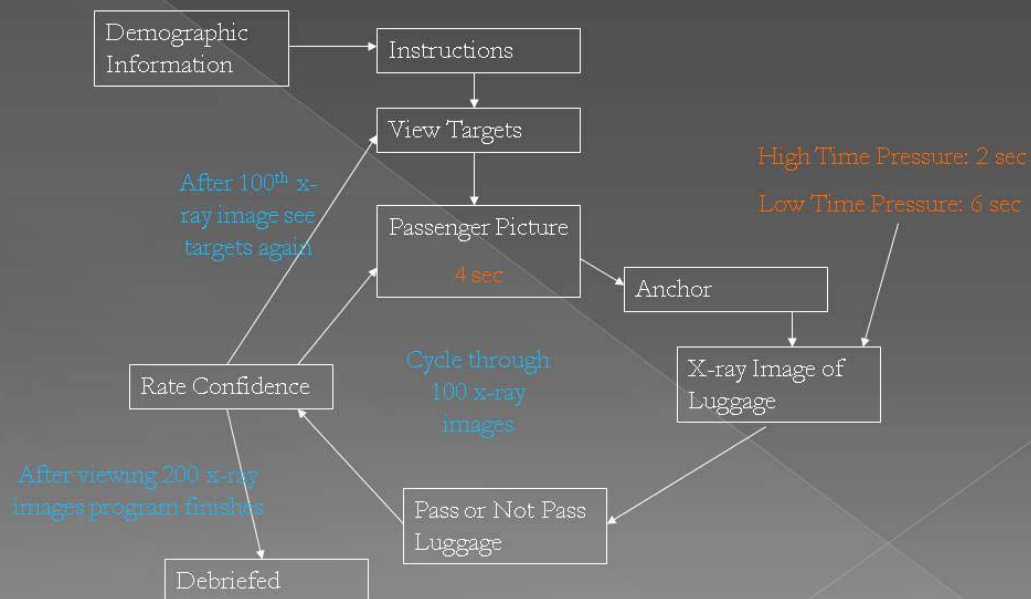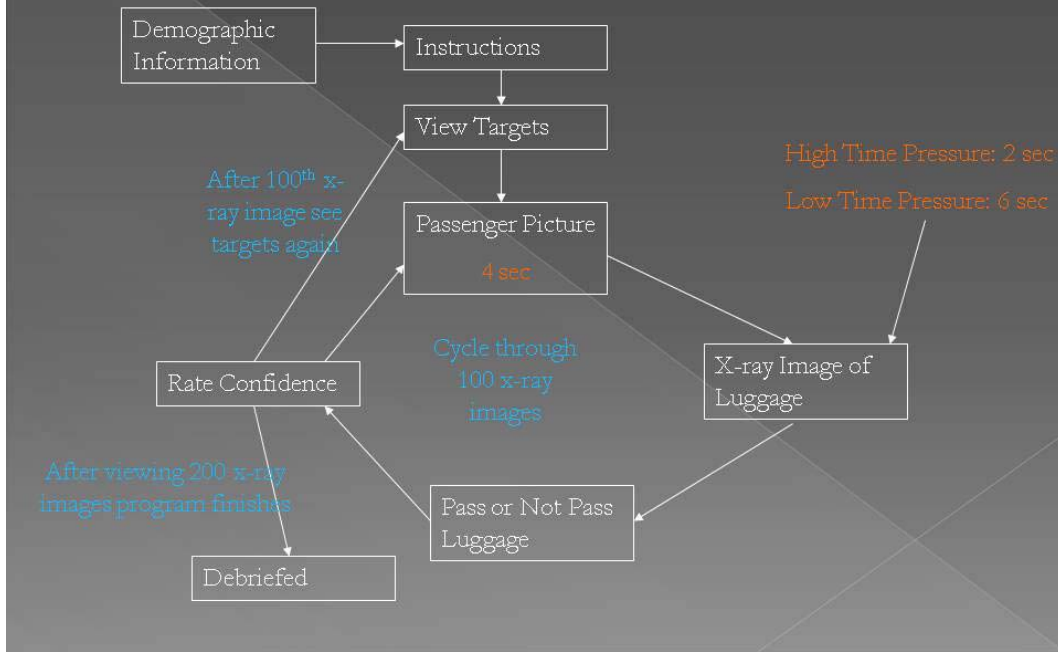
Experimental Groups

- Anchor
  - *"I think this person is attractive."*
  - *"I will most likely stop this person's luggage."*

  - Pre-anchor (n=24)
  - No-Anchor (n=24)
  - Post-Anchor (n=24)

Pre-Anchor Group Procedure
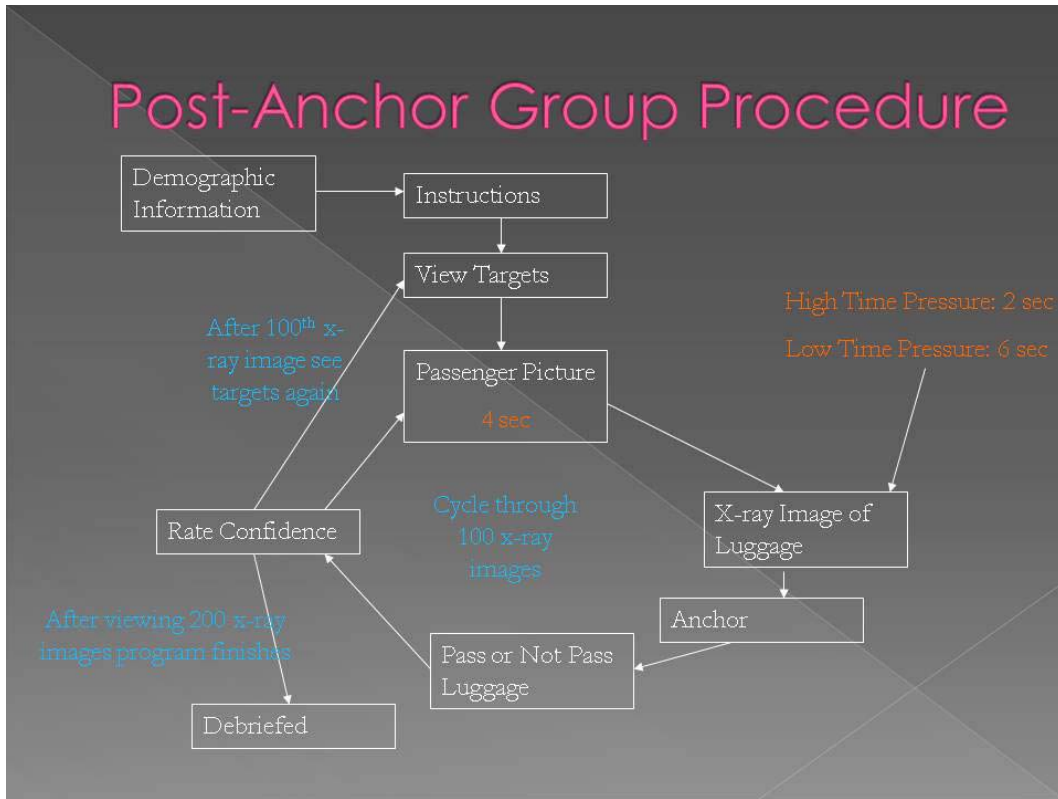
Variables of Interest

- Hit rate - the probability of correctly detecting a target.

- False alarm rate - probability of incorrectly saying there is a target present, when there is no target

- Sensitivity (d') – the perceptual ability to differentiate between a target and non-target.

- Response criterion setting (c) – the propensity to generate "yes" or "no" responses.

Sensitivity (d')
Response Criterion Setting
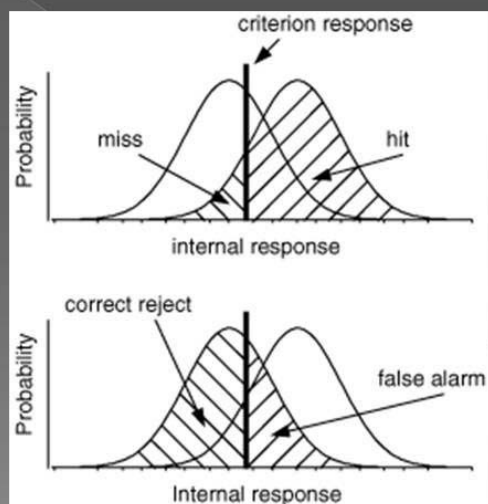
Results

- Performance Analysis
  - Hit Rate
  - False Alarm Rate
- Signal Detection Analysis
  - Sensitivity (d')
  - Response Criterion Setting (c)

Hit Rate

- 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA

- **Time Pressure:** $F(1,66) = 56.18$, $p \leq .001$, $\eta 2 = .46$

Hit Rate

- Non-significant main effects and interactions

Source				
Main Effects	**df**	**F**	**Sig**	**η^2**
Passenger Gender	1, 66	0.66	0.42	0.01
Passenger Race	4, 264	0.713	0.58	0.011
Interactions	**df**	**F**	**Sig**	**η^2**
Passenger Gender X Time pressure	1, 66	0.17	0.68	0.003
Passenger Gender X Anchor	2, 66	0.16	0.86	0.005
Passenger Gender X Time Pressure X Anchor	2, 66	1.01	0.37	0.03
Passenger Race X Time Pressure	4, 264	0.12	0.98	0.002
Passenger Race X Anchor	8, 264	0.96	0.47	0.03
Passenger Race X Time Pressure X Anchor	4, 264	0.71	0.68	0.02
Passenger Gender X Passenger Race	4, 264	1	0.41	0.015
Passenger Gender X Passenger Race X Time Pressure	4, 264	0.67	0.62	0.054
Passenger Gender X Passenger Race X Anchor	8, 264	1.89	0.062	0.054
Passenger Gender X Passenger Race X Time Pressure X Anchor	8, 264	1.71	0.095	0.05

Hit Rate:
Time Pressure Groups

- 2 (gender) X 5 (race) X 3 (anchor) mixed measures ANOVA
  - High Time pressure Analysis
    - All main effects and interactions ns

  - Low Time pressure
    - Statistically Significant interaction between passenger gender, passenger race, and anchor
    - $F(8, 132) = 2.071$, $p = .043$, η2 = .112
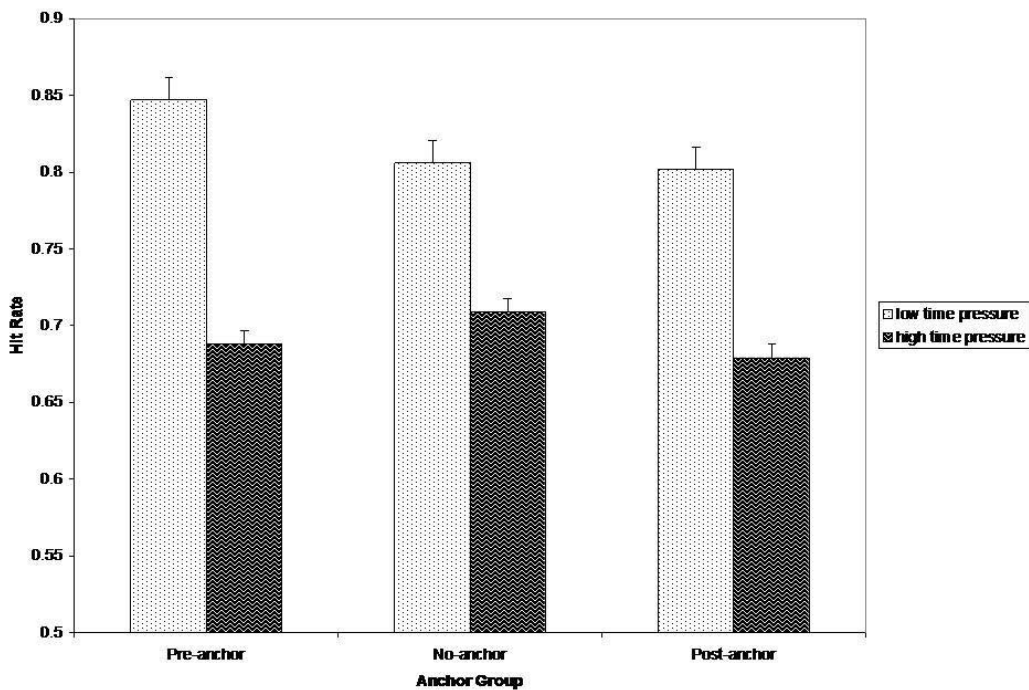
False Alarm Rate

- 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA

- passenger gender
  - $F(1, 66) = 7.81$, $p = .007$, $\eta2 = .11$

- time pressure
 - $F(1, 66) = 10.80$, $p = .002$, $\eta2 = .14$

802

False Alarm Rate

- ns main effects and interactions

Source				
Main Effects	**df**	**F**	**Sig**	**η^2**
Passenger Race	4, 264	0.273	0.9	0.004
Anchor	2, 66	2.01	0.14	0.057
Interactions	**df**	**F**	**Sig**	**η^2**
Passenger Gender X Time pressure	1, 66	1.93	0.17	0.03
Passenger Gender X Anchor	2, 66	0.133	0.88	0.004
Passenger Gender X Time Pressure X Anchor	2, 66	0.171	0.84	0.005
Passenger Race X Time Pressure	4, 264	0.454	0.77	0.007
Passenger Race X Anchor	8, 264	0.201	0.99	0.006
Passenger Race X Time Pressure X Anchor	4, 264	0.636	0.75	0.019
Passenger Gender X Passenger Race	4, 264	1.64	0.16	0.024
Passenger Gender X Passenger Race X Time Pressure	4, 264	1.71	0.15	0.025
Passenger Gender X Passenger Race X Time Pressure X Anchor	8, 264	0.306	0.96	0.009

False Alarm Rate: Time Pressure Groups

- 2 (gender) X 5 (race) X 3 (anchor) mixed measures ANOVA
- High Time Pressure
  - gender
    - $F(1, 33) = 8.395$, $p = .007$, $\eta2 = .20$
  - gender by race interaction
    - $F(4, 132) = 2.430$, $p = .051$, $\eta2 = .07$
- Low Time Pressure
  - passenger gender, passenger race, and anchor
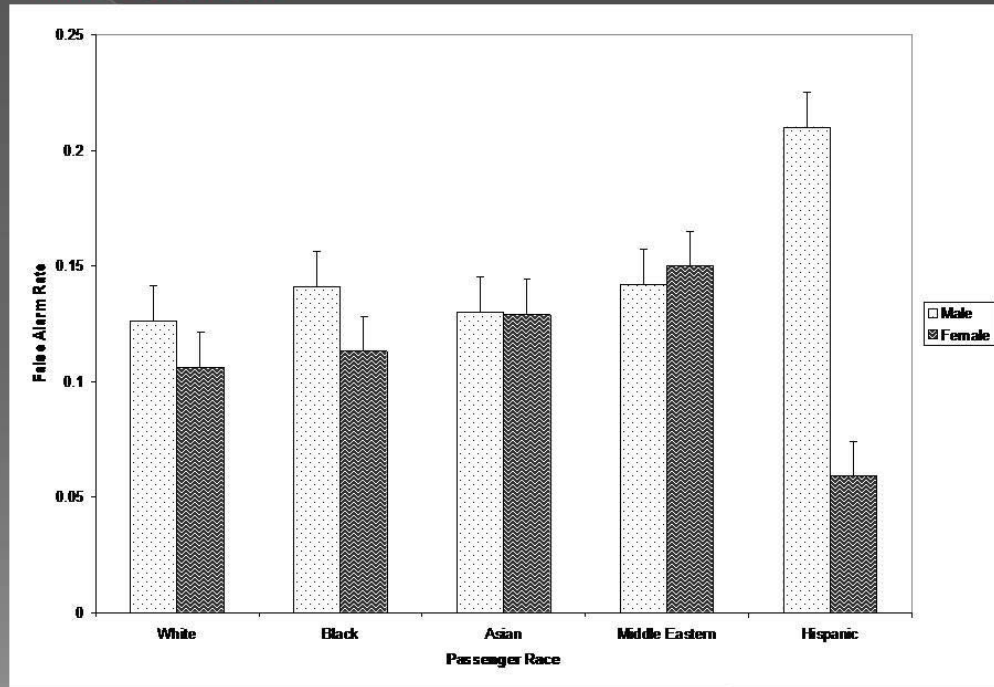   - $F(8, 132) = 2.03$, $p = .05$, $\eta2 = .11$

False Alarm Rate: High Time Pressure
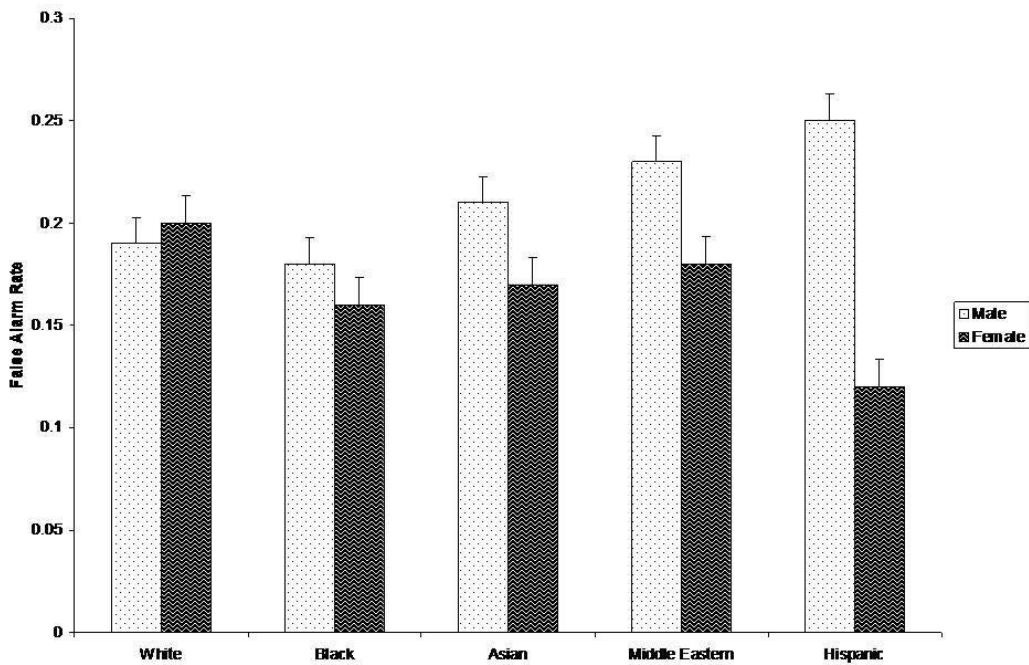
False Alarm Rate: Low Time Pressure interaction

- **Pre-anchor** 2 (gender) X 5 (race) mixed measures ANOVA
  - passenger gender
    - $F(1, 23) = 5.131$, $p = .033$, $\eta2 = .182$
  - passenger gender by passenger race interaction
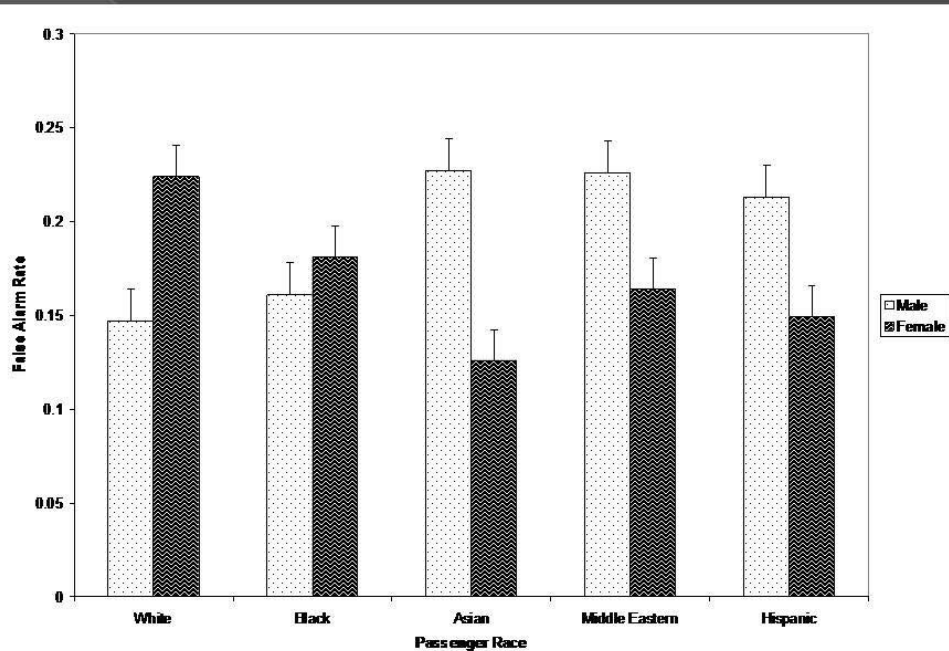   - $F(4, 92) = 3.120$, $p = .019$, $\eta2 = .119$

False Alarm Rate: Low Time Pressure interaction

False Alarm Rate: Low Time Pressure interaction

- No-anchor
  - › passenger gender by passenger race interaction
    - · $F(4, 92) = 3.221$, $p = .016$, $\eta2 = .12$

- Post-Anchor
  - › No significant main effects or interactions

805

False Alarm Rate:
Low Time Pressure interaction

False Alarm Rate:
Anchor Groups

- 2 (gender) X 5 (race) X 2 (time pressure) mixed measures ANOVA
- Pre-Anchor
  - › passenger gender by passenger race
    - $F(4, 88) = 3.132$, $p = .019$, $\eta 2 = .125$
- No-Anchor
  - › time pressure
    - $F(1, 22) = 6.958$, $p = .015$, $\eta 2 = .24$
  - › passenger gender by passenger race
    - $F(4, 88) = 3.145$, $p = .018$, $\eta 2 = .125$
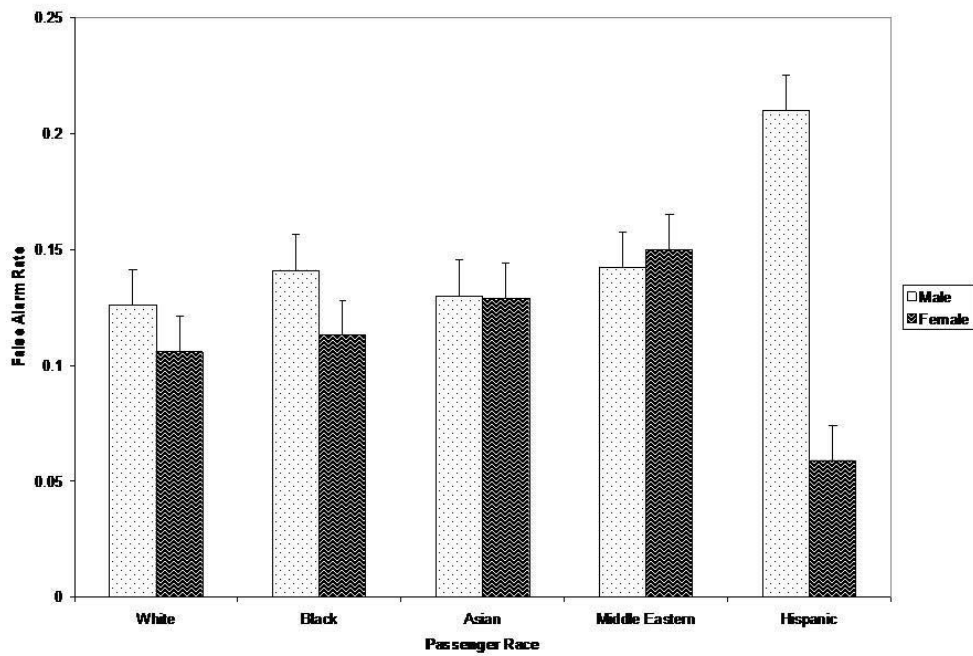- Post-Anchor
  - › No significant main effects or interactions

806

False Alarm Rate: Anchor Groups

False Alarm Rate: Anchor Groups

807
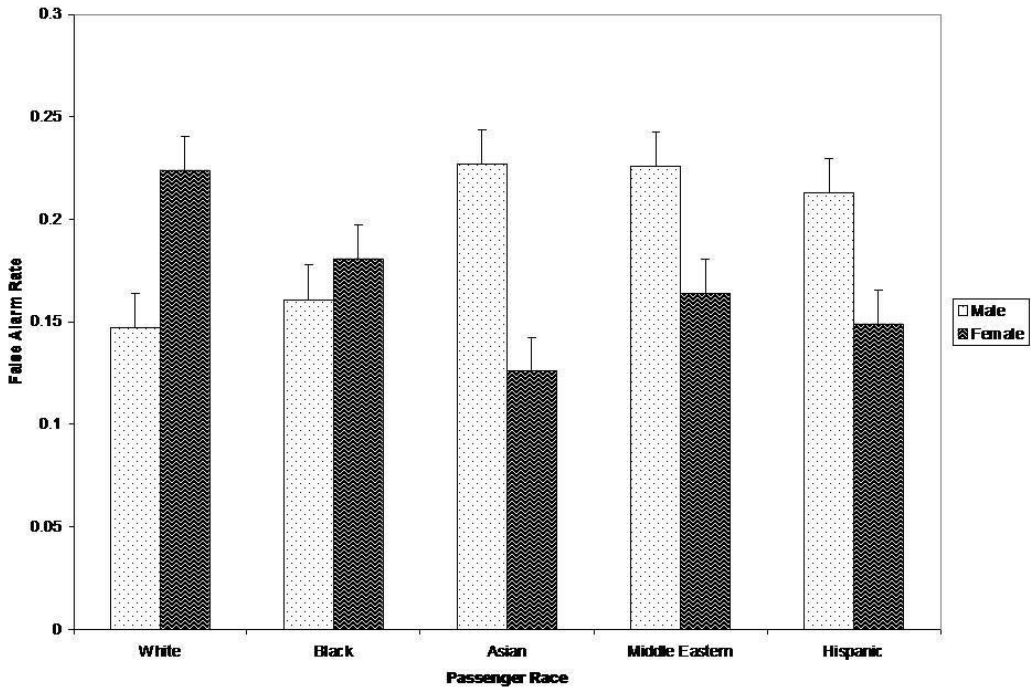
Sensitivity (d')

- 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA

- time pressure
  - $F(1, 66) = 47.34, p \leq .001, \eta 2 = .418$
- passenger gender, passenger race, and anchor
  - $F(8, 264) = 3.34, p = .001, \eta 2 = .092$

Sensitivity (d')

- Non-significant main effects and interactions

Source				
Main Effects	df	F	Sig	η²
Passenger Gender	1, 66	2.99	0.088	0.043
Passenger Race	4, 264	0.707	0.59	0.011
Anchor	2, 66	1.64	0.2	0.047
Interactions	df	F	Sig	η²
Passenger Gender X Time pressure	1, 66	0.329	0.57	0.005
Passenger Gender X Anchor	2, 66	0.32	0.73	0.01
Passenger Gender X Time Pressure X Anchor	2, 66	0.242	0.79	0.007
Passenger Race X Time Pressure	4, 264	0.283	0.89	0.004
Passenger Race X Anchor	8, 264	0.765	0.634	0.023
Passenger Race X Time Pressure X Anchor	4, 264	0.601	0.78	0.018
Passenger Gender X Passenger Race	4, 264	1.22	0.3	0.018
Passenger Gender X Passenger Race X Time Pressure	4, 264	1.65	0.16	0.024
Passenger Gender X Passenger Race X Time Pressure X Anchor	8, 264	1.4	0.2	0.041

Sensitivity (d'): Time Pressure Groups

- 2 (gender) X 5 (race) X 3 (anchor) mixed measures ANOVA
- High Time Pressure
  - passenger gender bypassenger race by anchor
    - $F(8, 132) = 2.144, p = .036, \eta2 = .115$
- Low Time Pressure
  - passenger gender by passenger race by anchor
    - $F(8, 132) = 2.607, p = .011, \eta2 = .136$

Sensitivity (d'): Anchor Groups

- 2 (gender) X 5 (race) X 2 (time pressure) mixed measures ANOVA
- Pre-Anchor
  - time pressure
    - $F(1, 22) = 24.068, p \leq .001, \eta2 = .522$
  - passenger gender by passenger race
    - $F(4, 88) = 2.963, p = .024, \eta2 = .119$
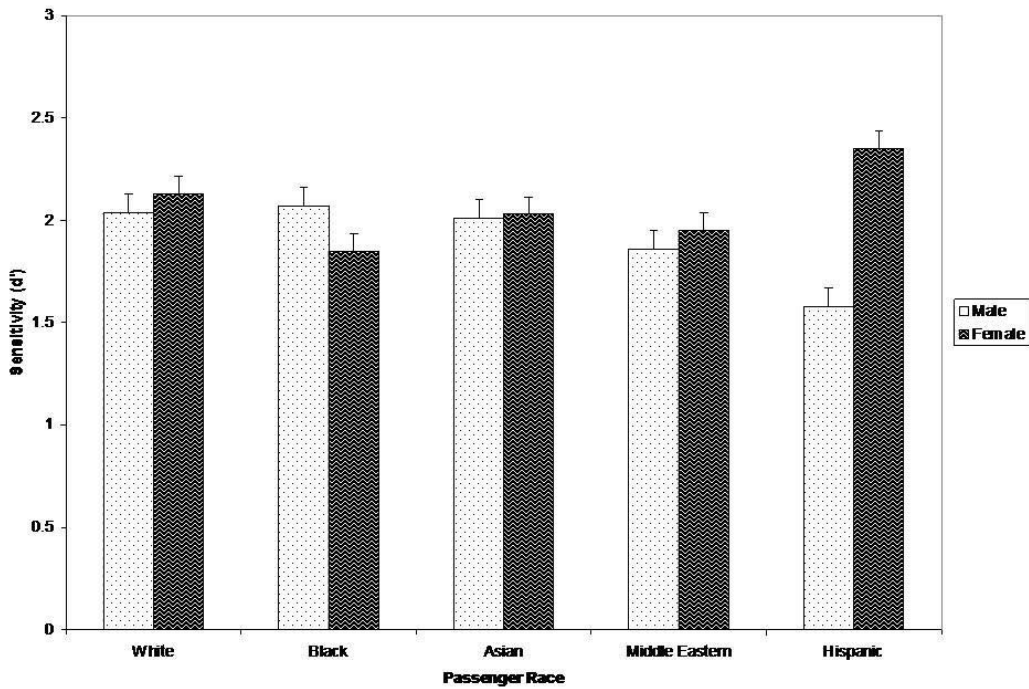
Sensitivity (d'): Anchor Groups

- No-anchor
  - time pressure
    - $F(1, 22) = 27.139, p \leq .001, \eta 2 = .458$
  - passenger gender by passenger race
    - $F(4, 88) = 2.56, p = .04, \eta 2 = .104$

Sensitivity (d'): Anchor Groups
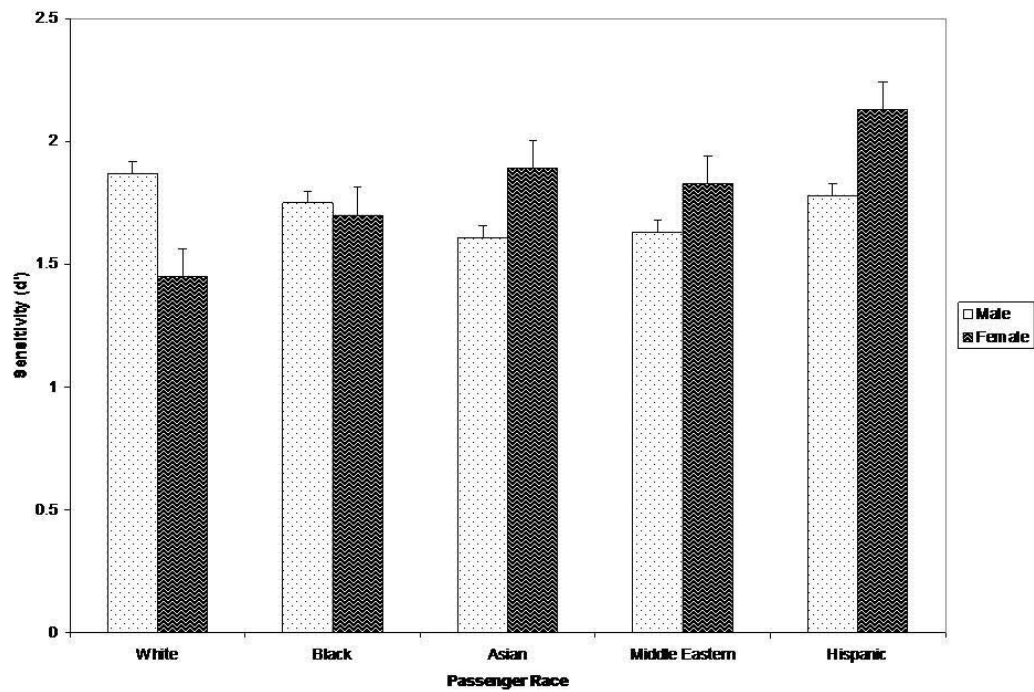
Sensitivity (d'): Anchor Groups

- Post-anchor
  - time pressure
    - $F(1, 22) = 9.717, p = .005, \eta2 = .306$
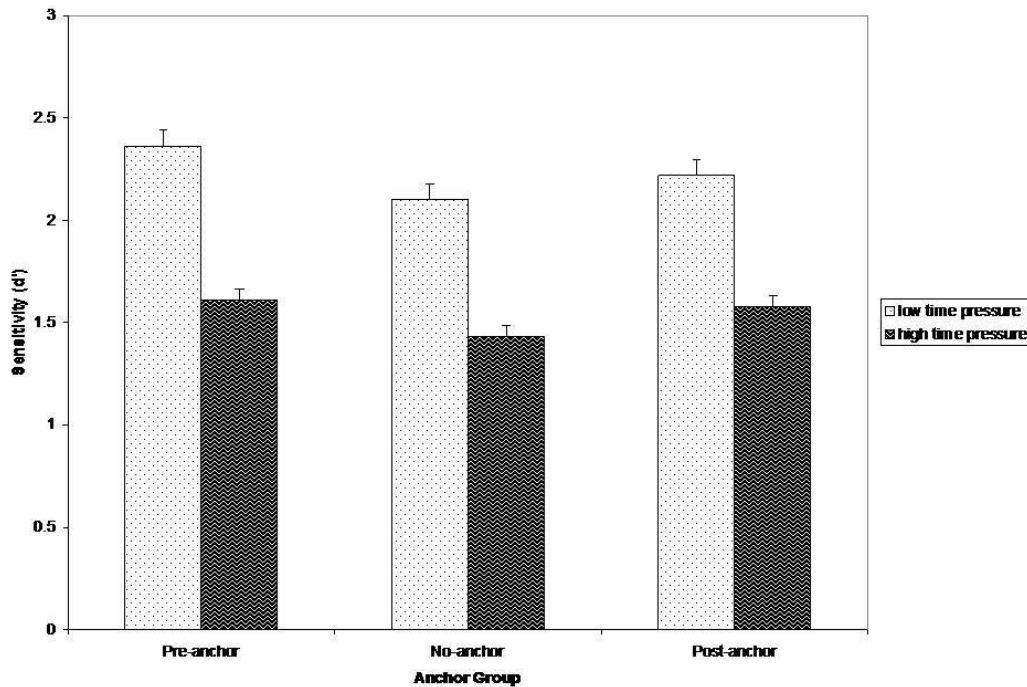
Response Criterion Setting (c)

- 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA

- **passenger race**
  - $F(4, 264) = 8.48, p \leq .001, \eta 2 = .114$
- **passenger gender, time pressure, and anchor**
  - $F(2, 66) = 3.50, p = .036, \eta 2 = .096$

Response Criterion Setting (c)

- Non-significant main effects and

Source				
Main Effects	**df**	**F**	**Sig**	**η²**
Passenger Gender	1, 66	1.8	0.184	0.027
Anchor	2, 66	0.784	0.46	0.023
Interactions	**df**	**F**	**Sig**	**η²**
Passenger Gender X Time pressure	1, 66	3.45	0.068	0.05
Passenger Gender X Anchor	2, 66	0.044	0.96	0.001
Passenger Race X Time Pressure	4, 264	1.33	0.26	0.02
Passenger Race X Anchor	8, 264	0.378	0.93	0.011
Passenger Race X Time Pressure X Anchor	4, 264	0.584	0.79	0.017
Passenger Gender X Passenger Race	4, 264	0.431	0.79	0.006
Passenger Gender X Passenger Race X Time Pressure	4, 264	1.92	0.11	0.028
Passenger Gender X Passenger Race X Anchor	8, 264	1.78	0.08	0.051
Passenger Gender X Passenger Race X Time Pressure X Anchor	8, 264	1.11	0.36	0.032

Response Criterion Setting (c): Time Pressure Groups

- 2 (gender) X 5 (race) X 3 (anchor) mixed measures ANOVA
- High Time Pressure
  - passenger gender
    - $F(1, 33) = 6.41$, $p = .016$, $η2 = .163$
  - passenger race
    - $F(4, 132) = 2.46$, $p = .048$, $η2 = .069$
  - anchor
    - $F(2, 33) = 3.523$, $p = .041$, $η2 = .176$
- Low Time Pressure
  - passenger race
    - $F(4, 132) = 6.56$, $p \leq .001$, $η2 = .166$

Response Criterion Setting (c): Anchor Groups

- 2 (gender) X 5 (race) X 2 (time pressure) mixed measures ANOVA

- ## Pre-anchor
  - > passenger race
    - $F(4, 88) = 2.837, p = .029, \eta2 = .114$

Response Criterion Setting (c): Anchor Groups
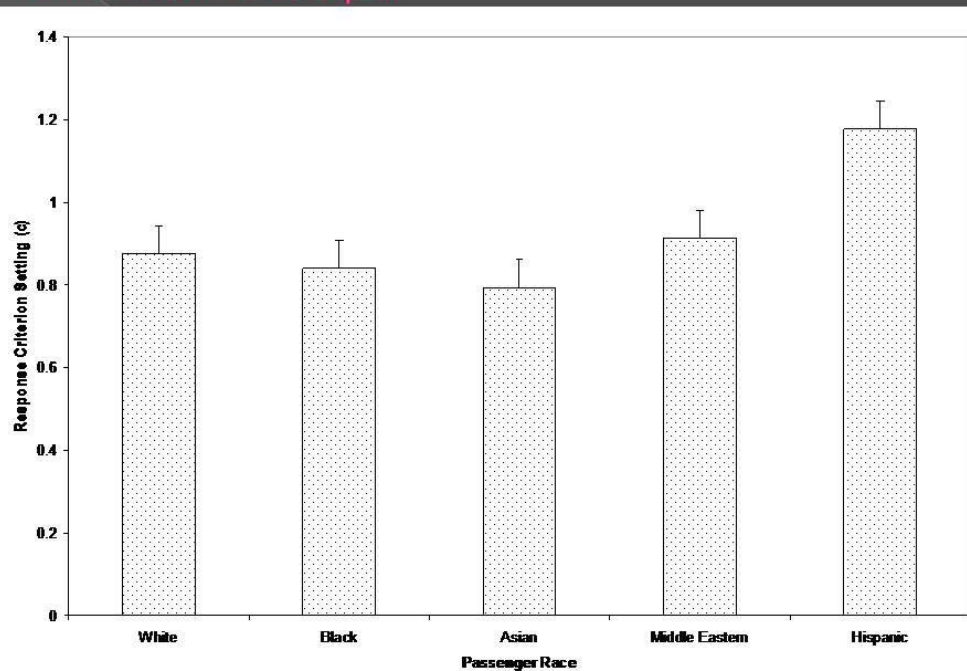
Response Criterion Setting (c): Anchor Groups

- No-anchor
  - passenger gender by time pressure
    - $F(1, 22) = 14.054, p = .001, \eta2 = .390$
  - passenger gender by passenger race by time pressure
    - $F(4, 88) = 14.054, p = .001, \eta2 = .390$).
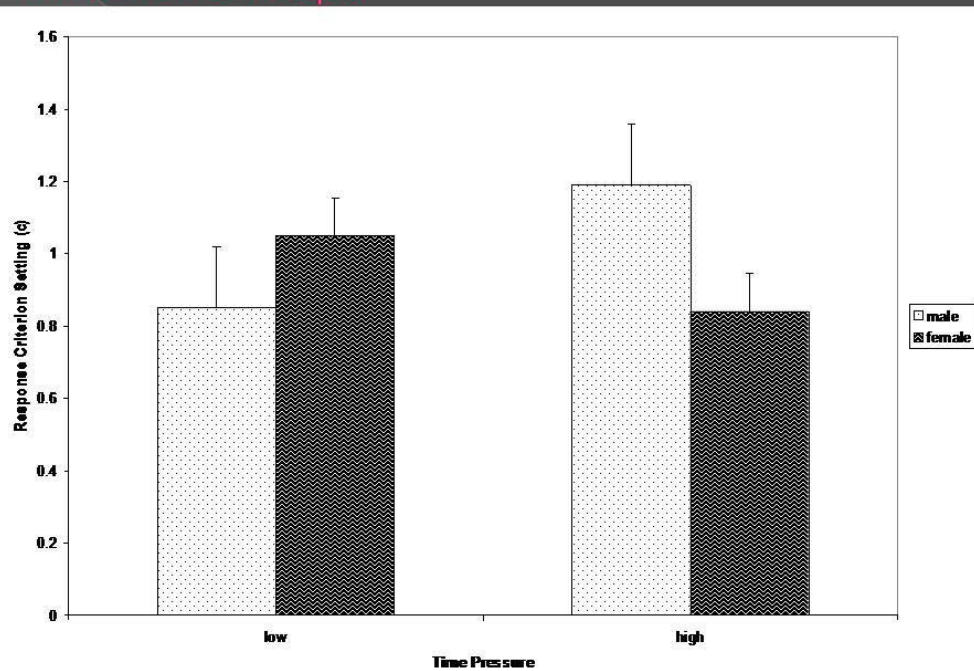
Response Criterion Setting (c): Anchor Groups
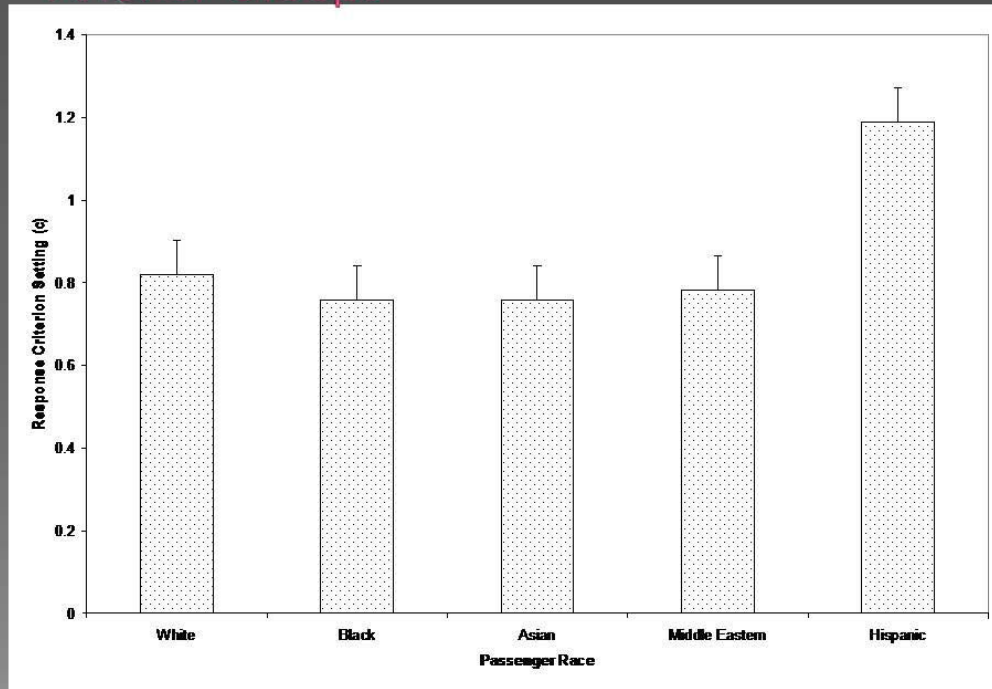
Discussion

- Role of time pressure

- Role of anchoring

Role of Time Pressure

- High time pressure
  - Fewer hits
  - Higher false alarms
  - Less time, can localize but not identify[2]
  - Higher response threshold for males
- Low time pressure
  - Lower response threshold for male passengers
    - Lower response criterion setting

Role of Anchoring

- Pre- or Post-anchor
  - Suppress social-cognitive biases
    - More salient when time pressure was low

- Decision heuristics
  - Gender
  - Race
    - Racial bias: higher false alarm rates for hispanics

Implications for Training and Design

- Luggage screening stations
  - › Increase number of screeners
  - › Block luggage screeners view of passengers
- Social-Cognitive Biases
  - › Can be mitigated
  - › Create training programs

Limitations and Conclusions

- Limitations
  - › Participants
  - › Simulation vs. real world screening
  - › Tangible consequences
- Conclusions
  - › Effects of social-cognitive biases
  - › Design and training

Future Questions Contact email:

jrbrown@odu.edu

5.18   EEG Artifact Removal Using A Wavelet Neural Network

EEG Artifact Removal Using A Wavelet Neural Network

Hoang-Anh T Nguyen, John Musson, Jiang Li and Frederick McKenzie
Old Dominion University
Hnguy025@odu.edu   jmuss003@odu.edu   JLi@odu.edu   rdmckenz@odu.edu

Guangfan Zhang and Roger Xu
Intelligent Automation, Inc.
gzhang@i-a-i.com   hgxu@i-a-i.com

Carl Richey and Tom Schnell
University of Iowa
richeyc@ccad.uiowa.edu   thomas-schnell@uiowa.edu

Abstract- In this paper, we developed a wavelet neural network (WNN) algorithm for EEG artifact removal without EOG recordings. The algorithm combines the universal approximation characteristics of neural network and the time/frequency property of wavelet. We compared the WNN algorithm with the ICA technique and a wavelet thresholding method, which was realized by using the Stein's unbiased risk estimate (SURE) with an adaptive gradient-based optimal threshold. Experimental results on a driving test data set show that WNN can remove EEG artifacts effectively without diminishing useful EEG information even for very noisy data.

1.0   INTRODUCTION

Electroencephalogram (EEG) recordings are known to be contaminated by physiological artifacts from various sources, such as eye blinking or movements, heart beating and movements of other muscle groups [1]. Such types of artifacts are mixed together with the brain signals, making interpretation of EEG signals difficult [2].

Eye movements or blinks usually produce large electrical potentials, which spread across scalp and contaminate EEG recordings. This class of potential generates significant electrooculographic (EOG) artifacts in the recorded EEG. Removal of EOG artifacts is nontrivial because these artifacts spread across the scalp, contaminate and overlap in frequency with the EEG. The effect of EOG artifacts on EEG activity is found most significantly in low frequency bands: Delta (1-4Hz), Theta (4-8 Hz) and Alpha (8-13 Hz) [3]. Eye blinking generates spike-like shapes with their peaks can reach up to 800uV and occur in a very short period, 200-400 ms [4]. Meanwhile, artifacts generated by eye movements are square-shaped, smaller in amplitude but last longer in time, corresponding to lower frequency components [5].

In recent years, there has been an increasing interest in applying various techniques to remove ocular artifacts from EEG [1, 2, 5, 6,8, 10, 13, 14-19]. The methods for removing EOG artifacts based on regression in time domain or frequency domain [8] were widely studied. All regression methods, both in time and frequency domains, rely on EOG recordings, which are however, not always available. Furthermore, these methods usually eliminate the neural potentials which are common to reference electrodes and other frontal electrodes.

Berg and Scherg [10] proposed principle component analysis (PCA) based technique for removing ocular artifacts. In this method, EEG and EOG signals were simultaneously collected. It was observed that PCA of the variance in these signals produced major components corresponding to various eye blinks and eye movements. The artifacts were removed by eliminating these contaminated components. Their experiments proved that PCA removes artifacts more effectively than regression based models. However, PCA models usually failed to completely separate artifacts from cerebral activity [11], and the

820

orthogonal assumption for data components in PCA is hardly satisfied [5].

Independent component analysis (ICA), which was developed for the blind source separation problems, the class of algorithms which decompose mixtures into original sources without any *a priori* knowledge about the mixing process or properties of those sources, has been used as an alternative method for EEG artifact removal [1, 12-14]. ICA for artifact removal usually requires a large amount of data and manual visual inspection to eliminate noisy independent components, making the method time-consuming and not suitable for real-time applications.

Recently, the wavelet-based methods [14-19] for EEG artifacts removal have received significant attention. Wavelet analysis has been used as an effective tool for measuring and manipulating non-stationary signals such as EEG. It provides flexible controls over the resolution with which neuroelectric components and events can be localized in time, space, and scale. The biggest advantage of using this method for EEG correction is that it does not rely on neither the reference EOG signal nor visual inspection.

This paper proposes a novel, robust, and efficient technique to remove EEG artifacts by combining the approximation capabilities of both wavelet and neural network methods. The method can be described briefly as the following (1) contaminated EEG signals are first decomposed to a set of wavelet coefficients, (2) low frequency wavelet sub-band coefficients are then passed through and corrected by a trained neural network and (3) the corrected coefficients are used to reconstruct clean EEG signals. The method was applied to correct EEG data contaminated by ocular artifacts and compared with other state-of-the-art methods including ICA and a wavelet thresholding method.

The rest of the paper is organized as follows: Section 2 shows other related works. Section 3 presents the proposed technique. Section 4 describes the experimental settings. Section 5 presents some of the achieved results. Section 6 provides discussions for the results and Section 7 concludes the paper.

2.0 RELATED WORK

2.1 EEG model
We assume the model for contaminated EEG signal as in the following form:

$$EEG_{rec}(t) = EEG_{true}(t) + k.EOG(t)$$

where $EEG_{rec}(t)$ is recorded contaminated EEG, $EEG_{true}(t)$ denotes the true EEG signal, and k.EOG (t) represents the propagated ocular artifact from eye to the recording site. The ultimate purpose of any artifact removal techniques is to recover $EEG_{true}(t)$ from $EEG_{rec}(t)$

2.2 Wavelet thresholding
Wavelet thresholding technique is built on the multiresolution analysis of wavelet transform, a tool that analyses signal in different time and frequency components [20]. These components, called approximations and details, are further processed by thresholding before reconstruction [14]-[18]. By selecting a 'good' mother wavelet, which resembles the shapes of the artifacts, large-valued coefficients are generated in the areas corresponding to the EEG artifacts at low-frequency sub-bands and are considered as an estimate of the ocular artifacts. Thus, shrinking the amplitude range of these coefficients by nonlinear thresholding functions would remove those artifacts. In this paper, a wavelet thresholding method was implemented as follow,

a. Use a butterworth lowpass filter to smooth the EEG signal before further processing
b. Apply Wavelet transform to the contaminated EEG signal
c. Utilize a thresholding function to automatically corrected high-valued coefficients at low-frequency sub-bands
d. Reconstruct the corrected EEG signal

2.3 Independent Component Analysis

Independent component analysis was first proposed by Herault and Jutten at a meeting in Snowbird Utah in 1986 [1, 11] to solve the blind source seperation problem (BSS). ICA aims at recovering independent source signals $s = \{s_1(t), s_2(t), \ldots s_N(t)\}$, from recorded mixtures $x = \{x_1(t), x_2(t), \ldots x_N(t)\}$ by an unknown matrix A of full rank. The basic problem of ICA is to estimate the mixing matrix $[A]$ or equivalently, the original independent sources (s) based on the following linear relationship $(x = As)$ while no knowledge is available about the sources or the mixing process. The method was developed based on several assumptions such as, the sources are statistically independent, the independent components must have non-Gaussian distributions and the matrix $[A]$ is assumed to be square and invertible. ICA identifies an unmixing matrix, $[W]$, which decomposes the multi-channel scalp data into a sum of temporally independent and spatially fixed components. ICA finds $(u = Wx)$, where the rows of the output data matrix represent time courses of activation of the ICA components [1, 9, 11]. Several algorithms have been proposed to implement ICA such as INFOrmation MAXimization approach (InfoMax), Fixed-point ICA, Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm and the Second Order Blind Identification (SOBI). In this research, the InfoMax algorithm was used to perform for EEG artifact removal.

3.0 PROPOSED METHOD

In this paper, we present a novel algorithm, Wavelet Neural Network (WNN), for EEG artifact revomal. In our method, the WNN is trained with simulated data resembling the properties in both time and frequency domains of EEG signal. The trained WNN is then used as the corrector for contaminated data. In both testing and training processes, the original signal is decomposed first with a wavelet to get different frequency components. The low frequency sub-band coefficients are then interpolated to maintain same lengths. A trained artificial neural network (ANN) is fed with such interpolated inputs to yield the corrected coefficients at its outputs. Finally, the corrected coefficients are downsampled for the wavelet construction to get the corrected signal y of original contaminated x as shown in Figure 1.

The core idea of the method, decomposing the signal in both time and frequency domains with wavelet and using an ANN to
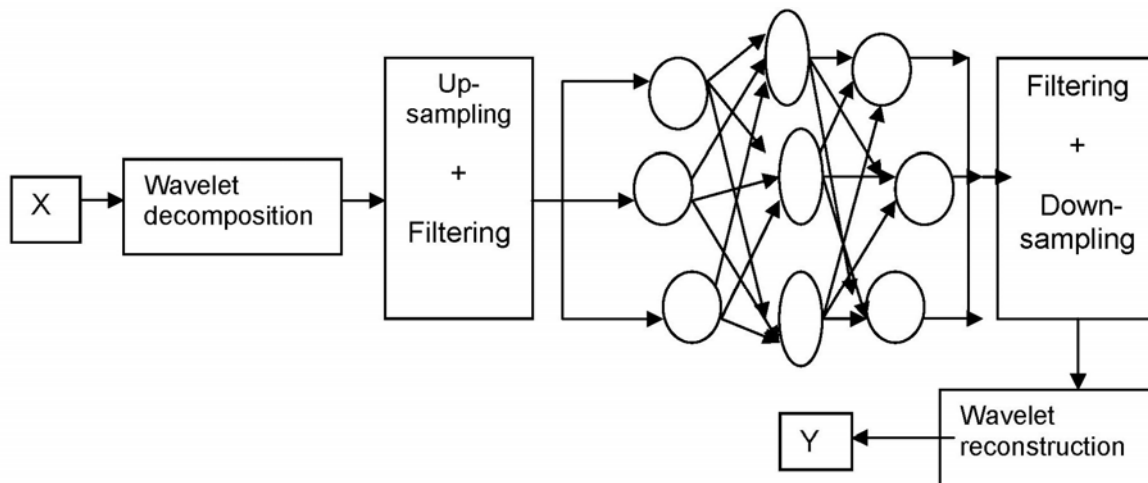

Figure 1. Proposed Wavelet Neural Network Structure.

correct them, can be viewed in a more succinct (and perhaps more precise) way. By combining the time/frequency property of wavelet and the universal approximation capability of neural network, we would be able to keep useful information related to cognitive activities while eliminate artifacts in EEG.

3.1 EEG data simulation

As described in [23], EEG signal can be simulated based on three assumptions, (1) Short segments of the spontaneous EEG can be described as linearly filtered, (2) non-stationary components in the spontaneous EEG can be simulated by changing the characteristics of this filtering process and (3) the spectral property of the simulated EEG data resembles that of actual signal. As shown in Figure 2, a set of Gaussian noises (GN) were generated and then filtered by a number of lowpass and bandpass filters with different cut-off frequencies that are similar to the spectral property of EEG frequency bands. Transients like eye blinks and eye movements, colleted from real signals were then filtered by lowpass filter and added to make the simulated data contaminated. Cutoff frequencies for those filters are

summarized in Table 1.

Table 1. EEG Frequency Band Specifications

Freq. bands	Lower (Hz)	Upper (Hz)
Delta	0.5	4
Theta	4	8
Alpha	8	13
Beta-1	13	20
Beta-2	20	30
Gamma	30	50

3.2 Neural Network Training

The backpropagation (BP) is used as the machine learning technique for multi-layer perceptron (MLP) neural nework. Experimental results show that the one hidden layer neural network structure 3-5-3 (3 inputs, 5 hidden units and 3 outputs) is good enough for EEG occular aritfact removal issue. The trained ANN's input and output are low frequency sub-band coefficients of the wavelet-decomposed simulated data and these coefficients after corrected, respectively. In this paper, the number of iterations for ANN training is set
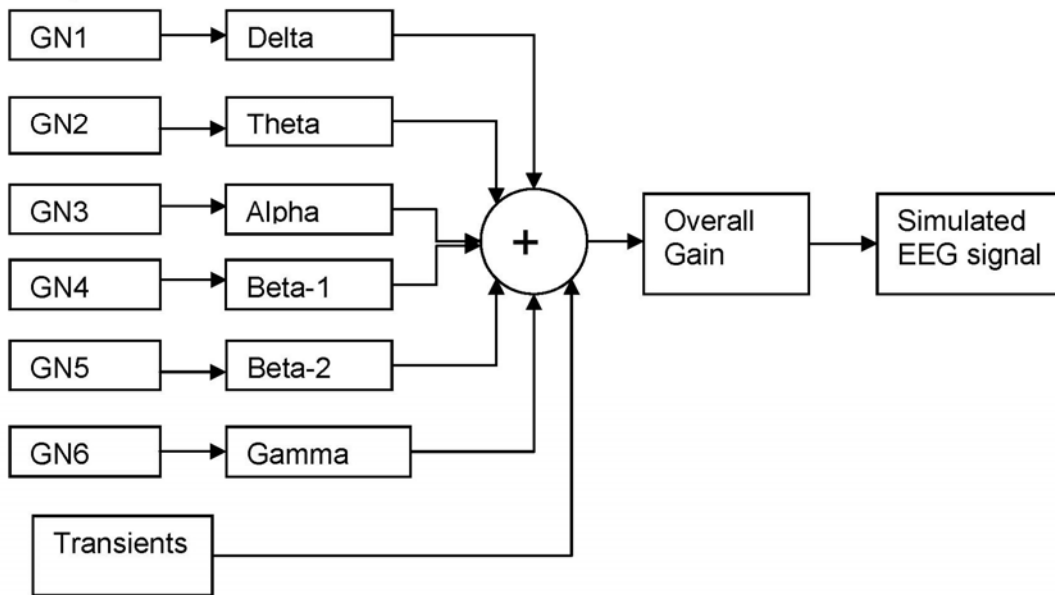

Figure 2. EEG Signal Simulation Model.

823

at 200, but this number might be optimized to improve the training accuracy.

3.3 Performance metric

We use two metrics, power spectrum density (PSD) and frequency correlation, to assess the proposed method. The PSD is a popular metric used to show information about the power spectrum of EEG signal at specific frequencies. Calculation of the correlation in frequency domain before and after artifact removal is equivalent to the correlation in time domain after filtering the time series with the corresponding frequency filter [14]-[22]. The frequency correlation between $\tilde{x}$ and $\tilde{y}$ is computed as in the following formula,

$$c = \frac{\frac{1}{2} * \sum_{w1}^{w2}(\tilde{x}^*\tilde{y} + \tilde{y}^*\tilde{x})}{\sqrt{\sum_{w1}^{w2}\tilde{x}\tilde{x}^* * \sum_{w1}^{w2}\tilde{y}\tilde{y}^*}}$$

where w1 and w2 are the lower and upper limits of the interested power spectrum region to be calculated, c is the correlation value that will be assigned to the frequency of (w1+w2)/2. If frequencies $\tilde{x}$ and $\tilde{y}$ are identical, c gets 1, otherwise, c obtains a value between 0 and 1. In this paper, the 'window size', w1-w2, is selected equal to 2.

### 4.0 EXPERIMENTS

4.1 Datasets

We validate our method on a data set, which was collected when participants were performing a driving test. The EEG information was collected by a 128-channel recording system at the sampling rate of 1000 Hz along with other information including description of the task, system dynamics related information, performance measures, physiological signals (ECG, respiration, etc.), and eye tracking. The workload was also analyzed according to the driving conditions (city-driving, stopped, highway passing, etc.). Due to the recording condition, the subject eye movements and blinks happen at high frequency making the data, especially at frontal recording channels, highly contaminated by ocular artifacts.

4.2 Experimental settings

We implemented three artifact removal methods for comparison, the ICA method, the wavelet thresholding algorithm and the proposed WNN technique. For each algorithm, we computed PSD and frequency correlation before and after artifact removal to illustrate the effectiveness of each of the algorithms. For the proposed method, we first simulated EEG signals to train an ANN and tested the trained model on a simulated signal and the driving test data set. For the wavelet thresholding method, we implemented it by following the instruction in [20] and for the ICA, we utilized the EEGLAB software.

5.0 RESULTS
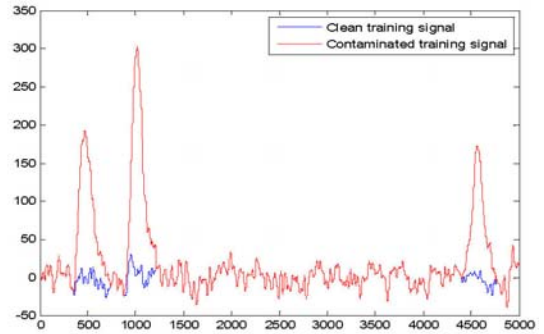
(a)
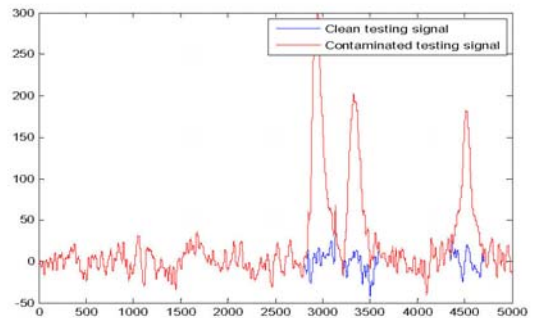
(b)

Figure 3. Clean and Contaminated Simulated Signal for (a) Training and (b) Testing.

For the proposed WNN algorithm, two simulated segments with a length of 5 seconds for training and testing at sampling rate of 1000 Hz were created as shown in Figure 3a and 3b, respectively, where the artifacts were taken from the driving test data set and added to the simulated data segments. Data in Figure 3a was then used to train the neural network in the proposed WNN algorithm. We applied the trained WNN model to the testing data segment (Figure 3b), and the corrected EEG signal is shown in Figure 4. Figures 5 shows PSD of the contaminated, corrected and the clean EEG signals. Figure 6 shows frequency correlations among those signals.
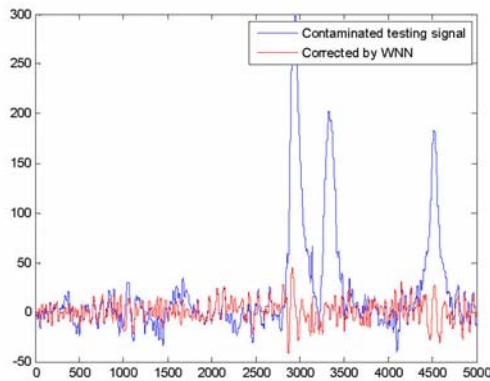

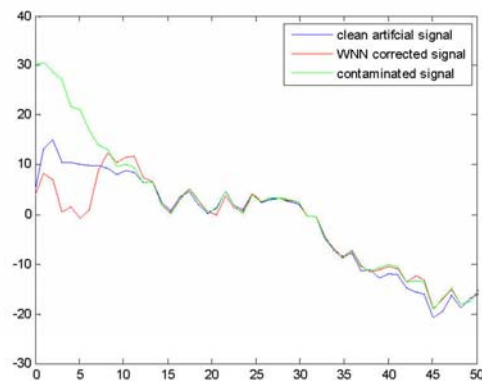
Figure 4. Contaminated Simulated and WNN Corrected Singals

Figure 5. PSD of Clean, Contaminated and WNN Corrected Signals for Testing

For the ICA algorihtm, it took a computer, equiped with Intel(R) Core(TM) 2 CPU 6400 @ 2.13 GHz and RAM 2.00 GB, 27 minutes

with 382 steps to remove the artifacts for one EEG segment in the driving data set.
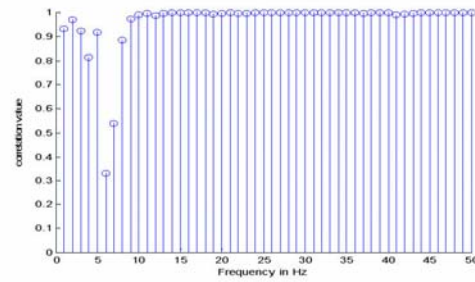
(a)
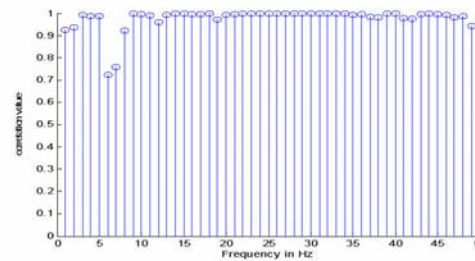
(b)

Figure 6. Frequency Correlation between (A) Contaminated and Corrected Simulated Signals and (B) Clean and Corrected Simulated Signals.

We then applied the trained WNN model to the driving test data set. We decomposed the EEG signal to 8 levels and 3 low frequency sub-band coefficients were corrected by the WNN algorithm.

The wavelet thresholding method was used to adaptively correct 4 low frequency sub-bands coefficients. For specific data segments, the corrections were repeated a number of times with various wavelets and at different levels of decompositions in order to make the corrected data most acceptable. The wavelets from Coiflet and Daubechies family were chosen because experiments show that they could extract the features of artifacts efficiently.

Figure 7 show PSD plots for one sample artifact removed segment in the driving test data by the three algorithms. Figure 8 shows the segment in time domain. Figure 9 shows frequency correlations between the contaminated and corrected segments.

825

Figure 7. PSD of Contaminated and Decontaminated EEG

(a) ICA

(a)

(b) Wavelet Thresholding

(b)

(c) WNN

Figure 8. Contaminated and Decontaminated EEG (a) Contaminated, ICA and WNN Corrected EEG (b) Contaminated, Wavelet Thresholding and WNN Corrected EEG
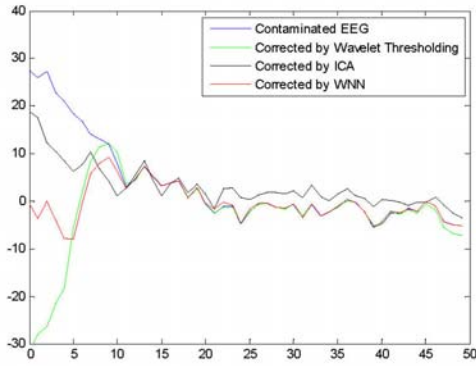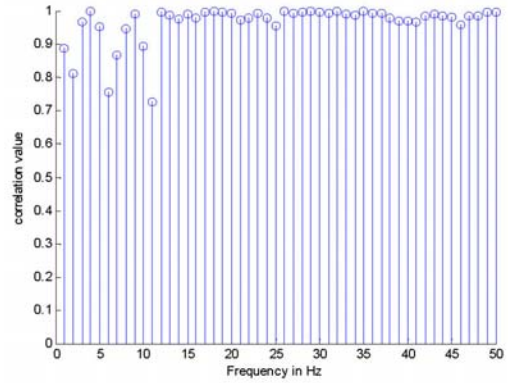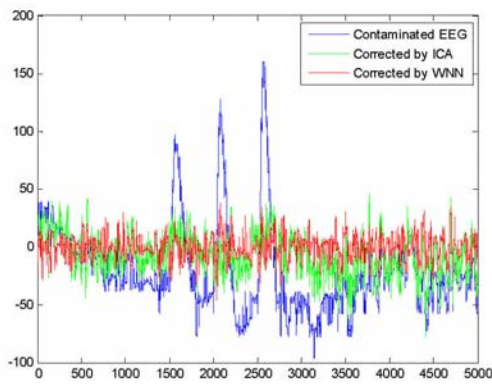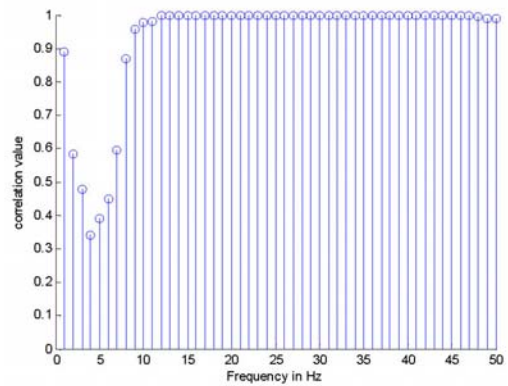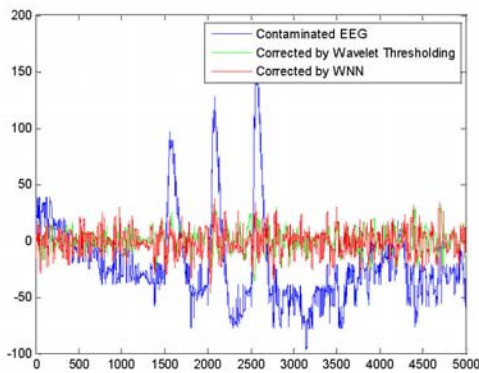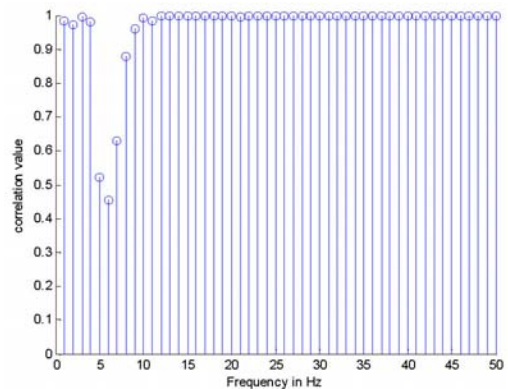
Figure 9. Frequency Correlation between Contaminated and Decontaminated EEG, (a) by ICA, (b) by Wavelet Thresholding and (c) by WNN

826

6.0 DISCUSSION

It is observed from various results that the WNN algorithm removes ocular artifacts efficiently while keeping cerebral background information. Like Wavelet Thresholding, WNN just needs one single channel data to perform correction that makes it advantageous over ICA, which needs to perform on the whole dataset. Furthermore, the method was proved through repeated experiments on various data segments for its effectiveness and stability, which is not true for the wavelet thresholding algorithm.

The PSD plot shows that the low frequency components are reduced significantly in the corrected signal. That is more evident if we look at the frequency correlation metric plot between contaminated and corrected signals: there is a slight difference in the range of low frequency components while in other ranges, the useful information is well-preserved.

The frequency correlation plots also show that the correction made by ICA spreads over the entire frequency range and the power of low frequency components are reduced not significantly. Mean while, the low frequency components in the signal were derogated by Wavelet Thresholding and WNN while high frequency components are well preserved by both.

ICA requires a lot more computing power and multiple channel data sources for artifact removal. It also demands either an automatic or a manual step to determine which independent component is artifact, making an online implementation of ICA difficult.

7.0 CONCLUSIONS

We proposed a novel algorithm, WNN, for artifact removal for EEG signal. The algorithm combines the time/frequency property of wavelet and the approximating capability of neural network to locate and eliminate artifacts. Experimental results ona driving data set show that WNN can effectively remove artifact and achieve better results than the wavelet thresholding algorithm. WNN is also much computationally efficient than the ICA algorithm making it possible an automatic online algorithm.

8.0 ACKNOWLEDGEMENT

9.0 REFERENCES

[1]. T. Jung, S. Makeig, C. Humphries, T. Lee, M. Mckeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," Psychophysiology, vol. 37, pp163-178, 2000.

[2]. E. Urretarazu, J. Iriarte, M.Alegre, M. Valencia, C. Vireri, and J. Artieda, "Independent component analysis removing artifacts in ictal recordings," Epilepsia, vol. 45, pp. 1071-1078, Sep. 2004.

[3]. Diego A. Pizzagalli, "Electroencephalography and High-Density Electrophysiological Source Localization", Chapter prepared for Cacioppo, J.T. et al., Handbook of Psychophysiology (3rd Edition)

[4]. D. Hagemann and E. naumann, "The effects of ocular artifacts on (lateralized) broadband power in the EEG," Clinical Neurophysiols., vol. 112, pp. 215-231, 2001.

[5]. C.A. Joyce, I. F. Gorodnitsky, and M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation," Psychophysiology, vol. 41, pp. 313 - 325 2004

[6]. Croft R J and Barry R J, "Removal of ocular artifact from the EEG: a revie," Clin. Neurophysiol. Vol. 30, pp 5-19, 2000.

[7]. Kandaswamy A, Krishnaveni V, Jayaraman S, Malmurugan N and Ramadoss K, "Removal of occular artifacts from EEG-a survey", IETE J. Res. Vol. 52 pp 121-130, 2005.

[8]. Grattton G, Coles M G and Donchin E, "A new method for off-line removal of ocular artifact," Electroencephalogram. Clin. Neurophysiol. Vol. 55, pp 84-468, 1983.

[9]. Woestenburg J C, Verbaten M N and Slangen J L, "The removal of the eye movement artifact from the EEG by regression analysis in the frequency domain," Biol. Psychol. Vol 16, pp 47-127, 1983

[10]. P. Berg, and M. Scherg, "Dipole models of eye activity and its application to the removal of eye artifacts from the EEG ad MEG". Clinical Physics and Physiological Measurements, vol.12(Supplement A), pp49–54, 1991b.

[11]. T. D. Lagerlund, F. W. Sharbrough, and N. E. Busacker, "Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition". Journal of Clinical Neurophysiology, vol. 14, pp73–82, 1997.

[12]. P. Common, "Independent Component Analysis, A new concept?". Signal Processing, vol. 36, pp 287-314, 1994.

[13]. Delorme A, Makeig S and Sejnowski T, "Automatic artifact rejection for EEG data using high-order statistics and independent component analysis," Proc. Third Int. ICA Conf. pp 9-12, 2001

[14]. V Krishnaveni, S. Jayaraman, L. Anitha and K. Ramadoss, "Removal of ocular artifacts from EEG using adaptive thresholding of wavelet coefficients," J. Neu. Eng. 3, pp 338-346, 2006

[15]. K.P. Indiradevi, E. Elias, P.S. Sathidevi, S. Dinesh Nayak and K. Radhakrishman, "a multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram", Com. In Bio. And Medicine 38, pp 805-816, 2008

[16]. L. Senhadji, J-L Dillenseger, F. Wendling, C. Rocha, and A. Kinie, "Wavelet analysis of EEG for three-dimensional mapping of epileptic events," Ann. Biomed Eng. Vol. 23(5), pp 543-552, 1995

[17]. V. J .Samar, A. Bopardikar, R. Rao and K. Swartz, "Wavelet analysis of neuroelectric waveforms: A conceptual tutorial", Brain and Language, Vol. 66, pp 7-60, 1999.

[18]. X. P. Zhang and M. D. Desai "Adaptive denoising based on SURE risk," IEEE Sig. Pro. Letters, Vol. 5, No. 10, Oct. 1998

[19]. V. J. Samar, A. Bopardikar, R. Rao, and K. Swartz, "Wavelet Analysis of Neuroelectric Waveforms: A Conceptual Tutorial," Brain and Language 66, 7-60(1999).

[20]. Daubechies Ingrid, "Ten lectures on wavelets".

[21]. Nicole A.M. de Beer et al, "Educational simulation of the electroencephalogram (EEG)"

[22]. A. Jung 'Dissertation on Statistical Analysis of Biomedical Data'

EEG Artifact Removal Using A Wavelet Neural Network

[1]Hoang-Anh T. Nguyen, [1]John Musson, [2]Guangfan Zhang, [2]Roger Xu, [3]Carl Richey, [3]Tom Schnell, [1]Frederic D. McKenzie and [1]Jiang Li

[1]Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA
[2]Signal Processing Group, Intelligent Automation, Inc., Rockville, MD
[3]Department of Industrial Engineering, University of Iowa, IA

1

Outline

➤ Introduction
➤ Related work
➤ Proposed method
➤ Experiment and results
➤ Discussion and conclusions

2

Outline

➢ Introduction

➢ Related work

➢ Proposed method

➢ Experiments and results

➢ Discussion

➢ Conclusions

3

EEG and artifacts

- **Electroencephalogram** (EEG):
  – Neural electrical signal
  – Recorded by using recording system
  – Important to many application fields: Computer control and communication, entertainment, education, military, commercial, etc.
- **Artifacts**:
  – Unavoidable non-cortical activities
  – Sources: Muscle activity, line noise, heart beating, eye movements and blinks, etc.
- **Electrooculogram** (EOG) artifact:
  – Generated by eye movements or blinks
  – Main artifactual portion of EEG recordings

830
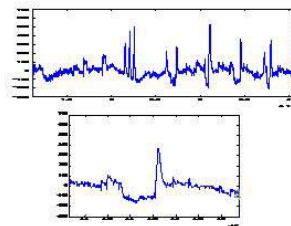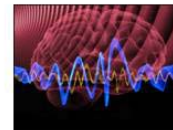
Contaminated EEG

5

Artifact removal techniques

- Regressions in time/frequency
- Principle Component Analysis (PCA)
- Independent Component Analysis (ICA)
- Wavelet Thresholding (WT)
- Wavelet Neural Network (WNN)?

Techniques comparison

Techniques	Regression	PCA	ICA	WT	WNN
Need EOG?	Yes	Yes	No	No	?
Single channel	No	No	No	Yes	?
Reliability	?	?	Yes	No	?
Automatic	Yes	No	No	Yes	?

6

Wavelet neural network brief description
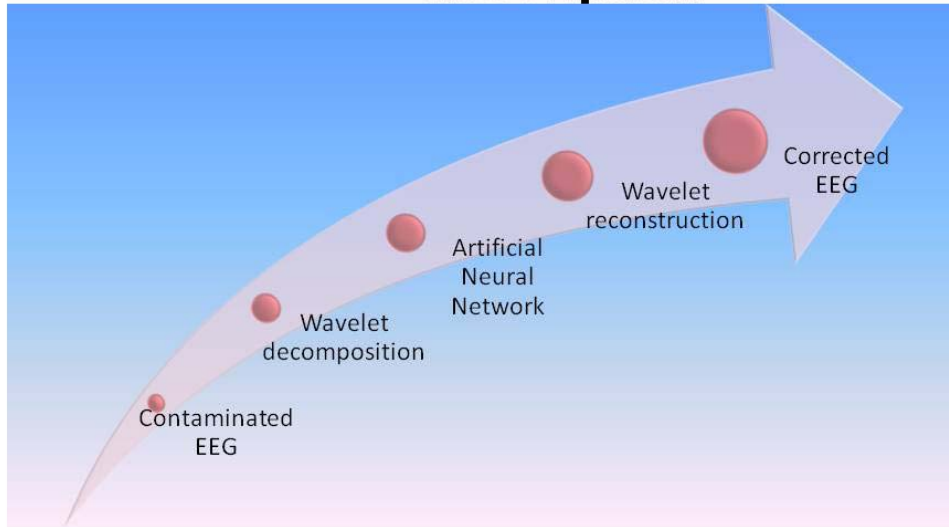
Contaminated EEG → Wavelet decomposition → Artificial Neural Network → Wavelet reconstruction → Corrected EEG

Outline

➤ Introduction
➤ Related work
➤ Proposed method
➤ Experiments and results
➤ Discussion
➤ Conclusions

EEG model

- EEG recordings
  - Contaminated EEG
  - Superposition of true EEG and propagated artifacts
- True EEG
  - Cortical signals excluding artifacts
- Propagated Artifacts
  - Non-cerebral activities
  - Contaminated EEG electrode recordings
  - Propagation factor k proportional to the recording electrode location

$$EEG_{true}(t) + k\_EEG_{art}(t) \rightarrow EEG_{rec}(t)$$
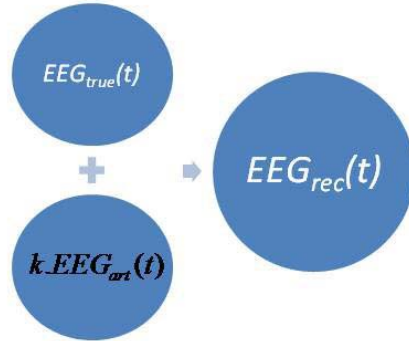
9

Wavelet transform

$$\psi_{a,\tau}(t) = \psi\left(\frac{t-\tau}{a}\right) \qquad W(a,\tau) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(t)\psi_{a,\tau}^{*}(t)dt$$
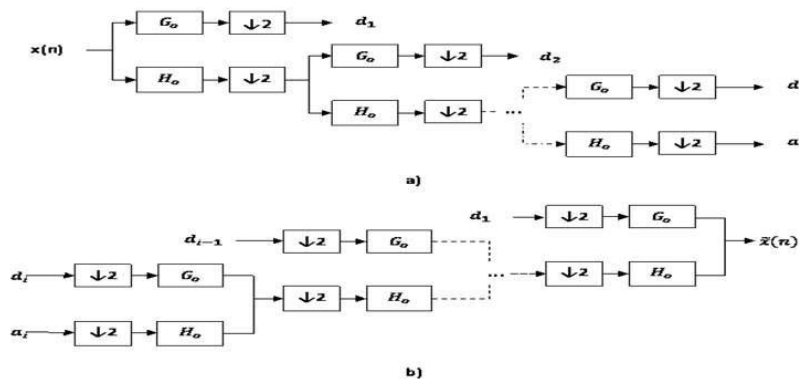
Figure 1. [a] Wavelet transform and [b] Wavelet reconstruction

10

833

Wavelet thresholding

Contaminated EEG
↓
Low-pass filtering
↓
Wavelet transform
↓
Thresholding
↓
Wavelet reconstruction
↓
Corrected EEG
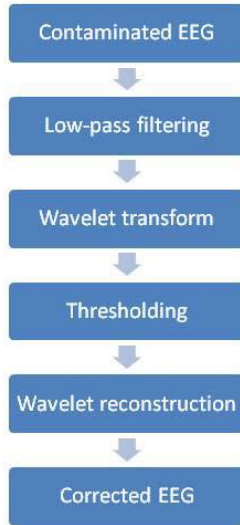
- Threshold function[18]:

$$w_k^c(w,t) = \begin{cases} w + t - \dfrac{t}{2k+1}, & w < t \\ \dfrac{1}{(2k+1)t^{2k}}\, w^{2k+1}, & |w| \le t \\ w + t + \dfrac{t}{2k+1}, & w > t \end{cases}$$

- Low-pass filter:
  - Butterworth
- Wavelet basis function selection:
  - Daubchies and Coiflet
  - Sensitive to time/frequency properties of EEG waves

11

Independent Component Analysis

- Notations:
  - **x** original mixtures
  - **s** blind source matrix
  - **u** estimated source
  - **A** mixing matrix
  - **W** un-mixing matrix, inverse of A
- ICA assumptions:
  - Source independence
  - Non-Gaussianality
  - A and W to be square and invertible
- Source independence definition:
  - Minimizing mutual information
  - Maximizing non-Gaussianality

s   x   u

A   W

Unknown mixing process   Blind Source Separation
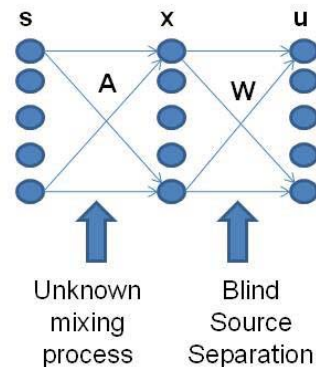
12

ICA in EEG artifact removal

$$\tilde{s} = Wx \qquad \tilde{x} = W^{-1}\tilde{s}'$$

Original Contaminated EEG → ICA → Zeros padding → Inverse ICA → Corrected EEG

$x \qquad \tilde{s} \qquad \tilde{s}' \qquad \tilde{x}$
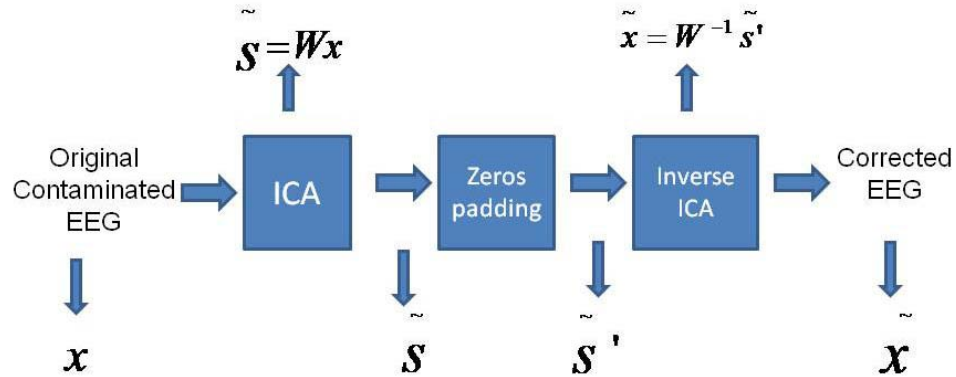
Being a batch algorithm, ICA needs to be performed on the whole data set with at least an adequate number of data points, so the computational power is expensive.

13

Outline

➢ Introduction

➢ Related work

➢ Proposed method

➢ Experiments and results

➢ Discussion

➢ Conclusions

14

Wavelet Neural Network
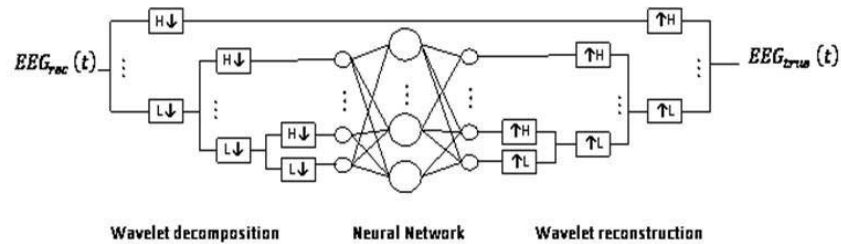
Figure 2. Wavelet Neural Network structure

15

Network Training
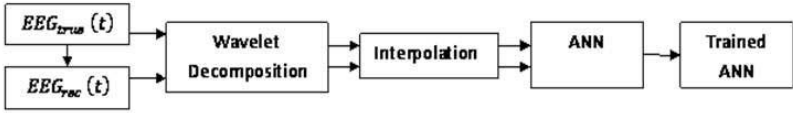
Figure 3. Neural network training procedure

16

EEG data simulation
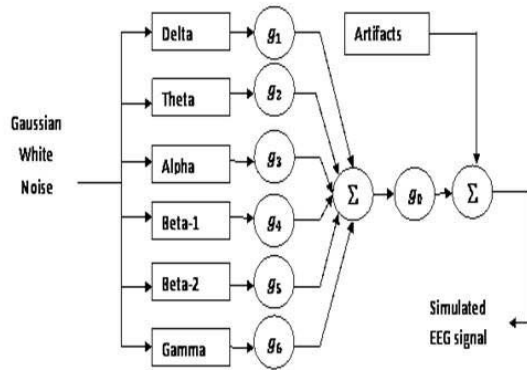
Figure 4. Simulated EEG generator model

Table 1. EEG frequency band specifications

Freq. bands	Lower (Hz)	Upper (Hz)
Delta	0.5	4
Theta	4	8
Alpha	8	13
Beta-1	13	20
Beta-2	20	30
Gamma	30	50

17

Simulated EEG

18

837
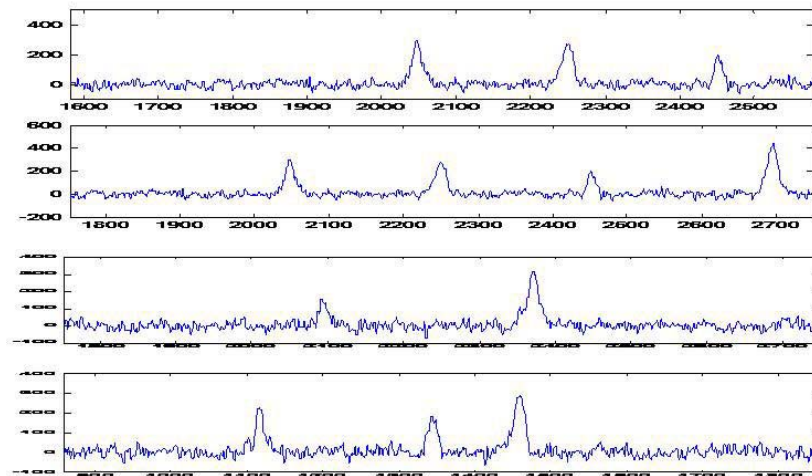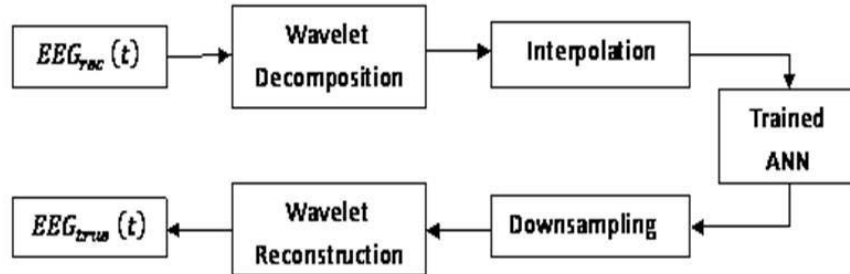
Network testing

Figure 5. Wavelet Neural Network testing procedure

19

Outline

➢ Introduction
➢ Related work
➢ Proposed method
➢ **Experiments and results**
➢ Discussion and conclusions

20

838

Experiments

- Data set:
  - Driving test
  - 128 recording channel system
  - Highly disturbed by multi-type artifacts
- Experimental settings:
  - Methods: WNN, WT and ICA
  - Validation metrics: PSD and frequency correlation
    - Frequency correlation mathematical formula:

$$c = \frac{\frac{1}{2} * \sum_{w1}^{w2} (\tilde{x}^* \tilde{y} + \tilde{y}^* \tilde{x})}{\sqrt{\sum_{w1}^{w2} \tilde{x}\tilde{x}^* * \sum_{w1}^{w2} \tilde{y}\tilde{y}^*}}$$

21

Results – Simulated EEG
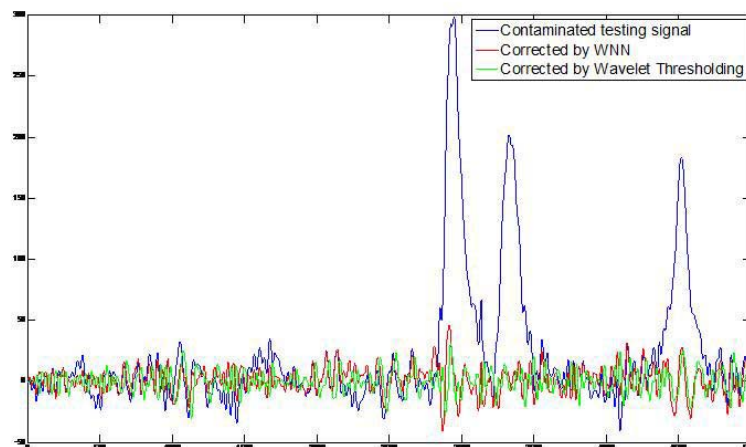
Figure 6. WNN performance on simulated data

22

Results – Simulated EEG
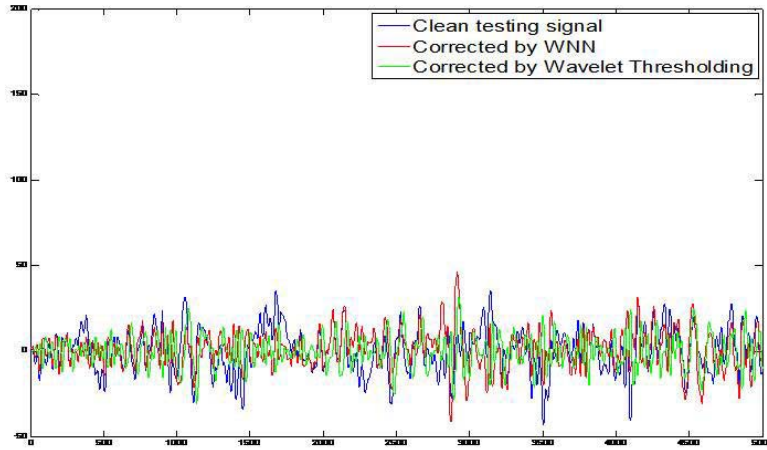
**Figure 7. Clean simulated and Decontaminated signals by
WNN and Wavelet Thresholding**

23

Error signals

Technique	RMSE
WNN	12.2473
Wavelet Thresholding	16.4942
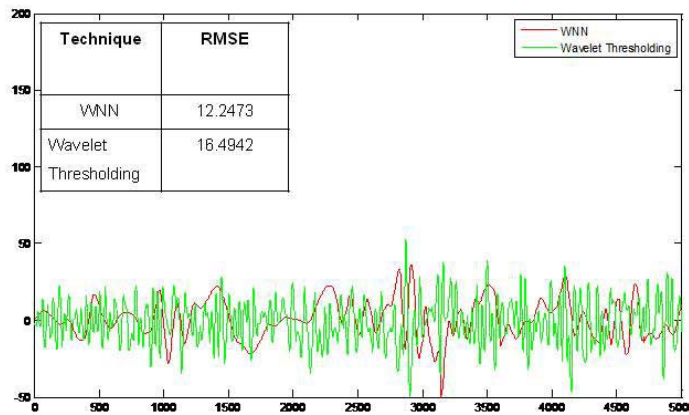
**Figure 8. Error signals, or differences between the 'ground
truth' and signals corrected by WNN and WT**

24

PSD and frequency correlation
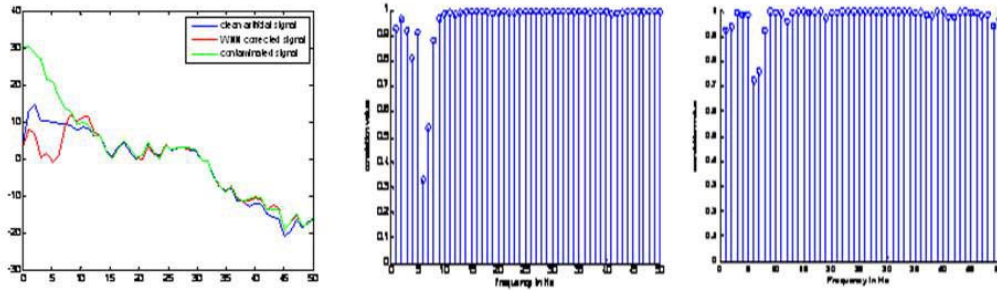
Figure 9. PSD (left) and frequency correlation between contaminated and corrected simulated signals (center) and clean and corrected simulated signals (right)
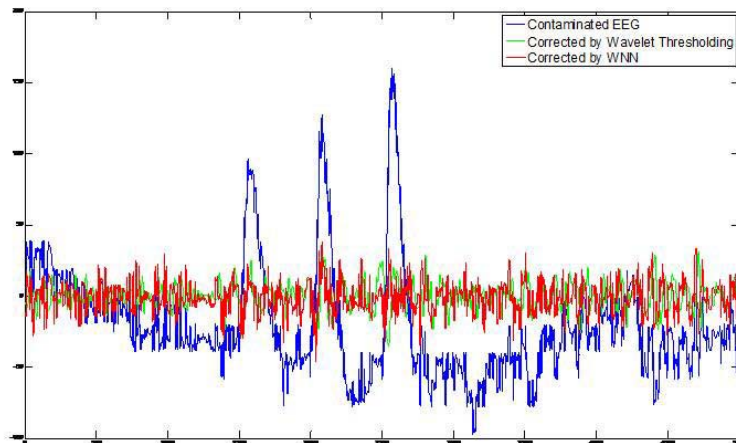
Results – Real EEG

Figure 10. Contaminated, Wavelet Thresholding and WNN Corrected EEG

Results – Real EEG
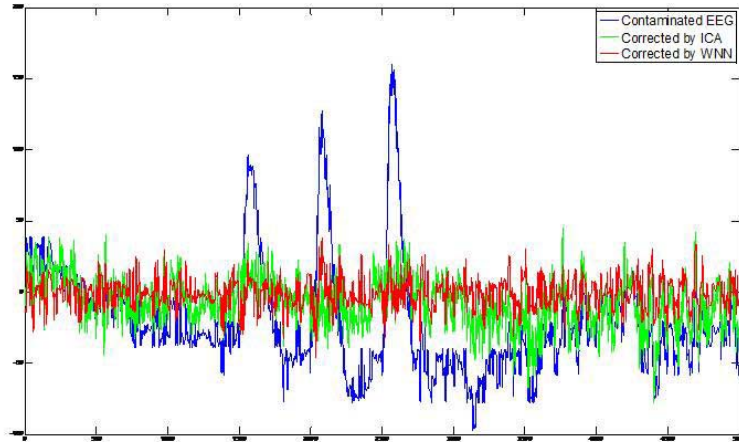
Figure 11. Contaminated, ICA and WNN Corrected EEG

Power Spectrum Density
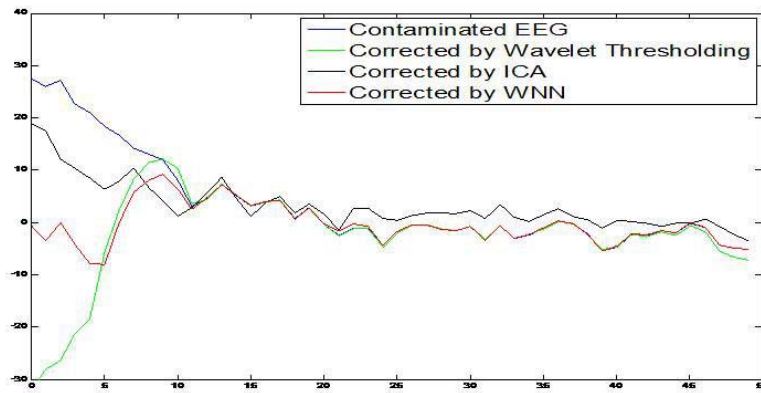
Figure 12. PSD of Contaminated and De-contaminated EEG

Frequency correlation

a)   b)   c)
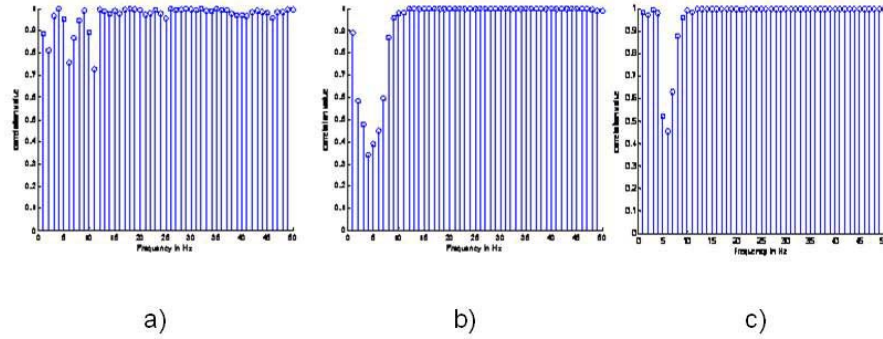
Figure 13. Frequency Correlation between Contaminated and Decontaminated EEG, (a) by ICA, (b) by Wavelet Thresholding and (c) by WNN

29

\

Outline

➤ Introduction

➤ Related work

➤ Proposed method

➤ Experiments and result

➤ Discussions and conclusions

30

Discussions

Techniques comparison

Techniques	Regression	PCA	ICA	WT	WNN
Need EOG	?Yes	Yes	No	No	No
Single channel	No	No	No	Yes	Yes
Reliability	?	?	Yes	No	Yes
Automatic	Yes	No	No	Yes	Yes

Wavelet Thresholding: It is sensitive to Wavelet basis function choice

ICA: computational complexity

31

Conclusions

❖ A novel and efficient method Wavelet Neural Network and its application to EEG artifact removal

❖ Make comparisons with several methods
  - ICA
  - Wavelet Thresholding

❖ Future work

32

References

1. T. Jung, S. Makeig, C. Humphries, T. Lee, M. Mckeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," Psychophysiology, vol. 37, pp 163-178, 2000.
2. E. Urretarazu, J. Iriarte, M. Alegre, M. Valencia, C. Vireri, and J. Artieda, "Independent component analysis removing artifacts in ictal recordings," Epilepsia, vol. 45, pp. 1071-1078, Sep. 2004.
3. Diego A. Pizzagalli, "Electroencephalography and High-Density Electrophysiological Source Localization", Chapter prepared for Cacioppo, J.T. et al., Handbook of Psychophysiology (3rd Edition)
4. D. Hagemann and E. naumann, "The effects of ocular artifacts on (lateralized) broadband power in the EEG," Clinical Neurophysiols., vol. 112, pp. 215-231, 2001.
5. C.A. Joyce, I.F. Gorodnitsky, and M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation," Psychophysiology, vol. 41, pp. 313 - 325 2004
6. Croft R J and Barry R J, "Removal of ocular artifact from the EEG: a revie," Clin. Neurophysiol. Vol. 30, pp 5-19, 2000.
7. Kandaswamy A, Krishnaveni V, Jayaraman S, Malmurugan N and Ramadoss K, "Removal of ocular artifacts from EEG-a survey", IETE J. Res. Vol. 52 pp 121-130, 2005.
8. Grattton G, Coles M G and Donchin E, "A new method for off-line removal of ocular artifact," Electroencephalogram. Clin. Neurophysiol. Vol. 55, pp 84-468, 1983.
9. Woestenburg J C, Verbaten M N and Slangen J L, "The removal of the eye movement artifact from the EEG by regression analysis in the frequency domain," Biol. Psychol. Vol 16, pp 47-127, 1983
10. P. Berg, and M. Scherg, "Dipole models of eye activity and its application to the removal of eye artifacts from the EEG ad MEG". Clinical Physics and Physiological Measurements, vol.12(Supplement A), pp49–54, 1991b.
11. T. D. Lagerlund, F. W. Sharbrough, and N. E. Busacker, "Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition". Journal of Clinical Neurophysiology, vol.14, pp73–82, 1997.
12. P. Common, "Independent Component Analysis, A new concept?". Signal Processing, vol. 36, pp 287-314, 1994.
13. Delorme A, Makeig S and Sejnowski T, "Automatic artifact rejection for EEG data using high-order statistics and independent component analysis," Proc. Third Int. ICA Conf. pp 9-12, 2001
14. V Krishnaveni, S. Jayaraman, L. Anitha and K. Ramadoss, "Removal of ocular artifacts from EEG using adaptive thresholding of wavelet coefficients," J. Neu. Eng. 3, pp 338-346, 2006
15. K.P. Indiradevi, E. Elias, P.S. Sathidevi, S. Dinesh Nayak and K. Radhakrishman, "a multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram", Com. In Bio. And Medicine 38, pp 805-816, 2008
16. L. Senhadji, J-L Dillenseger, F. Wendling, C. Rocha, and A. Kinie, "Wavelet analysis of EEG for three-dimensional mapping of epileptic events," Ann. Biomed Eng. Vol. 23(5), pp 543-552, 1995
17. V. J .Samar, A. Bopardikar, R. Rao and K. Swartz, "Wavelet analysis of neuroelectric waveforms: A conceptual tutorial", Brain and Language, Vol. 66, pp 7-60, 1999.
18. X. P. Zhang and M. D. Desai "Adaptive denoising based on SURE risk," IEEE Sig. Pro. Letters, Vol. 5, No. 10, Oct. 1998
19. V. J. Samar, A. Bopardikar, R. Rao, and K. Swartz, "Wavelet Analysis of Neuroelectric Waveforms: A Conceptual Tutorial," Brain and Language 66, 7-60(1999).
20. Daubechies Ingrid, "Ten lectures on wavelets".
21. Nicole A.M. de Beer *et al,* "Educational simulation of the electroencephalogram (EEG)"
22. A. Jung 'Dissertation on Statistical Analysis of Biomedical Data'

33

Thank You !

Q & A

34

845

6.0 K-20 STEM EDUCATION TRACK

6.1 Virginia Demonstration Project Encouraging Middle School Students in Pursuing STEM Careers

Virginia Demonstration Project Encouraging Middle School Students in Pursuing STEM Careers

Jane T. Bachman
Naval Surface Warfare Center Dahlgren Division

Dr. Dena H. Kota
Naval Surface Warfare Center Dahlgren Division
Aaron J. Kota
Naval Surface Warfare Center Dahlgren Division

Abstract. Encouraging students at all grade levels to consider pursuing a career in Science, Technology, Engineering, and Mathematics (STEM) fields is a national focus. In 2005, the Naval Surface Warfare Center, Dahlgren Division (NSWCDD), a Department of Defense laboratory located in Dahlgren, Virginia, began work on the Virginia Demonstration Project (VDP) with the goal of increasing more student interest in STEM education and pursuing STEM careers. This goal continues as the program enters its sixth year. This project has been successful through the participation of NSWCDD's scientists and engineers who are trained as mentors to work in local middle school classrooms throughout the school year. As an extension of the in-class activities, several STEM summer academies have been conducted at NSWCDD. These academies are supported by the Navy through the VDP and the STEM Learning Module Project. These projects are part of more extensive outreach efforts offered by the National Defense Education Program (NDEP), sponsored by the Director, Defense Research and Engineering. The focus of this paper is on the types of activities conducted at the summer academy, an overview of the academy planning process, and recommendations to help support a national plan of integrating modeling and simulation-based engineering and science into all grade levels, based upon the lessons learned.

1.0 INTRODUCTION

Since 2005, the Naval Surface Warfare Center, Dahlgren Division (NSWCDD), a Department of Defense laboratory, has been working on National Defense Education Program's (NDEP) Virginia Demonstration Project (VDP) with the goal of increasing student interest in Science, Technology, Engineering, and Mathematics (STEM) education. One of the VDP STEM focus events is conducting a summer academy for middle school-age students. During this week-long event, students participate in a variety of STEM activities. Provided in this paper is an overview of the academy planning process and a focus on the types of activities selected and conducted during the academy.

1.1 Academy Planning

The planning for a summer academy event is a year-long process. Planning consists of coordination of the event dates, facility, mentors, schools, activity selections, mentor and junior mentor training, scheduling, supply and/or inventory, and execution.

The first objective is to organize a STEM Academy Planning Team, consisting of scientists and engineers (S&Es), academia, and a middle school teacher.

1.2 VDP STEM Academy Planning Team

The team meets once a month, addressing planning activities identified in an academy timeline developed at Dahlgren. The main goals of the team are to select the academy dates, review and select the STEM activities, assign leads for each activity (most activities are conducted by the mentors, but during mentor training and academy week, the authors found it beneficial to have an assigned activity lead), selection of a facility to host the event, and train mentors and junior mentors. The planning team becomes the staff executing the academy event.

1.2.1 Director and Coordinator

The academy staff has a director and a coordinator. The director manages the collaboration between mentors, schools,

facility, and training as well as leads the monthly planning team meetings. The coordinator handles the coordination between publications, supplies, equipment, tools, inventory, and academy scheduling.

1.2.2 Academy Training
The VDP STEM Academy Planning Team conducts a three-day training event for all mentors participating in the academy. In addition, the junior mentors receive training during one of the three days. All mentors receive ethical training as well as training on inquiry-based learning techniques. The academy schedule, rules, procedures, and activities are discussed with the mentors. Exploratory laboratory time is worked into the training schedule to give mentors additional time to spend on any one of the activities discussed. Mentors are exposed to all student activities, including team building.

2.0 ACADEMY ACTIVITIES
Each STEM activity is reviewed and selected by the planning team with the goal of providing a wide range of STEM activities that cover multiple careers for the students. Once an activity is selected by the team, an activity plan is derived and finalized for the mentor manual. Upon complete selection of the academy activities, a schedule is formulated for the week. Students this year were given the opportunity to explore STEM careers in life science, robotics, tower design/construction/design presentation, and data collection and analysis through water rocket activities. Students also participate in laboratory demonstrations, listen to guest speakers discuss their STEM careers, and learn about team communication and collaboration skills. There is very little downtime during the day for the students. Students are equipped with a summer academy student manual containing their activities and the information they need during the week.

Older students participate in a Junior Mentor program during the academy. Half of their day is spent carrying out academy administrative duties, while the other half is spent working on an assigned robotics project. Test engineers are one of the roles that a junior mentor serves when conducting administrative duties. This year, the project was to build a robot that could navigate a maze.

2.1 Activity Descriptions
- Team Building. Mentors are provided several team-building activities that they can help facilitate. The team derives a name and constructs a poster that will host their mission completion tags.

- Life Science. Two life-science activities

Figure 1. Life Science Activity

are conducted to simulate the types of ongoing research at naval laboratories. Students learn about the spread of an epidemic and possible methods used by scientists to combat such types of warfare (see Fig. 1).
- Tower Building. This consists of several phases of work for the students. First, the student team decides on a design for their tower. Second, they construct the tower (see Fig. 2), followed by testing the strength-weight ratio. To conclude, the team formulates its design into a presentation that is given in front of an invited panel. At the conclusion of the team's presentation, the panel conducts a question-and-answer period.

848

Figure 2.   Tower Construction Activity

- Water Rockets. Students construct a water rocket and conduct several trial tests to gather data. Following data analysis, the students decide on final measurements and conduct a final water rocket test to achieve the highest launch possible (see Fig. 3).

Figure 3.   Water Rocket Activity

2.1.1 Robotics

The robotic activity contains eight missions. Mission rules are established to provide some boundaries; however, team creativity is encouraged. Robotics boards contain a challenge mat denoting the home base location, island, troop rotation, humanitarian aid drop area, ship rescue area, and dry dock. The 'Map the Underwater Surface' has its own table designed to represent underwater terrain. Teams can test as many times as they want (see Fig. 4) prior to a test engineer witnessing the final mission test.

Figure 4.   Robotic Missions

- Rescue the Swimmer. Team members will need to rescue a swimmer. Starting from home base, the robot must be capable of maneuvering around the island, grabbing the swimmer from a known location, and bringing the swimmer to shore by any means. The robot may use any sensor available to the team including the rotation sensors built into the motors.
- Troop Rotation. Team members must transport troops from home base to a specific troop location across the water that already contains a group of troops on a platform. Troops must not touch the water during transport. The troops must be wholly within the drop-off area, and the robot may not run over them on the return trip. The robot must return the original troops stationed at this location to home base. There must always be a minimum of five troops at the designated location., and troops are not to be mixed during the transfer. The robot must use at least two sensors.
- Recover the Ship. Team members will recover the damaged ship and bring it back to home base by any means. If the robot turns the ship on its side or flips it, the mission must be reattempted. The robot must use at least two sensors.
- Create an Early Warning Structure. Team members will need to design a stationary early warning structure containing an NXT brick that will act as a signaling device. The structure will be placed where the lighthouse is located on the challenge board. A test engineer

will then start a robot from home base and direct its movement toward the tower. As a robot approaches the structure, a series of signals must indicate the distance from the robot to the tower. These distances should be broken into three range groupings each of which corresponds to a unique signal: more then 10 inches away, 10-4 inches away, and less than 4 inches. Teams may select whatever signaling method they would like. Examples include three colored lights, changing the NXT display, or variations in tone. Note: Each team should have, for the test engineer, a programmed robot able to approach the early warning structure at a slow-to-moderate pace during testing.

- Recover Beacon. Team members will need to detect and recover an infrared (IR) beacon located somewhere in the water. The test engineer may place the beacon anywhere within the beacon placement area (see map of robotics board). Each team searches the area and captures the beacon. The robot must use at least an IR seeker.
- Map the Underwater Surface. Team members will need to create a map of underwater terrain using the ultrasonic sensor and the data collection features of the robotic software. The map must provide depth and inches traveled in inches. It should be scaled properly and reported to a test engineer on graph paper.
- Dry Dock. The teams' robots will need leave from home base, drive up onto the top level of the dry dock, display a message, remain there for five seconds, drive off of the dry dock, and return to home base. The robots may use any sensor available to the teams including rotation sensors built into the motors.
- Humanitarian Aid. Each team must deliver five crates of humanitarian aid to the designated location. The robot may use any sensor available to the team including the rotation sensors.

2.1.2 Academy Token Plan
The VDP STEM Academy Planning Team generated a plan where tokens serve as the students' form of currency for the week. Each time a team attempts to complete a robotic mission, a token must be paid to the test engineer. The tokens provide students with an incentive program during their summer academy week. The role of 'token master' is served by a junior mentor who, under the supervision of the academy director, is in charge of distributing tokens. Teams begin the week with 10 tokens and the plan details events throughout the week in which teams can earn tokens based on their accomplishments and teamwork.

3.0 LESSONS LEARNED
Following the academy, the VDP STEM Academy Planning Team's first meeting is to identify the lessons learned and to discuss the "do's and don'ts" for next year. Below is a list of this year's lessons learned.

- Facility. A gym that provides enough space for the robotic boards and sixteen team tables is ideal venue for the event.
- Teams. Seven-member teams are less favorable than five-member teams.
- Career speaker. Guest speakers opening morning sessions with a 10-15 minute brief on their careers is beneficial.
- Early arrivals. Students can watch LabTV while waiting for other buses to arrive to the academy site.
- Robotics. Add robotics refresher training time for junior mentors.
- News flash. A daily academy news board that junior mentors can coordinate and monitor should be considered.
- Rocket launcher. Teams discovered that this year's launcher operated better than last year's (see Fig. 5).

Figure 5.   Water Rocket Launcher

- Internet access. Student teams use the Internet for research into tower building and for generating their briefs.
- Junior mentor project. The academy team agreed that the maze navigation was a good project for this group.
- Laboratory demonstrations. The academy team used this activity, which the students enjoyed, as one of the focal points (see Fig. 6) on the last day of the academy.

Figure 6.   Student Laboratory
Demonstration Participation

4.0 CONCLUSION

The NDEP VDP STEM Summer Academy program provides middle school students with "hands-on" activities that contain challenges in both robotics and engineering problems. Teaming middle school teachers with practicing S&Es at NSWCDD is one

way the VDP STEM Academy generates student interest in math and science. With respect to integrating modeling and simulation (M&S)-based engineering into the program, the planning team used simulation in two life science activities: directing students build a tower model and organizing the robotic challenges so that the students built and programmed a robot to simulate a Navy initiative using sensors and motors. The academy team recommends the following considerations:

1) Involve a math and/or science teacher of the targeted grade early in the activity decision-making process. The teacher can assess the activity with the skill set the students have and determine whether the M&S activity is acceptable.
2) Run through the exercise prior to the planning team's activity review.
3) Generate an M&S activity plan to include purpose, design, analysis, and test.
4) Prepare training material for the mentors that includes all pertinent information regarding the activity.
5) Evaluate the results from first use of the M&S activity, and determine if adjustments are needed or if the activity should be removed from the curriculum. The VDP STEM Academy planning team decided to change the junior mentor robotics project to maze navigation this year, which the junior mentor team successfully concluded.
6) Create team roles that allow students to hold a responsible lead for the team. Roles may include data manager, team supply manager, robotic maintenance manager, and water rocket data recorder.
7) Collect feedback from mentors at the end of the event and use for future improvements.

In conclusion, providing a working environment experience where students can

sense the why, what, and how things are done through interaction with S&Es and math and science teachers can benefit them when they begin making career decisions.

6.2   Using Game Development to Engage Students in Science and Technology

Using Game Development to Engage Students in Science and Technology

John Wiacek
ECPI College of Technology
jwiacek@ecpi.edu

Abstract. Game design workshops, camps and activities engage K-12 students in STEM disciplines that use game engine and development tools. Game development will have students create games and simulations that will inspire them to love technology while learning math, physics, and logic. By using tools such as Gamemaker, Alice, Unity, Gamesalad and others, students will get a sense of confidence and accomplishment creating games and simulations.

1.0 NOMENCLATURE

STEM : Science, Technology, Engineering, and Math

NPC : Non-Player Character, computer controlled character

GUI : Graphical User Interface

STEP : Science and Technology Enrichment Program.

2.0 WHAT'S THE PROBLEM?

In this day and age of science and technology, students are struggling more than ever with math, science, and technology. There are several countries across the world that have noticed a decline in their math and science scores over the last couple of decades. Almost 30 percent of students in their first year of college are forced to take remedial science and math classes because they are not prepared to take college-level courses[1]. The United States has addressed this by creating *"A National Action Plan for Addressing the Critical Needs of the U.S. Science, Technology, Engineering, and Mathematics Education System."* [1] Other countries have taken similar steps in addressing this decline.

"The United States possesses the most innovative, technologically capable economy in the world, and yet its science, technology, engineering, and mathematics (STEM) education system is failing to ensure that all American students receive the skills and knowledge required for success in the 21st century workforce.... To succeed in this new information-based and highly technological society, all students need to develop their capabilities in STEM to levels much beyond what was considered acceptable in the past.... Strengthening STEM education across the nation is critical to maintaining a high quality of life for our citizens and ensuring that Americans remain competitive in international science and technology. Public awareness and action are critical to addressing this crisis." [1]

To respond to this many schools have taken a look at their curriculum and teaching methods and modified their methods to focus on test questions more than the course material [2]. This is not a long-term solution; it just addresses the symptoms (the test scores) rather than the cause (the lack of student understanding).

2.1 Why is this happening?

Over the last fifty years or so what has changed? Why is this becoming a problem? The reasons differ quite a bit for each region in the world. However, as reviewed below, some reasons transcend regional differences.

One of these reasons is the perception that STEM topics are too hard and the student will fail at them. The test is not the problem, but the subject. Before even getting started, many students have already given up. Math has the biggest stigma; most students start their struggle here and this leads to a similar problem in their future science, engineering and technology study since they are based on math.

Another big reason is the real-world examples used in current textbooks and classes. The examples used are not necessarily real-world from the student's point of view and can be hard to understand.

Students are affected greatly by peer pressure, and the majority of the STEM curriculum falls in an unpopular category. This gives students even less motivation to want to succeed in these areas of their studies.

School systems everywhere are being strained with smaller budgets and more students, making it difficult for students to get the help they need. Students learning the material don't get enough help figuring out how to solve the problems that are given to them and how to estimate possible solutions. In the end, many students just turn in a number hoping that it's correct.

These are just a few of the bigger reasons for the decline in STEM education and a need for improvement. This does not account for many other reasons that might be political, cultural, or other in nature.

2.2 Overcoming the Problem
These problems of getting the next generation to like and be good at STEM and see it as a viable possibility for their future careers can be overcome. We have noticed while teaching game and simulation development for over ten years that many of these problems can be overcome fairly easily with the right tools and activities. We take a look at alternative solutions that get students to learn while having fun and making their own games.

2.2.1 The Subject of Game Development
Game development has a great appeal to the majority of the younger students, and this by itself is enough to increase students' confidence in themselves. They consider themselves experts at games from day one. Younger students also have a key knowledge of how a video game works, and

they know what needs to be on the screen to verify that everything is working properly, and can check their own solutions. The average student in the United States has over five thousand hours of gameplay experience [3] giving them confidence when checking their solutions. All this time and experience with video game start them off with a mental preparation for success and very often a passion for video games. Many of them have also played video games that they modified or created new levels for with the games built-in editors. These built-in editors, however, are not very good for teaching because they have been simplified to let anyone create their own levels easily and often lack documentation.

The subject of video games also creates a peer-pressure environment that motivates students to do better. They strive to create games that will impress their peers in the classroom and friends outside of class. Often students will put in more time in making and polishing a game or a simulation than they would anything else in their studies.

We have seen at ECPI that the game and simulation courses have an attendance rate ten percent higher than other technical classes in their major. This also improves their pass rate in these courses by a similar percentage. We have also seen this same trend with camps that we have offered to high school students compared to camps in other fields [4].

Other institutions such as Purdue University, MIT, NASA and more have created games or tools to appeal to the younger generations. They have created these games and tools to bring new blood to their respective industries. Robert Morris University did a survey of fourth graders before and after a STEP camp. The students came to the college four hours every week for eighteen weeks, with their time split equally between technology and sciences [5]. For young students attending camps at ECPI, we found that 36% of the

854

students surveyed before the camp were interested in a career that involved STEM compared to 68% after the camp. This significant change has to be contributed largely to the students starting to like the subject especially at such an impressionable age.

2.2.2 Math the Building Block

Math, being the building block for science, engineering and technology, needs to be addressed first. Game development can be used to teach math using simple classic games such as Breakout, Space Invaders, and Asteroids. The math, logic, and rules in these games are simple making them the ideal choice. We also start by using a game engine that has a graphical programming language requiring no programming experience, such as Gamemaker or Gamesalad. All this allows us to focus on the logic and basic math.

Recreating and expanding many of the early games can be used to introduce the following math concepts:

- Unit Conversion – by converting screen coordinates to world coordinates in game

- Cartesian Coordinates – Using basic movement to calculate character positions

- Functions – can be taught by using the built-in functions and creating new one for power-ups in games

- Sine, cosine – for player rotation

- Statistics and probability – for creating random distributions for NPC, power ups, and dealing damage

These touch just some of the simpler concepts; many more math concepts can be introduced in similar ways. Some of the more advanced math, such as matrix manipulation and the complex number system, may require a more advanced

game engine such as Unity or Torque with some C++ like programming.

Unlike many word problems in math where students struggle to understand the problem, the calculations, and the process for checking a solution, the task is clear and the student can focus on solving the problem. They can check their solution by playing the game and immediately know if there is a problem with their solution. Having these visual cues from the game, they can see what is wrong and can usually solve the problem by themselves most of the time, along the way learning the concepts in greater depth.

In Fig 1 you see one of the games we used to teach the Cartesian coordinate system to fourth and fifth graders based on the classic Breakout game. The students loved the lesson and did not even realize that this involved math. The students set up the ball movement, bat, and blocks in the game using basic GUI commands and using screen coordinates that are very similar to
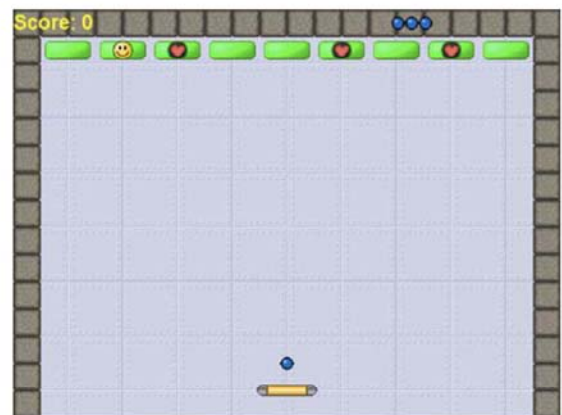
Figure 1

the Cartesian coordinates. Math is where we had the greatest improvement for our third and fourth graders from our STEP program in comparison to their peers from over two hundred students surveyed [5].

2.2.3 Science

Game development requires the direct use of physics, while the other sciences can be included at best indirectly. Most game engines have physics support built in, such as PhysX® and we take advantage of this in our lessons.

The easiest way to start teaching science with games is with Newtonian physics and remaking games such as Super Mario Brothers, Gran Turismo, Donkey Kong, and Lunar Lander. These games can be used to introduce many concepts and formulas such as gravity, acceleration, momentum, friction, mass, force, terminal velocity, torque, and levers. The concepts can be used to compare and contrast using actual physics versus guesstimating, allowing the students to see how the difference affects the behavior of the game. This means more to the student than "real-world examples" or just reading a problem and finding a number that is the solution. When customizing these games, students will be able to manipulate variables such as force, mass, acceleration and time, and get a better understanding of these concepts. Using the game engine will allow them to experiment with variables that would not be safe or even possible in a lab environment.

Other sciences such as biology can also be taught by having the students develop serious games to teach the concepts. These games can range from simple games based on the action genre that students create, to MMO/Adventure type games that will incorporate the science as part of the gameplay that students design. Here are a few games that have used these approaches that students can create:

- "Moon Base Alpha" - NASA's MMO

- "Spore" – Electronic arts

- "Math Blaster" – Davidson (Nintendo and Sega later)

- "The Incredible Machine" – Sierra

- "Marble Madness" - Atari

- "Critical Mass" – Purdue University

Creating some of these games and simulations will require more advanced features in the tools than the most basic game engines can offer, requiring a more advanced tool such as Unity.

Figure 2

Fig. 2 was created by ECPI students. It shows the USS Monitor floating in rough seas. Here not only did the students have to create the ship itself, but had to work out the buoyancy and placing the mass of the ship, turret, and engine in the correct locations. After putting all this together, they saw that the stern of the ship sits lower than the bow. They thought that they had made an error but after looking at historical pictures, they saw this was correct.

2.2.4 Engineering

Engineering is a great place to apply the math and physics that the students learned using game development tools for the initial lessons. Game development tools have support for many of the principles needed for engineering. This allows students to create games and/or simulations for building bridges, ships, car, ecosystems, and more. We have had our students create all of these as simulations or games.

However, only some of the examples can be done with basic tools, creating most of these engineering examples require more advanced tools such as Unity, Torque, or Alice. The ones that can be implemented easily are:

856

- A 2D bridge building game where students have a challenge of building a bridge that spans a river or valley.

- An ecosystem simulation that you set up for a pond where the user has to balance the system so that nothing dies out.

- Simulating traffic at an intersection, stretch of highway, or parking lot to see what bottlenecks are there and look at possible fixes.
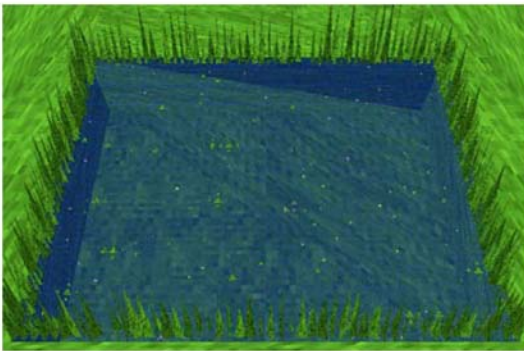
Figure 3

Here you see in Fig 3. a pond simulation that that our students have completed. They have to do the research and get help with biology, not only programming. While creating it, they learn about ecosystems, and after completing the pond, they can experiment with trying to balance it or seeing how easily, an ecosystem can be broken.
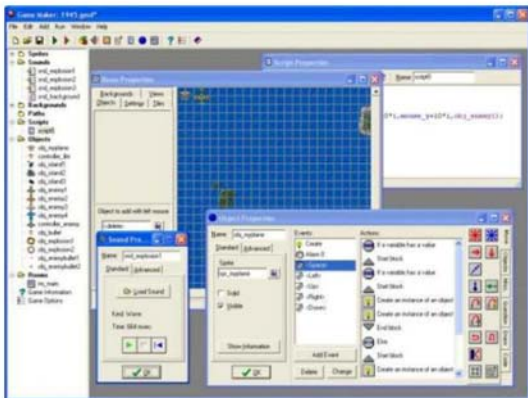
Figure 4

2.2.5 Technology

Using game development tools for teaching the science, engineering and math, teaches technology at the same time by using game development tools. This helps build computer literacy and logic skills with every lesson that the students do.

The computer literacy skills that the students learn in the lessons include: file management, file types, compiling projects, image editing, web research and more. Whatever field the student eventually enters, these skills will be required. While learning the game and simulation tools, the students learn logic, design and programming. These tools introduce students to programming concepts such as functions, variables, inheritance, arrays, and conditional statements.

3.0 RESULTS
With STEM subjects not being the most glamorous in students' eyes, games and game development help address this. The first question to be answered is "Do these activities change students outlook on STEM fields?" We see this happening with all our students, but the ones that are influenced the most are the fourth graders. This is supported by before and after surveys of fourth and fifth graders that show an increase of 32% in their views of science and technology as a career and 38% increase in liking science and technology. These are our most easily influenced group.

The high school students we had in our game development camps were there voluntarily. Camps offering game development were filled to capacity with a waiting list while camps on other STEM topics had seats open, leading to the conclusion that the addition of game development into the curriculum increased student participation. Looking from this perspective game development works to attract high schools students to STEM.

The attendance for ECPI game and simulation classes is just over 10% higher than in other computer science classes for the same students, and their tardiness is reduced as well. All the classes compared started and ended at the same time.

4.0 CONCLUSION

The biggest problem is getting faculty that can teach with the game development tools, especially for the advanced lessons that require the higher-level game and simulation tools. Game development and games will inspire the next generation to like technology. It will make them better at the sciences, engineering and math. The exact size of the benefit is still unknown. However, additional studies need to be done to measure the long-term effectiveness of the initiatives.

5.0 REFERENCES

1 – Author: NSF Date: 10-30-2007 Title: "A National Action Plan for addressing the critical needs of the U.S. science, technology, engineering and mathematics education system". Available: http://www.nsf.gov/nsb/documents/2007/stem_action.pdf

2 – Author: Deubel, P. Date: 9-27-2007 Title: "Test prep and math realities" Available: http://www.thejournal.com/articles/21329

3 – Author: Kris Graft Date: 8-19-2009 Title: "Study: 60 Percent Of Western World Plays Games" Available: http://www.gamasutra.com/view/news/24859/Study_60_Percent_Of_Western_World_Plays_Games.php

4 – Author: ECPI Dates: 2008 – 2010 Camps, registration number for all the camps at ECPI-Newport News

5 – Author: Robert Morris University Dates: 2000 – 2002 STEP Program Survey

6.0 ACKNOWLEDGMENTS

6.3 Computational Experiments for Science and Engineering Education

Computational Experiments for Science and Engineering Education

Charles Xie

The Advanced Educational Modeling Laboratory, The Concord Consortium

qxie@concord.org

Abstract. How to integrate simulation-based engineering and science (SBES) into the science curriculum smoothly is a challenging question. For the importance of SBES to be appreciated, the core value of simulations—that they help people understand natural phenomena and solve engineering problems—must be taught. A strategy to achieve this goal is to introduce computational experiments to the science curriculum to replace or supplement textbook illustrations and exercises and to complement or frame hands-on or wet lab experiments. In this way, students will have an opportunity to learn about SBES without compromising other learning goals required by the standards and teachers will welcome these tools as they strengthen what they are already teaching. This paper demonstrates this idea using a number of examples in physics, chemistry, and engineering. These exemplary computational experiments show that it is possible to create a curriculum that is both deeper and wider.

1.0 INTRODUCTION

Before starting this paper, I feel obliged to explain how I use the terms "modeling" and "simulation." When studying the function or behavior of a system that involves time-varying properties, I prefer to use "simulations." If only the structure or configuration of a system is concerned, I prefer using "models." For example, a protein model describes a protein structure in a stable state, whereas a protein dynamics simulation describes a protein folding or binding process. But this distinction is just personal. For most of the part, the words "modeling" and "simulation" or "model" and "simulation" can be used quite interchangeably.

Simulation-based engineering and science (SBES) is defined as the discipline that provides the scientific and mathematical basis for simulating natural and engineered systems [1]. SBES is increasingly important in accelerating research and development because of the analytical power and cost effectiveness of computer simulation. Advanced simulation tools based on solving fundamental equations in physics are routinely used to understand natural phenomena and solve engineering problems. SBES is an interdisciplinary subject indispensable to the nation's continued leadership in science and technology [2].

SBES, however, has virtually no place in the current science and engineering curriculum frameworks at the secondary level. Despite the fact that modern simulation tools can run on an average computer and be used just like an ordinary application, it is still commonly thought that SBES mandates advanced mathematics and science, uses abstruse jargon, requires monster supercomputers, works only through the esoteric command line, and cannot be possibly taught or used at the secondary level. As a result, most students are not informed of the modern concepts of SBES and are deprived of an opportunity to develop an interest in it earlier in their education. The consequence of this deficiency may have contributed to the erosion of the nation's leadership in SBES and engineering and science in general [1].

One of the purposes of this multi-disciplinary conference is to create a dialogue that will lead to the development of a national plan for integrating SBES into K-20 education, a literacy framework for SBES, and a research agenda.

How to integrate SBES into the secondary curriculum is a tricky, open question. The U.S. science curriculum is often criticized to be "a mile wide and an inch deep." [3] Incorporating SBES into the existing curriculum must reconcile with the growing national consensus around the need for "fewer, higher, clearer" education standards [4]. The majority of science educators will need to be convinced that the integration of SBES into their curricula will be realistic, constructive, and helpful.

This paper suggests an implementable integration strategy that uses the products of SBES to help teachers achieve their goals in deepening students' conceptual understanding, as set by the new National Science Education Standards [4], and in doing so, conveys the core values of SBES to both students and teachers. To be practical, I will demonstrate how this strategy may work using a few concrete examples in physical science and engineering. Each of these examples shows how students can use a visual, interactive, and constructive simulation tool to conduct and design a sequence of computational experiments to explore a broad set of concepts in great depth. These examples are based on my work on creating simulation tools for science education using research-grade computational methods for solving fundamental physical laws. Resorting to first principles in physics to build educational tools may be considered as overkill by some educational developers, but it is essential to bringing learning experience with authentic science to the classroom. Of more importance, it opens up new, profound opportunities for deeper learning that will not exist otherwise, as the examples will show. Having a strong root in SBES, the examples may be used as stimulating introductions of the theory and practice of the corresponding computational methods behind the scene. By employing well-established pedagogical principles such as design-based learning [5] or constructionism [6] in the curriculum to encourage students to create simulations to answer questions, solve problems, or design systems, learning secondary science and learning SBES can have a significant overlap and a mutual enhancement can consequently occur.

2.0 WHAT IS A COMPUTATIONAL EXPERIMENT?

A computational experiment is a computer simulation of a real experiment or a computer implementation of a thought experiment. Computational experiments complement analytical theories and real experiments by providing a tool for explaining
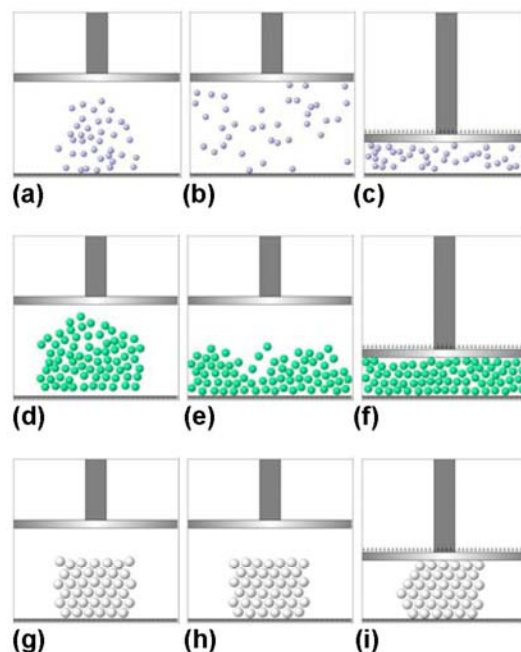
Figure 1: A computational experiment for studying states of matter. The images show if and how molecules of a gas, liquid, and solid students place into a container will fill the space and if they can be compressed. Gas: a, b, c; liquid: d, e, f; solid: g, h, i.

what was observed and predicting what will happen. With all these explanatory and predictive power, computational models, the machinery of computational experiments, have become a key element of science [7]. They are now also considered a pillar of science education—in parallel to mental and physical models—if appropriate user interfaces are provided to make them accessible to *every* student.

Figure 1 shows a computational experiment for investigating the molecular mechanisms underlying some key macroscopic properties of a gas, liquid, and solid. Students can add different types of molecules into a container maintained at a given temperature (Fig. 1a, d, g). If the molecules form a gas, they will quickly fill up the entire container and move rapidly and chaotically (Fig. 1b). If they form a liquid, they will fill up the lower part of the container and move randomly around each other (Fig. 1e). If they form a

solid, they will stay together to form an ordered structure (Fig. 1h) and vibrate around some fixed positions. Now if we apply some pressure through a piston, the gas will be significantly compressed (Fig. 1c), whereas the liquid and solid can hardly be (Fig. 1f, i).

As you can see from Fig. 1, the most important properties of the three states of matter are *all* explained on the molecular basis in a single computational experiment using just the mouse. All the related concepts such as the Kinetic Molecular Theory, diffusion, lattice vibration, and pressure, however disparate they may be in a textbook, are linked and unified in the very same computational model and can be manifested or inquired through the computational experiment. With a variety of visualization and analysis tools, many more details can be discovered. For instance, students can select a molecule and visualize its trajectory to examine how it moves in different states. The force vector, velocity vector, or kinetic energy shading on each atom can be shown to provide further information about molecular collisions.

3.0 GOING DEEPER, REACHING WIDER

The example shown in Fig.1 demonstrates how a computational experiment can transform the way we teach states of matter. The new way is simple enough to be applicable even to lower grades. But the computational model has a lot more to offer. The example shows only the tip of the iceberg. What lies beneath is a great number of opportunities to go deeper and reach wider.

The computational experiment shows how matter in different states behaves without further explaining what makes them behave so. Figure 2 shows the equilibrium conformations of a molecular system corresponding to three different strengths of interatomic interactions, which students can adjust through a graphical user interface shown in Fig. 2d. This step takes students down to the level of studying the states of matter in terms of interaction and energy. It offers an
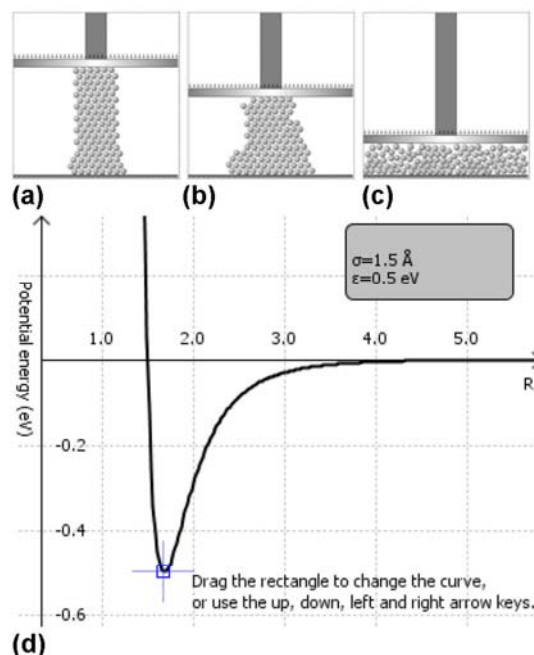
Figure 2: The effect of interatomic interactions. The potential well depths are: (a) 1 eV, (b) 0.5 eV, and (c) 0.01 eV, respectively. A shape of the interatomic potential function is shown in (d).

intuitive explanation of the idea that stronger attractions result in stronger materials.

With a few more mouse clicks, many other concepts can be studied. The material strength can be tested by increasing the pressure on the piston and observe how the solid is deformed, which explains plasticity. By increasing the temperature, the solid will become more ductile, which then explains the business of a blacksmith. When the temperature continues to rise, the solid will melt down into a liquid to fill the container, showing a phase change. If the temperature keeps rising, the liquid will turn into a gas and start to push the piston up, causing a dramatic volume expansion. This explains how heat can do work. Gas laws can also be studied. Students can discover that under the same pressure, higher temperature will result in greater volume, and under the same temperature, higher pressure will result in smaller volume. Students can even ask questions not covered in most curricula about gas laws. For example, the mass of

the molecules can be changed to see if it affects the equilibrium volume of the gas under the same pressure and temperature. Visually, it seems that the mass of molecules should affect the volume, as more massive molecules appear to move more slowly. The frequency at which they bump into the piston is, therefore, lower. But a computational experiment simply shows that molecular mass has no effect, just as suggested by the Ideal Gas Law: $PV=nRT$. This is not easy to figure out just by thinking. Even if one can reason correctly that a more massive molecule will deliver a greater impact when it collides with the piston, some mathematical work is still needed to prove that the two effects cancel out exactly. Where the mathematical skill prevents the majority of students from investigating further, the computational experiment helps them move forward. Beyond investigating the effect of mass, students can even adjust the atomic radius and the van der Waal attraction to explore how the equation of state deviates from the Ideal Gas Las. A comparison of two gases that differ only in atomic radius or van der Waals potential energy shows the effect of excluded volume or intermolecular attraction.

It is probably not an exaggeration to assert that the possibility of inquiry is only limited by the imagination of the experimenter. The breadth and depth of the science embodied in this computational experiment, along with the ease of inquiry afforded by a graphical user interface, suggest the feasibility of creating a curriculum that is both wide and deep using this type of simulations.

This prospect would not have been possible without using authentic science to build the educational tool. The scientific power demonstrated above originates from the application of the molecular dynamics method [8], which is an important tool in SBES for studying nanoscale science and engineering [9]. The computational experiment described above was designed and conducted using the *Molecular Workbench* software,

which has a classical molecular dynamics tool tailor-made for science education [10].

I hope you are reasonably inspired by this introductory example. In the following section, I will discuss more about the need to use true science to build educational tools and the implication of this to SBES education. Then I will show more examples in other disciplines in later sections.

4.0 WHY USE ROCKET SCIENCE TO BUILD EDUCATIONAL TOOLS?
The educational software market is largely dominated by cartoon movies, animations, and games. Most of these media were usually produced with multimedia effects as the paramount design goal in the developers' minds. Although many claim to offer computer models, most are insufficiently intelligent to have the desired predictive power. For example, an animation that the user cannot change has only illustrative power but no predictive power at all. An interactive model or game designed to have a limited number of outcomes scripted by the developer can explain the preset causality but nothing beyond. In making the rules for determining the outcomes, many developers seldom perceive a need to exploit the "rocket science"—advanced mathematics and computation based on first principles in science. In the following, I will explain why there is such a need.

A first principle is a foundational scientific law from which many phenomena can be explained and many propositions can be derived. For example, Newton's equation of motion is the first principle in classic mechanics—everything in the domain of classic mechanics can be explained by solving it analytically or numerically. The classical molecular dynamics method that powers the computational experiment shown in the previous sections is based on solving Newton's equation of motion for a system of interacting particles that model atoms and molecules. It is responsible for all the simulations in the computational experiment that explain the myriad of concepts. There is no need for

862

students to program all those—everything just emerges from the number crunching done by the computational engine according to the ruling equations.

The molecular dynamics method was, however, not originally intended for education. It was developed to help scientists and engineers explore nanoscience and engineer nanosystems [11]. The generations of computational scientists who contributed to the theory and practice of the method presumably did not anticipate that one day the method would find its place in thousands of schools all over the world. But this should not be surprising at all. In fact, science education and scientific research share a common goal: to understand how the world works. It is, therefore, no wonder that a research tool can be successfully converted into a learning tool.

Perhaps the single most important reason for using first principles to build educational tools is that all the power of explanation, prediction, and creation embodied in them will then be given to *every* student. What else is more important in education than passing students the greatest power and deepest wisdom brought to us by the most brilliant figures in the history of science and engineering? Now that the information technology has empowered us to deliver them through computing, an unprecedented opportunity to revitalize science and engineering education using this enabling technology is right upon us.

Unfortunately, this opportunity is often underappreciated in the educational world. Using first principles to build interactive media is not part of the design guidelines for the mainstream. There are many more domains of science and engineering where the curriculum needs to be transformed in a way similar to what was described in the previous sections. Enormous volumes of literature have existed for how to simulate real world problems by numerically solving fundamental equations such as the Navier-Stokes equation for fluid dynamics and the

Maxwell equations for electrodynamics and photonics. Sadly, there has been little investment and interest in making those powerful methods usable by students and the public at large.

5.0 NEW STANDARDS, NEW OPPORTUNTIES

There is now a chance for SBES to prove its value in education at a large scale. The new National Science Education Standards has put forward a "more coherent vision" of science education [4]. The framework calls for educators to focus on a limited number of core ideas and give time for students to engage in scientific investigations and achieve depth of understanding. It emphasizes that learning about science and engineering involves the integration of both content knowledge and the practices needed to engage in scientific inquiry and engineering design. It recognizes learning as an ongoing developmental progression. Exactly how this vision will turn into actions is a critical question. Given the fact that the results from the 1996 Standards have been disappointing [12], the development of creative ideas to implement the new framework in the curriculum will be more important than ever.

The recommendation of learning from core ideas is not an overstatement. Richard Feynman once noted: "I am inspired by the biological phenomena in which chemical forces are used in repetitious fashion to produce all kinds of weird effects (one of which is the author)." Indeed, the unity of science—that everything can be derived from some basic rules however their appearances and representations may differ—is probably the most profound nature of science. For students to achieve deeper learning, the curriculum must be structured to reflect this nature. Learning should focus on the basic rules as suggested by the principle of Occam's razor and science should be taught as a way of thinking based on them rather than a large collection of facts.

The idea that complexity arises from unity is the holy grail of SBES, too. A simulation

863

program uses the same code in repetitious fashion to produce all kinds of results to explain the weird effects observed in the real world. The parallelism between the inner workings of a simulation and the conceptual structure of knowledge it simulates makes it an ideal cognitive tool. Being the scientific discipline about simulations, SBES can be a

(a)

(b)

Figure 3: A computational experiment for studying the inverted single (a) and double (b) pendulums on an oscillatory

cornerstone for building the technological foundation of the new science framework.

Although the framework literally stresses the importance of simulations as a creative engine that drives the scientific and engineering enterprise, simulations are considered more as individual expressions of concepts than as possible systematic solutions to realize the vision. In fact, a simulation tool can be used not only as an inquiry tool to teach known facts as shown by the example of the states of matter, but also as a research tool to explore the unknowns. The latter provides opportunities to teach students to think and practice like scientists and engineers, a wish reiterated in the new standards. A simulation tool constitutes a computational laboratory in which students will ask questions, identifies problems, find solutions, and analyze results. The following computational experiment about the inverted pendulum, a classic problem in dynamics and control theory, shows an example of how this may work for students.

A pendulum is what every student learns in physics. An inverted pendulum is what we get when we turn it upside down after it has stopped swinging. We know this upright position will not be stable. Any blow will knock the mass off from that position. But there are ways to stabilize it. One way is to fix the pivot on a base and rapidly oscillate the base up and down. If the oscillation is simple harmonic motion, the pendulum's motion is described by the Mathieu equation, which has very complex solutions that tell how high the frequency and how large the amplitude should be in order to maintain stability.

An interesting question that immediately follows is: what will happen if we invert the double pendulum? This is a question that would almost instantaneously excite any mathematician or physicist who knows the importance of a double pendulum in nonlinear dynamics and chaos theory. The study of an inverted double pendulum may well worth a Ph.D. thesis. But never mind about the intimidation of the mathematical com-

864

plexity, a simulation tool can easily bring students to where the mathematicians or physicists stand. Fig. 3a shows a computational experiment designed to test the stability of the inverted pendulum. The amplitude and frequency of the oscillation of the base and the perturbation on the mass can be adjusted to study the dynamic stability. If students want to test an inverted double pendulum, they can just append another mass to the mass as shown in Fig. 3b. Can this chaotic system be stable in the face of the butterfly effect? In other words, does unpredictability necessarily imply instability?

I will leave this interesting question for you to ponder. The central point of this example is that it demonstrates the enormous educational value of a simulation tool in supporting *all* levels of scientific investigations, which in this case range from a well-known problem (a single pendulum) to a less-known problem (an inverted pendulum) and then to an unknown problem (an inverted double pendulum).

6.0 ENGINEERING DESIGN
Engineering is considered an integral part of the new standards. Engineering design is a creative and iterative process for identifying and solving problems under various constraints. It is a core element to engineering like inquiry to science.

Modern engineering methodologies heavily involve SBES. Computer simulations are often used to screen solutions to a particular problem or optimize a design before building the real system. One of the most successful applications of SBES to solve engineering problems is computational fluid dynamics (CFD).

Deeply at the core, CFD involves sophisticated numeric methods for solving the Navier-Stokes equation such as the finite difference method or the finite element method. While it may be inappropriate to teach the nuts and bolts of these numerical methods at the secondary level, it is desirable to teach how engineers use these tools to

solve problems. This is similar to teaching how to use CAD tools to design structures without teaching the computational geometry under the hood. In fact, simulated fluid flows, when visualized, are intuitive enough for students to understand. For example, the Kármán vortex street is mathematically complicated but probably not incomprehensible as similar patterns are not uncom-
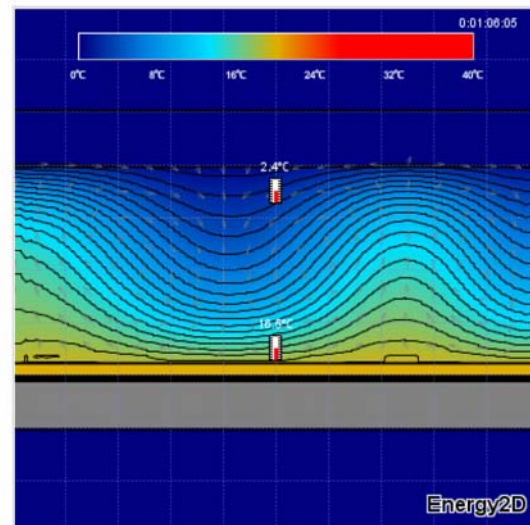
(a)
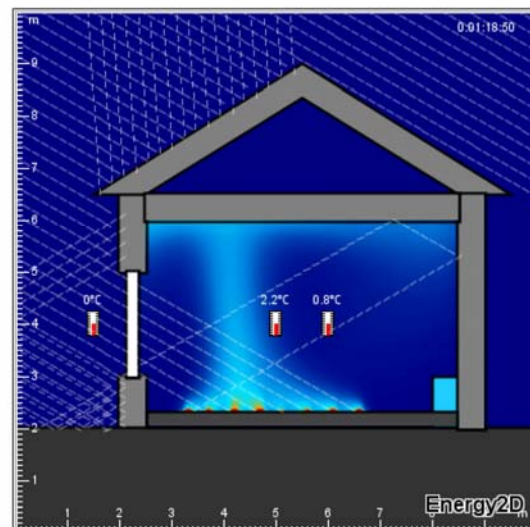
(b)

Figure 4: (a) A computational experiment for studying the Rayleigh-Benard convection pattern of a fluid between a cold plate and a hot plate. (b) A computational experiment for studying solar heating of a house through a window.

mon in everyday life. The key is to develop the user interfaces that will make these tools visually and manually accessible to students and, importantly to engineering, make it possible for students to design with them.

Ironically, while there are numerous CFD tools developed for professionals to tackle engineering problems, little has been done to make a CFD tool for average secondary students to learn with the powerful method. With all the dramatic, artistic effects of flow that worth the volume of a book [13], CFD has enormous potential to bring fun and enjoyable learning experience to the classroom. This potential should not be left untapped any longer.

Figure 4 shows two computational experiments designed using an educational CFD tool called *Energy2D* I created to move towards the goal of providing a versatile CFD laboratory for students. The tool allows the user to set up a 2D thermal system such as a house and run CFD simulations to assess the energy flow within it. Ultimately, this tool will be integrated into a 3D environment for students to evaluate and optimize their designs for real world applications such as a green building, an internal combustion engine, or a cooling system for a CPU. But even the 2D simulations show the richness of science and engineering concepts students can explore. For example, the temperature difference between the hot plate and the cold plate in Fig. 4a can be adjusted to test when the convective pattern becomes turbulent. The angle of sunlight can be changed to investigate solar heating of a house at different times of the day, as shown in Fig. 4b. Virtual thermometers can be placed in the model house to monitor temperature changes at any locations to check if a passive solar design meets the requirements of thermal comfort. When the house is heated internally, a virtual thermostat can be added to maintain the indoor temperature. Students can evaluate energy costs under various conditions and constraints. For example, if the environmental

temperature is one degree colder, how much more energy will be needed to keep the house as warm? If the sun is shining into the house through a window, how much energy can be saved? With a learning environment like this, many engineering problems and design challenges can be posed to students.

7.0 DISCUSSIONS
In this section, I would like to make a few further suggestions on how to foster SBES's role in the upcoming reform of science and engineering education.

7.1 Blurring the line between research and education
Modern personal computers have become fast enough to run serious simulations that involve intense computation. The ubiquity of multicore processors in the near future will only make computers even more powerful. What can science education benefit from personal computers that rival supercomputers only one or two decades ago?

Scientific simulation software programs, the direct products of SBES, can capitalize from ubiquitous multicore computing [14]. Powerful simulation tools running on powerful multicore computers have the potential of becoming one of the most powerful scientific investigation tools for education, just like their supercomputing counterparts to scientists and engineers but only much more accessible. Recent studies revealed that children are born investigators, capable of reasoning in a surprisingly sophisticated way about the natural world based on direct experiences with the physical environment [15]. If easy-to-use graphical user interfaces, or even the more modern touch interfaces, are provided, there is fundamentally nothing that can prevent them from becoming amateur scientists and engineers. With powerful simulation tools at students' fingertips, the line between research and education will be blurred and science can then be taught as the way it is. When the difference between learning and investigation diminishes, the curriculum can become a fantas-

tic journey to discover the jewels of science and engineering.

7.2 Learning from games but not counting on them

Outside education, game developers have adopted first principles far more quickly and aptly. Games need to have realistic look-and-feels in order to be competitive in the market that always demands better realism. Major graphics libraries already provide excellent lighting functions. Realistic motions and flows powered by physics engines such as *Lagoa Multiphysics* and *Maya Fluid Effects* are now available for animators. Real-time physics games such as *Algodoo* and *Crayon Physics* are making inroads into classrooms at the lightning speed. There is a lot to learn from the success of games.

But game developers are only interested in technologies that can entertain the player and are not necessarily willing to invest on things like quantum mechanics, genome dynamics, or climate modeling. The future of science and engineering education cannot rely on the good will of the game industry. It lies in the hands of a strong alliance between scientists, engineers, and educators. The SBES community is among the foremost groups that can lead the charge and bridge the gap.

7.3 Learning by creating simulations

One of the most important affordances of open-ended simulation tools such as the *Molecular Workbench* software is the ability for students to create their own simulations. Only through the creation process can learning be maximally deepened and personalized. This kind of simulation tools offers an important method to implement the theory of constructionism [6] or learning by design [5] for the scientific fields they grow out from. A good user interface will allow students to design any computational experiments to test their own hypotheses. In our field tests with the *Molecular Workbench* software, we found students became very creative once they were given creative tools. We were initially concerned that stu-

dents would just copy each other's design or duplicate a demo, thus invalidating the pedagogy. But this did not happen. On the contrary, it turned out that a surprisingly high percentage of students came up with creative solutions that even professional scientists had never thought of.

Learning science by designing new simulations has a substantial overlap with learning the practice of SBES. Both aim at using simulations to prove a concept or test a design. The only difference is that the mission of the SBES professionals is to explore the unknowns on behalf of the society whereas students are only exploring on their own behalves what is probably only new to themselves. But this difference is not really fundamental, except for a cutting-edge research task requires a higher skill level and a broader scope of knowledge. If an educational tool employs true SBES, some of the modeling skills and knowledge students learn from creating their own simulations in classrooms may end up transferring into SBES literacy and skills. For instance, students may learn some basic data analysis skills that are commonly needed to understand a simulation in both a research setting and an educational setting.

Creating simulations for learning science provides the necessary contextualization for SBES to be adopted in the science curriculum, as well as the driving force for engaging students and teachers to pursue SBES. Nothing is more rewarding than seeing one's own simulations at work. And nothing is more satisfying than seeing one's own students succeeding in doing impressive work. As such, students are more likely to be motivated to learn more deeply and dig under the hood of SBES in order to improve their own simulations. And teachers are more likely to adopt the tools if they see their potential.

8.0 CONCLUSIONS

This paper suggests a strategy for integrating SBES into the science curriculum using computational experiments as the facilita-

tors. How to implement the strategy under the conceptual framework of the new National Science Education Standards was discussed and substantiated by a number of concrete examples in physical science and engineering. It was elucidated that powerful scientific simulations can serve as cognitive tools for learning science and engineering more profoundly. An important outcome of adopting computational experiments in the science curriculum will be that they will also provide pathways to teach the principles and practices of SBES.

9.0 ACKNOWLEDGMENTS

10.0 REFERENCES

[1] NSF Blue Ribbon Panel on SBES, "Simulation-based engineering science: Revolutionizing engineering science through simulation," NSF, Washington, DC, 2006.

[2] S. C. Glotzer, S. Kim, P. T. Cummings, A. Deshmukh, M. Head-Gordon, G. Karniadakis, L. Petzold, C. Sagui, and M. Shinozuka, "International Assessment of Simulation-Based Engineering and Science," Baltimore, MD, 2009.

[3] W. H. Schmidt, C. C. McKnight, and S. A. Raizen, "A splintered vision: An investigation of US science and mathematics education," Boston, MA/Dordrecht, The Netherlands/London, UK: Kluwer Academic Press, 1997.

[4] The National Academies, "A Framework for Science Education: Preliminary Public Draft," The National Academies, 2010.

[5] J. L. Kolodner, D. Crismond, B. B. Fasse, J. T. Gray, J. Holbrook, M. Ryan, and S. Puntambekar, "Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting a learning-by-design curriculum into practice," *Journal of the Learning Sciences,* vol. 12, pp. 495–548, 2003.

[6] S. Papert, "Situating Constructionism," in *Constructionism*, I. Harel and S. Papert, Eds. Norwood, NJ: Ablex Publishing Corporation, 1991.

[7] P. Thagard, *Computational philosophy of science.* Cambridge, MA: MIT Press, 1988.

[8] D. C. Rappaport, *The Art of Molecular Dynamics Simulation.* Cambridge, UK: Cambridge University Press, 1997.

[9] E. K. Wilson, "Computational nanotechnology: modeling and theory are becoming vital to designing and improving nanodevices," *Chemical & Engineering News,* vol. 81, pp. 27-29, 2003.

[10] Q. Xie and R. Tinker, "Molecular dynamics simulations of chemical reactions for use in education," *Journal of Chemical Education,* vol. 83, pp. 77-83, January 2006.

[11] K. E. Drexler, *Nanosystems: Molecular Machinery, Manufacturing, and Computation*, 1st ed.: Wiley, 1992.

[12] B. Alberts, "Reframing Science Standards," *Science,* vol. 329, p. 491, 2010.

[13] M. Van Dyke, *An Album of Fluid Motion.* Stanford, CA: The Parabolic Press, 1982.

[14] J. Larus and D. Gannon, "Multicore Computing and Scientific Discovery," in *The Fourth Paradigm: Data-Intensive Scientific Discovery*, T. Hey, S. Tansley, and K. Tolle, Eds.: Microsoft Research, 2009.

[15] R. A. Duschl, H. A. Schweingruber, and A. W. Shouse, *Taking Science to School: Learning and Teaching Science in Grades K-8*. Washington DC: The National Academies Press, 2007.

6.4 Digi Island: A Serious Game for Teaching and Learning Digital Circuit Optimization

Digi Island: A Serious Game for Teaching and Learning Digital Circuit Optimization

Michael Harper, Joseph Miller, and Yuzhong Shen
Old Dominion University
Department of Modeling, Simulation, and Visualization Engineering
mharp011@odu.edu, jmill103@odu.edu, yshen@odu.edu

Abstract. Karnaugh maps, also known as K-maps, are a tool used to optimize or simplify digital logic circuits. A K-map is a graphical display of a logic circuit. K-map optimization is essentially the process of finding a minimum number of maximal aggregations of K-map cells with values of 1 according to a set of rules. The Digi Island is a serious game designed for aiding students to learn K-map optimization. The game takes place on an exotic island (called Digi Island) in the Pacific Ocean. The player is an adventurer to the Digi Island and will transform it into a tourist attraction by developing real estates, such as amusement parks and hotels. The Digi Island game elegantly converts boring 1s and 0s in digital circuits into usable and unusable spaces on a beautiful island and transforms K-map optimization into real estate development, an activity with which many students are familiar and also interested in. This paper discusses the design, development, and some preliminary results of the Digi Island game.

1.0 INTRODUCTION

Electronic games are a pervasive aspect of American culture and entertainment: as many as 65 percent of American households play games [1]. With a revenue of $20.2 billion in 2009 [2], the game industry has evolved into an important sector that is larger than the film industry. The passion for games can be exploited for more vital purposes, such as education, training, and marketing, via "serious games." Game-based learning uses serious games with defined learning outcomes and objectives. The values of game-based learning have been recognized by organizations such as National Science Foundation and National Research Council [3-5]. NSF considers games as an important form of cyberlearning platform and technology [6]. The latest ground-breaking game technologies, such as Nintendo Wii and Microsoft Kinect, have significant impact on gamers, transforming gameplay into a more positive and healthy experience. Now the gamer demographics cover every age group [1].

Digital circuits are embedded in almost all electronic equipment and devices in use today, such as computers, MP3 players, and digital cameras. Digital circuit optimization, or simplification, is a process to reduce the complexity of the digital circuits so that electronic devices will have a smaller size (thus less weight) and less power consumption (thus prolonged battery life). Various techniques have been developed in the last several decades for digital circuit optimization. Among them, the Karnaugh map is the standard method to teach digital circuit optimization in introductory digital circuit courses because its graphical representations facilitate logic simplification, providing an intuitive and systematic way for circuit optimization. However, many students have difficulties learning circuit optimization using Karnaugh maps merely because it is the first time for them to be exposed to Karnaugh maps and class lectures do not provide enough coverage and exercises. A serious game that exploits students' interest and curiosity with games would be helpful for learning circuit optimization using Karnaugh maps.

As part of a Senior Design Project at the Department of Electrical and Computer Engineering of Old Dominion University, the authors developed a serious game, Digi Island, to aid teaching and learning digital circuit optimization using Karnaugh Maps. This paper discusses the design and

development of the game and presents some preliminary results.

2.0 BODY

2.1 Digital Circuit Optimization Using K-maps

There are two different types of electronic circuits: analog circuits and digital circuits. Analog circuits represent and process information in continuous or analog form, while in digital circuits, information is represented and processed in discrete (most commonly binary) forms. Most components of modern electronic devices are digital circuits and the transformation from analog to digital is still underway. Use the media for storing music as an example. The traditional audio cassette tapes store music as analog signals, while the MP3 music players that became extremely popular in the last decade store music as digital signals. Compared with analog systems, digital systems have many advantages in terms of flexibility, programmability, computational capability, numerical accuracy, information storage and retrieval, error detection and correction, and miniaturization [7].

The same logic function can be implemented by different digital circuits with varied complexities. Thus, it is necessary to find the optimal digital circuit with minimum complexity for the desired function. Such process is called digital circuit optimization or simplification, which is important to reduce the size and weight of electronic devices and prolong their battery life (just think about the evolution of cell phones since their inception in terms of size, weight, and battery life). Circuit optimization is an important theoretical concept covered in introductory digital circuit courses. Various techniques have been developed in the last several decades for digital circuit optimization, including Boolean algebraic manipulation and minimization, Karnaugh maps, Quine-McCluskey, Petrick's algorithm, Espresso, and others [7]. Among them, the Karnaugh map is the standard method to teach digital circuit optimization in introductory digital circuit courses because it is a graphical representation that facilitates logic simplification, providing a standard and systematic way for circuit optimization.

Karnaugh maps, also known as K-maps, are graphical representations of logic circuits, which can also be represented by Boolean algebraic expressions. The sides of a K-map represent circuits inputs, while each cell of a K-map represents the corresponding circuit output with values of 1 or 0. Figure 1 shows a K-map representing a circuit with 4 inputs. K-map optimization is essentially the process of finding a minimum number of maximal aggregations of K-map cells with values of 1 according to a set of rules. Circuit simplification using K-maps requires understanding of several key concepts, including implicant, prime implicant, and essential prime implicant [7-8]. To find the optimized expression of a K-map, all prime implicants must be identified first. The optimized expression is the logic sum of all essential prime implicants and other prime implicants consisting of minterms not included in the essential prime implicants. The remaining nonessential prime implicants can be determined using a selection rule that minimizes the overlap among prime implicants [8]. K-maps are introduced in introductory digital logic circuit courses, such as ECE 241 *Digital Logic Circuit* at Old Dominion University. Without understanding and using K-maps proficiently, students are likely to fail in this introductory course and more advanced digital circuit courses.

AB\CD	00	01	11	10
00	1	0	0	1
01	0	1	1	0
11	1	1	0	1
10	1	0	0	1

Figure 1 A 4-variable K-map

870

2.2 Game Design

The goal of the game Digi Island is to provide both a formal introduction to K-maps and an engaging game setting that encapsulates the K-maps. To this end, Digi Island is designed as a construction based strategy game [9-10]. The game has three modes: Tutorial Mode, Practice Mode, and Play Mode. The Tutorial Mode provides several tutorials about the K-map through exemplary circuits with 2, 3, and 4 inputs; it identifies implicants, prime implicants, and essential implicants in these circuits, illustrates the procedure of selecting essential prime implicants and nonessential implicants using the selection rule, and finally generates the Boolean expression for the optimized logic circuit. The Practice Mode first displays rectangles in K-maps with 2, 3, and 4 input variables and asks the player to identify them as implicants, prime implicants, or essential prime implicants. The player needs to find all implicants, prime implicants, and essential prime implicants directly. User interfaces should be provided to start and end identification of these terms. The player also needs to generate the optimized Boolean expression. User interfaces should be provided to the player to enter input variables, their complements, and the logical OR operation. Both the Tutorial Mode and Practice Mode provide instructions and practices using standard representations of K-maps that are used in regular classrooms.

The Play Mode is the fun part of the Digi Island game. There are no more K-maps in the Play Mode and instead the player sees a beautiful island (called Digi Island) in the Pacific Ocean. In the game, the player is an adventurer to the Digi Island and will transform it into a tourist attraction by developing real estates, such as amusement parks and hotels. However, not all places can be exploited by the player, such as rocks and reserves for wild lives. The player is given a map of Digi Island that labels all usable and unusable spaces. Large buildings bring about better financial outcomes as they provide more efficient utilization of the space. Thus, the goal of the player is to construct a minimum number of building as large as possible covering all the usable spaces, while satisfying the regulations of the Pac Republic, which exercises sovereignty over the Digi Island. Some sample regulations are listed below.

- All the spaces occupied by each building must be adjacent to each other.
- The number of usable spaces (blocks) in each building must be a power of 2.
- Sharing spaces between adjacent buildings is allowed, but should be minimized.

Initially, the player has a certain amount of cash that can be used to exploit the island and construct buildings on the usable spaces. Depending on the importance of the buildings, they have different values and are represented differently, e.g., single house, skyscrapers. Larger buildings are more valuable and generate more profit for the player and thus more points.

In addition to the game design discussed above, the game has a number of other requirements. The game must be deployed and playable on personal computers and Microsoft Zune HD media players. The touch screen of the Zune HDs must be utilized as input device. Sound effects must be included to provide user feedback. Voice instructions should be provided as well.

2.3 Game Development

The development of the game Digi Island contained two major components: front end and back end. The front end mainly contains a graphical user interface that displays menus, tutorials, and K-maps. The front end has different user input modes for K-map manipulation and provides user feedback for their actions. The back end contains the major logic for digital circuit optimization, that is, for an input digital circuit, the back-end generates the optimized Boolean expression, compares that with the player's solution (answer), and

871

provides feedback to the player. In the following, the back end is discussed first.

2.3.1 Logic Circuit Optimization

Since the goal of the game is to help the player learn logic circuit optimization using K-maps, the game must know the final optimized Boolean expression for an input logic circuit. Although K-maps are effective visual tools used by humans for manual circuit optimization, they are not suitable to be implemented on computers for automatic circuit optimization. Several other logic circuit optimization algorithms have been developed and here we mainly discuss two widely used methods: Espresso algorithm and Quine-McCluskey algorithm.

The Espresso algorithm is the de facto industry standard for circuit optimization and it was initially created by Brayton et al. [11] and later revised by Rudell of University of California at Berkeley [12]. The Espresso algorithm is a heuristic but effective algorithm in terms of memory usage and computational complexity. Although it does not guarantee to produce an optimized circuit, in practice it always leads to a solution that is either optimal or very close to optimal. The source code of the Espresso algorithm in C programming language is available for downloading from the University of California at Berkeley [13]. There were several options to utilize the Espresso algorithm in the game: 1) porting the Espresso C source code to C# in the game, 2) compiling the source code into libraries and calling the libraries in the game, and 3) calling the executables of Espresso directly. Considering the time restriction (only one semester) of this senior design project and that the Espresso is a complex algorithm, it was not feasible to port the Espresso source code in C to C# in this project. Options 2 and 3 worked for the game running on the personal computers (PCs), but didn't for the game on the Zune HD player since there is no C compiler that can generate object code for the Zune HD platform. Considering these obstacles, the

Espresso algorithm was not deemed suitable for this project.

The Quine-McCluskey algorithm is another widely used method for circuit optimization [14-15]. The Quine-McCluskey algorithm is designed to work similarly to the human brain's pattern recognition. It is a systematic method that guarantees to produce the optimized Boolean expression and its tabular form makes it suitable for computer implementation. The Quine-McCluskey algorithm first finds all implicants with n variables, then combines some of them to implicants with $n - 1$ variables, and continue this combination process until all prime implicants are found. The algorithm then identify all essential prime implicants using a prime implicant chart. However, the circuit is not fully optimized or minimized yet as the remaining prime implicants may still have overlap. A covering procedure is then utilized to select a minimum number of remaining non-essential prime implicants in the prime implicant chart so that the circuit function is fully covered. Unlike the Espresso algorithm, No authoritative source code exists for the Quine-McCluskey algorithm. Since it is a systematic method that is straightforward to implement, the team decided to develop the C# code for the Quine-McCluskey algorithm from scratch. The Microsoft .NET Framework and C# programming language were utilized to develop the code. C# is an object-oriented programming language drafted by Microsoft and has been approved as a standard by ISO.

2.3.2 Game Play

Microsoft XNA Game Studio is a game development toolkit for Windows, Xbox 360, Zune HD players, and Windows phones. The XNA Game Studio consists of two parts: XNA Framework and a set of tools and templates for game development. The XNA Framework is an extensive set of libraries for game development based on the Microsoft .NET Framework. It encapsulates low-level technical details so that game developers can focus more on

872

content and high-level development [16]. XNA provides templates for common tasks, such as development of games, game libraries, and game components. It also provides utilities for cross-platform development, publishing, and deployment. Developers can make use of both the XNA Framework and the .NET Framework in the game with the former for game-specific tasks such as graphics rendering and managing inputs and latter for more general programming tasks.

XNA Game Studio is a powerful tool for rapid development of cross-platforms and it was selected as the tool to implement the graphical user interface and game play. XNA Game Studio 3.1 was used and the latest version is 4.0 Beta. Object oriented programming (OOP) was utilized and a number of classes were developed to represent different game scenes, graphical user interface, user inputs, K-map, and logic circuit optimizations. XNA Game Studio provides a fundamental class Game that handles game logic update and drawing. A class diagram of the game is shown in Figure 2.
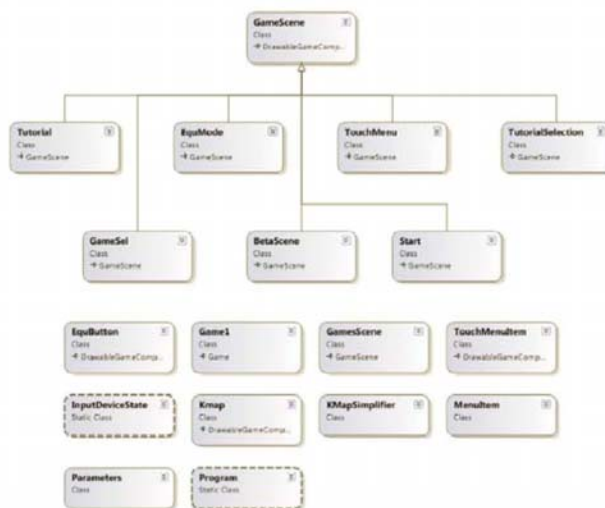
Figure 2 A class diagram of the game.

2.4 Preliminary Results

A preliminary prototype has been developed as a product of this Senior Design Project in the Spring Semester of 2010. Some screen captures of the game are shown in Figure 3.
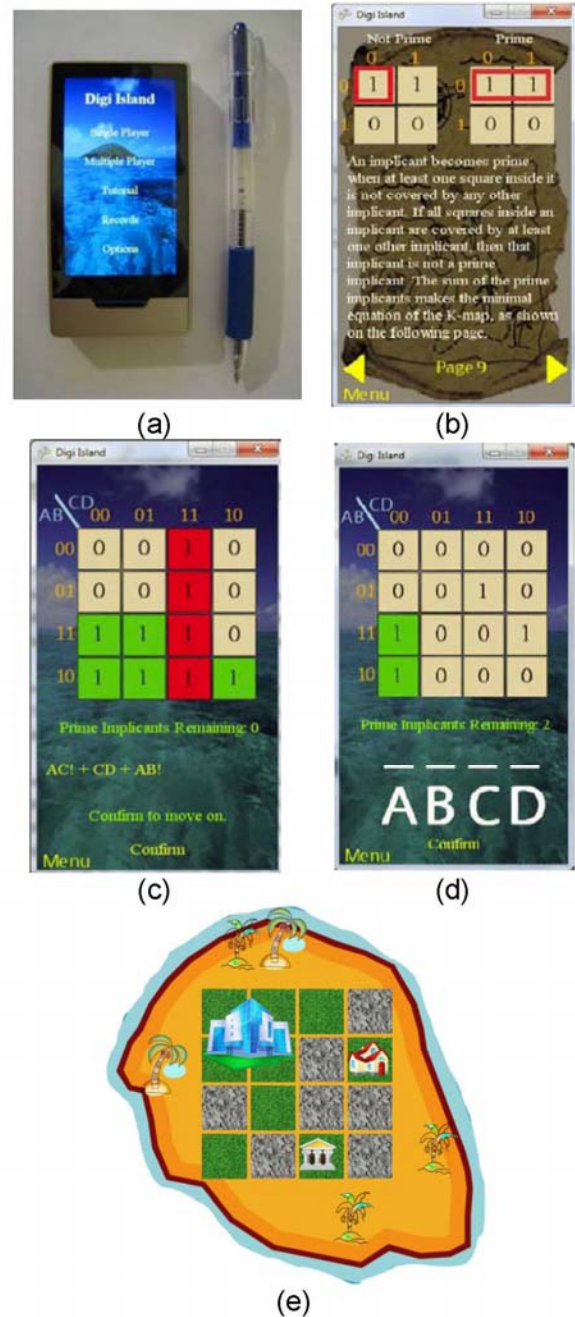
Figure 3 Screen captures of Digi Island. (a) Welcome screen. (b) Tutorial. (c) Practice mode with direct implicant selection. (d) Practice mode with equation input. (e) Play mode.

The game has been deployed on the Zune HD media player, Figure 3(a); it has a multimedia tutorial, Figure 3(b); the player

873

can select different input mode (direct implicant selection and equation input) to enter the answer, Figure 3(c); the Digi Island map is shown in Figure 3(d).

3.0 DISCUSSION

The limited prototype of the game already demonstrated many advantages of the game. For example, the typical K-maps in the textbooks contain many overlapping rectangles representing different primary implicants, making them visually confusing and difficult to understand. In the Digi Island game, each primary implicant can be selected and highlighted individually through user interactions, leading to a much clearer representation and better understandings. Advanced rendering techniques, such as transparency control using alpha maps, can be used to align two K-maps to facilitate groups of minterms, which is not possible using just plain textbooks or paper and pencil methods.

One important principle of learning is to connect new concepts and understanding to pre-existing knowledge [17]. The *Digi Island* game elegantly converts boring 1s and 0s in digital circuits into usable and unusable spaces on a beautiful island and transforms K-map optimization into real estate development, an activity with which many students are familiar and also interested in. The rules for K-map optimization exhibit themselves as construction regulations for real estate development. Players will be more engaged when they deal with real assets such as skyscrapers and amusement parks rather than blocks of abstract 1s and 0s.

Currently, the K-maps are randomly generated by the game and the player has no control of the K-map generation. In the near future, we will add another mode in which the player can generate his/her own K-maps. However, this may have some unintended effects, e.g., the player can use this game to solve homework problems. An area of future development is to include "don't cares" in the K-maps, which are the outputs of certain inputs that do not matter. They "don't cares" can be treated as either 1 or 0. Usually including the "don't cares" in the K-map optimization can further simplify the circuit.

The game can be further expanded to have multiplayer mode to form more competitive and engaging game plays. Players can face off against each other to see who can solve a K-map the fastest. With networking, a game server can also be setup to store player configuration and performance. The game can be easily ported to smartphones using Windows Phone 7 with very minimal effort.

4.0 CONCLUSION

K-maps are an important tool for teaching and learning digital circuit optimization and simplification, which is critical to reduce physical size and power consumption of electronic devices and prolong their battery life. This paper discussed the design and development of Digi Island, a serious game for teaching and learning K-map optimization. The game was developed as a product of a senior design project at Old Dominion University. The Digi Island is a fun and engaging game that offers many advantages over traditional teaching methods of K-maps and it will be further expanded and enhanced in the near future.

5.0 REFERENCES

[1]   Entertainment Software Association, "Industry Facts," *http://www.theesa.com/facts/index.asp,* 2010.

[2]   NPD Group, "2009 U.S. VIDEO GAME INDUSTRY AND PC GAME SOFTWARE RETAIL SALES REACH $20.2 BILLION," *http://www.npd.com/press/releases/press_100114.html,* 2010.

[3]   The National Academies, "Improving Undergraduate Instruction in Science, Technology, Engineering and Mathematics," 2003.

[4]   Project Kaleidoscope, "Recommendations for Action in

Support of Undergraduate Science, Technology, Engineering, and Mathematics," 2002.

[5] American Association for the Advancement of Science, "Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering and Mathematics Education," 2005.

[6] NSF Task Force on Cyberlearning, "Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge," National Science Foundation2008.

[7] V. P. Nelson, H. T. Nagle, J. D. Irwin, and B. D. Carroll, *Digital Logic Circuit Analysis & Design*: Prentice Hall, 1995.

[8] M. M. Mano and C. R. Kime, *Logic and Computer Design Fundamentals*. Upper Saddle River, NJ: Pearson Prentice Hall, 2008.

[9] T. Fullerton, *Game Design Workshop: A Playcentric Approach to Creating Innovative Games*: Morgan Kaufmann, 2008.

[10] J. Schell, *The Art of Game Design: A Book of Lenses*: Morgan Kaufman, 2008.

[11] R. K. Brayton, A. L. Sangiovanni-Vincentelli, C. T. McMullen, and G. D. Hachtel, *Logic Minimization Algorithms for VLSI Synthesis*. Norwell, MA: Kluwer Academic Publishers, 1984.

[12] R. L. Rudell, "Multiple-Valued Logic Minimization for PLA Synthesis," *Memorandum No. UCB/ERL M86-65,* 1986.

[13] University of California at Berkeley, "Espresso," *http://embedded.eecs.berkeley.edu/pubs/downloads/espresso/index.htm*, 2010.

[14] W. V. Quine, "The Problem of Simplifying Truth Functions," *American Math Monthly,* vol. 59, pp. 521-531, 1956.

[15] E. J. McCluskey, "Minimization of Boolean Functions," *Bell System Technical Journal,* vol. November, pp. 1417-1444, 1956.

[16] Microsoft, *XNA Game Studio Documentation*, 2009.

[17] Board on Behavioral, Cognitive, and Sensory Sciences and Education, *How Students Learn: History, Mathematics, and Science in the Classroom*. Washington, DC: The National Adademies Press, 2005.

MODSIM WORLD
Conference & Expo

Digi Island: A Serious Game for Teaching and Learning Digital Circuit Optimization

Michael Harper, Joseph Miller and Yuzhong Shen

Department of Modeling, Simulation and Visualization Engineering

Old Dominion University

mharp011@odu.edu, jmill103@odu.edu, yshen@odu.edu

9/13/2010

1

MODSIM WORLD
Conference & Expo

Outline

- Introduction
- Digital Circuit Optimization Using K-maps
- Game Design
- Game Development
  - Logic Circuit Optimization
  - Game Play
- Results
- Discussion and Conclusion

9/13/2010

2

Digital Circuits and Optimization

- Digital circuits are present in almost all electronics equipment and devices today
  - Computers, MP3 players, digital cameras, etc.
- Optimization of digital circuits reduces complexity.
  - Less weight, longer battery life

Necessity of Optimization

- Same logic function can have multiple solutions of varying complexities.
- Necessary to find the solution that has the least complexity.
- Reduction in complexity allows for reduction in weight and size of circuits, and increase in battery life.
- Process to find the least complex circuit is known as digital circuit optimization.

Methods of Optimization

- Various methods of optimization have been developed over the years.
  - Boolean algebraic manipulation and minimization
  - Karnaugh maps
  - Quine-McCluskey algorithm
  - Espresso algorithm
  - Petrick's algorithm
- Karnaugh Map is the standard method for teaching digital circuit optimization.

Karnaugh Map Overview

- Karnaugh maps (K-maps) graphically represent logic circuits intuitively and systematically.
- Sides of a K-map represent circuit inputs.
- Each cell of a K-map represents circuit output with values of 1 or 0.

878
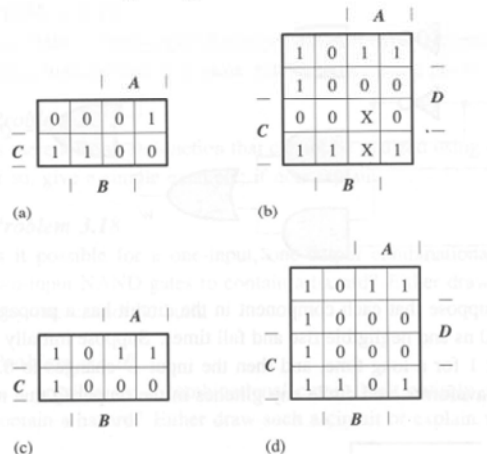
K-map Optimization (1/2)

- K-map optimization is the process of finding a minimal number of maximal aggregations of K-map cells.
  - Cells must have a value of 1.
  - Grouped according to a set of rules
- Key concepts to understanding K-maps
  - Implicant
  - Prime Implicant
  - Essential Prime Implicant

K-map Optimization (2/2)

- To find the optimized expression, identify all prime implicants.
- Optimized expression is the sum of essential prime implicants and prime implicants not contained in essential prime implicants.
  - Remaining prime implicants determined by selection rules
  - Selection rules minimize implicant overlap.

Digi Island Overview

- Passion for games can be exploited for educational purposes via "serious games."
- Goal of the game is to formally introduce K-maps in an engaging manner.
- Designed as a construction based strategy game.
- Digi Island has three modes
  - Tutorial Mode
  - Practice Mode
  - Play Mode

Game Requirements

- Must be playable on personal computers and the Microsoft Zune HD.
- On the Zune, the touch screen is the input device.
- Sound effects and voice instructions must be included.

Tutorial Mode

- Provides several tutorials by providing circuits with 2, 3 or 4 variables.
- Identifies implicants, prime implicants and essential implicants.
- Illustrates procedure for selecting prime implicants and nonessential implicants.
- Generates final Boolean expression for the circuit.

Practice Mode

- Displays K-maps with 2, 3 or 4 variables.
- Asks the player to identify implicants, prime implicants and essential prime implicants.
- Player needs to find all implicants directly.
- Player must also generate the optimized Boolean expression.

Play Mode

- Play mode is where Digi Island evolves from a teaching tool to a fun, engaging game.
- Player is an adventurer to Digi Island, and challenged to develop the island into a tourist attraction by developing real estates.
- Player cannot utilize the entire island.
  - Rocky terrain
  - Wildlife preserves
  - Regulations from the local government
- Player must maximize available space with buildings to produce the greatest income.

Game Components

- Game development contained front end and back end components.
- Front end component deals with graphical user interface.
  - Menus, tutorials, K-maps
  - Different user input modes and user feedback
- Back end component contains major logic for digital circuit optimization.
  - Generates optimized Boolean expression(s)
  - Compares possible solutions to player's answer

Logic Circuit Optimization

- To effectively teach students about K-maps, the game must know the final optimized expression.
- K-maps work well visually, but do not lend well to implementation on a computer.
- Several algorithms have been developed for logic circuit optimization.
  – Quine-McCluskey algorithm
  – Espresso algorithm

Espresso Algorithm

- Heuristic algorithm that is highly efficient
- Doesn't guarantee optimized circuit, but close.
- Available from University of California at Berkeley.
- Not feasible, as implementation options either too time consuming or incompatible with Zune HD.

Quine-McCluskey Algorithm

- Uses pattern recognition similar to human brain.
- Guarantees optimal solution, but inefficient.
- Team developed code for the algorithm from scratch.
  - Microsoft .NET Framework
  - C# programming language

Programming Tools

- Game play developed using Microsoft's XNA Game Studio, which supports cross platform development and includes
  - Tools and templates for rapid game development
  - Optimized libraries based on Microsoft .NET Framework
- XNA Game Studio provides "Game" class to handle game logic updates and drawing.
- Utilized object oriented programming to develop a number of classes.

Class Diagram
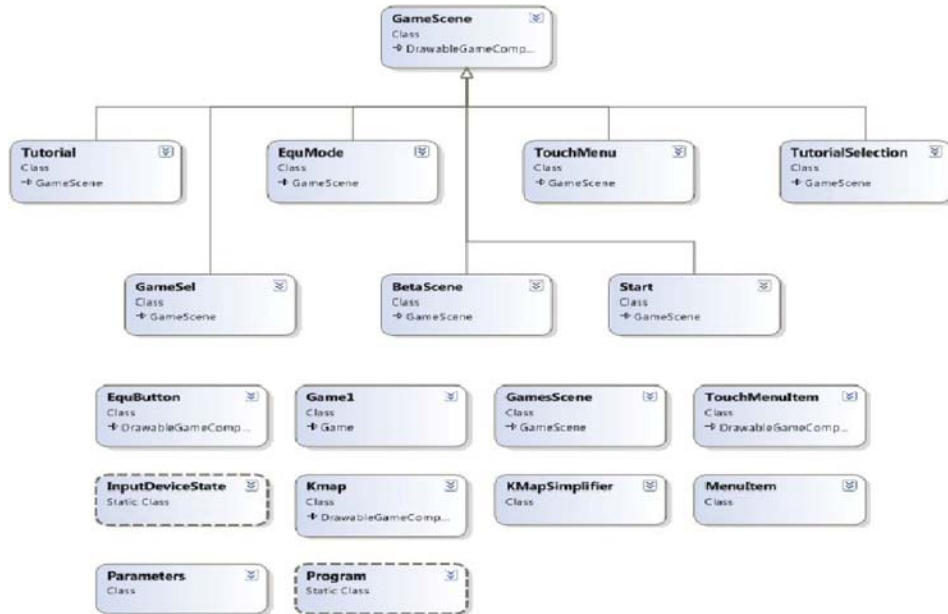
Preliminary Results

- Prototype deployed on the PC and Zune HD.
- Tutorial and practice modes completed, with minor aesthetic changes possible.
- Quine-McCluskey algorithm developed from scratch.
- Play mode yet to be implemented.

Zune Implementation

Tutorial Mode

886

Practice Mode

Advantages of Digi Island

- Textbook examples of K-maps are visually confusing, which renders them difficult to understand.
- Digi Island allows users to highlight specific answers, leading to clearer understanding.
- Digi Island converts K-maps into an island that players transform to earn points, which engages learning.

Future Developments

- Include "Don't Cares", or outputs that do not matter and can be either 0 or 1.
- Multiplayer mode
- Port to smartphones with Windows Phone 7.

Conclusion

- K-maps are important for teaching digital circuit optimization.
- Digital circuit optimization is necessary to reduce size and increase battery life of circuits.
- Digi Island is a fun, engaging game that offers many advantages over traditional teaching.
- Digi Island will be further enhanced and expanded.

Questions?

Thank you!

6.5 Longitudinal Study: Efficacy of Online Technology Tools for Instructional Use

Michael D. Uenking
Thomas Nelson Community College
uenkingm@tncc.edu

Studies show that the student population (secondary and post secondary) is becoming increasingly more technologically savvy. Use of the Internet, computers, MP3 players, and other technologies along with online gaming has increased tremendously amongst this population such that it is creating an apparent paradigm shift in the learning modalities of these students. Instructors and facilitators of learning can no longer rely solely on traditional lecture-based lesson formats. In order to achieve student academic success and satisfaction and to increase student retention, instructors must embrace various technology tools that are available and employ them in their lessons. A longitudinal study (January 2009-June 2010) has been performed that encompasses the use of several technology tools in an instructional setting. The study provides further evidence that students not only like the tools that are being used, but prefer that these tools be used to help supplement and enhance instruction.

1.0 INTRODUCTION

Technology is becoming more prevalent in our society and our regular day-to-day activities. With online chat, video, and other tools, we are bridging the gap between peoples of different cultures, ethnic backgrounds, and languages without having to spend thousands of dollars to travel to foreign countries. With tight budget constraints, businesses are choosing to conduct meetings using these and other tools instead of sending their employees overseas to have face-to-face conferences.

Even children of today are able to fight a "virtual" war or play "virtual" sports using an online gaming system (i.e. XBox, Playstation, Wii, etc.) and team up with people all across the world to accomplish their various tasks and missions. But as time progresses, as these children are growing up in their respective countries using these and other computer technologies, they are becoming increasingly more technologically savvy. But is this possibly creating a paradigm shift in the way these children are learning? Are educators encountering difficulties teaching children of the 21st century especially if they do not use technology in their classrooms? Could this possibly be contributing to some of the behavioral problems that educators are facing? The answers to all of these questions are not entirely clear. But what is clear is that students enjoy having technology as part of the learning experience and educators also find that technology provides them with a rewarding experience as well. In the very least, according to Reference [17], learning in an online environment has been overwhelmingly proved to be just as effective as that in traditional classrooms (Tallent-Runnels et al., 2006, Spring, p. 116).

The following report encapsulates a longitudinal study that occurred from January 2009 to June 2010 in which an online tool (Adobe® Captivate®) was used to conduct a mechanical engineering technology lesson. Quantitative data was collected from the students and qualitative data was collected from fellow instructors during this timeframe. The next section will provide the body of this report.

2.0 BODY

The first main section is a literature review that imparts the background for this study. This section will be broken down into the following sub-sections: effects of online gaming, characteristics of an online student, advantages of online learning, and issues that exist with online learning tools. The second section will discuss the method used in the study. The third section will

provide a brief description of the participants used during the study. The fourth and fifth sections will introduce the reader to the quantitative analysis and results, respectively. Finally, the sixth section will provide some qualitative comments provided by the instructors who were given a chance to review the modules.

2.1 Literature Review

2.1.1 <u>Effects of Online Gaming</u>

As mentioned previously, there are children across the world who are engaged in various forms of online gaming. They use various forms of gaming devices to include their personal computer and/or some or other commercially available gaming console such as XBox Live, Playstation, Wii, or others. Whenever a child engages in these forms of online gaming systems, it is obvious that learning is also occurring. Not only do the children have to learn how to use the system, but also embedded within the individual games are certain techniques, skills, and strategies that must also be learned in order for the student to become proficient in the game and be more competitive with and against other players who are in the system. So, if learning is truly occurring in the gaming world, then how is that being translated to the real world; more specifically, how is it being translated in the educational environment of these students?

Reference [15] provides a detailed study that addresses this very question. This study provided the following results, in that online gaming:

- provides learners multiple avenues of support and communication;
- provides learners opportunities to access vital information via social networks and construct knowledge as the result of social collaboration;
- promotes deliberate, functional epistemology toward the acquisition of knowledge and the development of performance;

- affords various degrees and types of interactivity, each supporting the development of expertise in unique and interesting ways;
- and provides a structured context intended to promote the necessary skills to accomplish complex, goal-based tasks (Schrader & McCreery, 2008, December, pp. 570-571).

The study also states the individual learners "are empowered through a dynamic, interconnected process that scaffolds both technological skills sets and content knowledge [which, in turn,] provide substantial support and developmental tools for focused goal oriented learning at all levels of expertise" (p. 571). So, it is easy to see how students of today are using technology to not only provide cognitive engagement, but how that same technology also enhances their higher order thinking skills. What characteristics are then commonplace in these types of students who are now becoming online learners and are engaging in online learning environments?

2.1.2 <u>Characteristics of an online student</u>

References [2], [3], [11], and [17] all agree that a successful online learner is one that is already proficient in the basic use of a computer and has either prior online learning experience or is fluent/proficient with using the Internet and various online tools (Cramer, Cramer, Fisher, & Fink, 2008, p. 35, December; Dabbagh and Bannan-Ritland, 2005, p. 39; Menchaca & Bekele, 2008, pp. 246-249; Tallent-Runnels et al., 2006, Spring, p. 116). But reference [3] provides an even more detailed description of the ideal online learner:

- Exhibiting a need for affiliation
- Understanding and valuing interaction and collaborative learning
- Possessing an internal locus of control
- Having a strong academic self-concept

891

- Having experience in self-directed learning or the initiative to acquire such skills (Dabbagh and Bannan-Ritland, 2005, p. 39)

With these skills being applied in the online environment, there are definitely some advantages of learning online.

2.1.3 Advantages of Learning Online

There are many advantages to learning online which are as follows:

- References [4], [8], [9], [10], [11], [13], [14], and [19] all show that online learning contributes to not only higher achievement rates, but also to higher satisfaction levels, and higher levels of engagement (D'Arcy, Eastburn, & Bruce, 2009, Winter, p. 62; Jackson et al., 2006, May, p. 433; Krentler & Willis-Flurry, 2005, July/August, p. 319; Lim, Kim, Chen, & Ryder, 2008, June, p. 119; Menchaca & Bekele, 2008, pp. 246-249; Rogers and Cox, 2008, January/February, p. 38; Saadé & Kira, 2004, Winter, p. 362; Wang & Reeves, 2007, p. 190).
- Reference [2] states that students may feel "more connected, more challenged, and more engaged in learning than ever before…self-confidence [can also be developed as well]" (Cramer, Cramer, Fisher, & Fink, 2008, December, p. 35).
- From a more learning theory-based approach, online learning also, according to Reference [7], helps to support the "constructivist learning" modality "which encourage, and are focused on, users creating, or constructing, their own content" (Hsu, 2007, p. 71). These tools also "emphasize student interaction, group learning, and collaboration, rather than the more traditional classroom mode…[especially] where the emphasis is on student communication, where students have access to technology, and where creative output and thinking

is encouraged" (p. 85). Reference [5] also points out the need for this "constructivist learning" environment to be more "learner-centered" as well (Hannum, Irvin, Lei, & Farmer, 2008, November, p. 223).
- References [6], [11], and [20] also address the fact that learning online provides more flexibility of where and when the learning will occur (i.e. home, work, vacation, or on travel for business) and more specifically, for rural areas where a traditional instructor is hard to acquire (Hannum, Irvin, Banks, & Farmer, 2009, pp. 13-14; Menchaca & Bekele, 2008, pp. 246-249; Zhao, Alexander, Perreault, Waldman, & Truell, 2009, pp. 210-211).
- References [1], [4], and [14] address the efficacy of using online quizzes in that not only do they provide repetition, but also instant feedback to the students and, in turn, they better prepare the students for unit exams. Faculty and students are also able to focus on discussion and hands-on activities (Bartini, 2008, p. 10; D'Arcy, Eastburn, & Bruce, 2009, Winter, p. 57; Saadé & Kira, 2004, Winter, p. 361).

2.1.4 Issues with Learning Online

References [11], [12], and [17] promote the notion that it is important for students to have some sense of community whether it is a face-to-face contact session or some means to make connections with the faculty and their peers. This, in turn, helps to enhance the learning process (Menchaca & Bekele, 2008, pp. 246-249; Nicholas & Ng, 2009, p. 323; Tallent-Runnels et al., 2006, Spring, p. 116). The main issue that many students have is in how the online course is formatted and designed; so it is important, according to References [10], [11], [16], and [19], that instructors provide means for practice, feedback, and improvement for the course, that technical issues are directly addressed, and that they ensure that the

online tools that are being used are updated and current (Lim, Kim, Chen, & Ryder, 2008, June, p. 119; Menchaca & Bekele, 2008, p. 249; Sitzmann, Kraiger, Stewart, & Wisher, 2006, Autumn, p. 654; Wang & Reeves, 2007, pp. 186-190).

One additional issue that Reference [18] mentions is that there exist "differences in perception about online learning…between faculty students…which may be due to the heterogeneous points of view and motivations for online learning between faculty and students" (Tanner, Noser, & Totaro, 2009, p. 36). The next section will now go into detail about the study that was performed.

2.2 Method

In the Spring 2009 semester, a three-part mechanical engineering technology online module was developed that addressed the three basic methods of truss analysis (i.e. method of joints, method of sections, and method of members). The students were given the module in lieu of regular class meetings over a ten-day period. They were not allowed to obtain any assistance from the instructor during this timeframe nor were they allowed to elicit support from their classmates. The only tools that they were allowed to use during this timeframe included their textbook, a calculator, writing utensils, the module, and their respective computers with Internet access. All student participants were given a pre-test that was comprised of five truss analysis problems and were instructed to not prepare for it prior to the exam. This pre-test provided a baseline assessment score that was compared to a final assessment score in the final analysis component. These two assessment scores, Likert Scale values, and demographical data were the primary forms of quantitative data. A second session was attempted in the Fall 2009 semester, but various college Internet technical issues prevented the students from completing the module, so their data was removed from consideration for the longitudinal study. But the module was also presented to other

instructors via email transfer during the Spring 2010 semester and was also presented at the 2010 Virginia Community College System (VCCS) New Horizons conference. Qualitative data was collected from all of the instructors.

The software program that was used to develop the module was Adobe® Captivate® which enables the instructor to incorporate animation (text and graphic), PowerPoint slides, user-input text fields, instant feedback quiz generation (which can also send the results to the user's email address), music, and recorded voice. The program also allows the instructor to create multiple formats that can be incorporated into various media outlets (i.e. Flash video, HTML, and a standalone executable). A snapshot of the user interface is shown below in Figure 1:

Michael D. Uenking
Associate Professor in
Mechanical Engineering Technology
Thomas Nelson Community College
Spring Semester 2009

Figure 1. Snapshot of Adobe Captivate User Interface

2.3 Participants

A total of ten students (comprised of three females and seven males whose average age was approximately 25) that were enrolled in MEC131 (Applied Statics in Engineering Technology) participated in the study. There were also five instructors from different colleges across the United States that provided qualitative feedback via email. The last group was comprised of six additional instructors in the VCCS who were given a demonstration of the module, were provided results from the Fall 2009 student

data, and were given an opportunity to interact with the module. Qualitative data was collected for this last group as well.

2.4 Student Quantitative Analysis
The student analysis is provided below:

Hypothesis: The average score of the post-test will be higher than the average score of the pre-test.

Null hypothesis: There is no difference in the average scores of the pre-test and post-test.

Test used: one-tailed t-test

Value for alpha: p = .05

Table 1 below provides the data that was collected for the pre-test and post-test data values:

Table 1. Student Pre-Test and Post-Test Data Values

Count:	10	10
Averages:	32.50	55.00
Median:	40.00	58.75
Std Deviation:	25.658007	30.29943
Variance:	658.33333	918.0556
Var/(number-1):	73.148148	102.0062
Total:	175.15432	
Square Root:	13.234588	
estimated t-value:	1.7000907	
Degrees of Freedom:	18	

2.5 Student Quantitative Results
Based on the values shown in Table 1, since the estimated t-value approaches the table value of 1.734, it can be said that the average test scores increased somewhat significantly from the pre-test to the post-test based on the module intervention (t=1.70,dof=18,p<.05); therefore, the

students did improve statistically in their test scores due to the module intervention.

A post-test questionnaire was also provided that utilized a Likert Scale format. The significant results are provided below:

Question: What is your overall feeling about the STAMINA modules that you participated in for Chapter 5?

Answer: 70% of the students liked the modules.

Question: Did you like or dislike the addition of music to the presentation?

Answer: 70% of the students liked the addition of music.

Question: Would you consider these modules to be excellent tools as a SUPPLEMENT to your regular classroom time; that is, would these tools be considered a great addition to your regular class?

Answer: 90% of the students would consider these modules as an excellent supplement to the regular course.

Question: If these modules were given as a SUPPLEMENT to my MEC131 course, I would use them to enhance my learning.

Answer: 80% of the students would use the modules to enhance their learning of the content material.

Question: In your opinion, could you use these modules as a STANDALONE learning tool; that is, could these modules be used instead of having a regular classroom environment?

Answer: 100% of the students disagreed that this module can be used as a standalone learning tool.

Question: Was the user interface (Adobe Captivate) in your browser easy to load and navigate?

Answer: 100% of the students agreed.

2.6 Instructor Qualitative Results
Several comments were provided by the instructors in both the email group and the face-to-face group that provide excellent feedback for this study. Some samples of the comments are provided below:

"I commend you for developing the modules listed below and I expect it took you quite some time to complete them. I would be interested to learn what your plans are for the modules going forward. I would also be interested to know what textbook you are currently using for this course."

"I REALLY liked your first presentation!..I'm going to have to learn how to create one like it."

"Great work here! I like the interactive and interesting visual and experiential components. I am developing any helps for my statics classes to make it more interactive."

"Overall you have done well with the presentations. The main benefit to students will be with the ability to review the demos more than once."

"I'm not sure how well the pre-assessment part would work, but the other stuff would be good for students who might want some passive learning experiences."

"Really like the graphics and colors and user interface…"

"Excellent: warm-ups, graphics, music…"

3.0 DISCUSSION
This study was limited in that the study was only administered to one group of students (95% confidence level, confidence interval +/- 30.98). Also, the group was not supervised to ensure that all three modules were fully viewed by each of the students even though the number of times each

student accessed the modules was catalogued electronically. It was hoped that the second group of students in the Fall 2009 semester would have provided an additional set of data to increase the validity of this study, but because of the Internet connectivity problems, this was not possible. Due to other commitments, course scheduling, and time constraints, the study was not able to be administered to other groups of students during the Spring 2010 semester.

Also, an issue that is encountered by engineering technology students is that this particular discipline has been traditionally taught in a lecture format only. Introducing technology into these types of courses creates a sort of paradigm shift in that not only do the students have to learn how to use this technology as an integral part of their learning process, but engineering technology instructors will also need to learn how to incorporate different forms of technology into their curriculum to help make their courses more robust.

4.0 CONCLUSION
What can be concluded from this study is that students not only liked the technology that was used, but prefer to have some form of technology to supplement their learning experience. This agrees with the literature review that was previously provided. Test scores did in fact improve significantly, so it is possible that if an instructor wanted to use the module as a standalone tool for implementation in a distance or hybrid version of the course, then there might be some usefulness in doing so (even though the students who participated in the study were against using it in this fashion). It still may be in the best interest of the instructor and the students to use tools like this to primarily further reinforce concepts taught in the classroom. Giving students the ability to review the video an unlimited amount of times gives them further practice in understanding the concepts that are provided which may, in turn, help better prepare them for unit exams. This definitely

frees up the instructor from having to use additional time in class to go over the same concepts and puts the onus on the student to take more responsibility of their own learning experience (i.e. being more learner-centered).

Fellow instructors also liked how the module was designed and seemed encouraged to want to try implementing some form of online tool in their respective courses. These instructors provided helpful feedback that will be used to revise and/or modify the modules should they be implemented again for future sections of the course.

5.0 REFERENCES

[1]. Bartini, M. (2008, March). An Empirical Comparison of Traditional and Web-enhanced Classrooms. Journal of *Instructional Psychology, 35*(1), 3-11. Retrieved July 28, 2010, from ProQuest Psychology Journals database. (Document ID: 1467176201).

[2]. Cramer, S., Cramer, S., Fisher, D., & Fink, L. (2008, December). Online or Face-to-Face? Which Class to Take. *Voices From the Middle, 16*(2), 25-36. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 1618382801).

[3]. Dabbagh, N. & Bannan-Ritland, B. (2005). *Online learning: Concepts, strategies, and application.* Upper Saddle River, NJ: Prentice Hall, Inc.

[4]. D'Arcy, C.J., Eastburn, D. M., Bruce, B. C. (2009, Winter). How Media Ecologies Can Address Diverse Student Needs. *College Teaching, 57*(1), 56-63. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 1622306051).

[5]. Hannum, W.H., Irvin, M.J., Lei, P., & Farmer, T.W. (2008, November). Effectiveness of using learner-centered principles on student retention in distance education courses in rural schools. *Distance Education, 29*(3), 211-229. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 1618071811).

[6]. Hannum, W.H., Irvin, M.J., Banks, J.B., & Farmer, T.W. (2009). Distance Education Use in Rural Schools. *Journal of Research in Rural Education (Online), 24*, 1-15. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 1648626241).

[7]. Hsu, J. (2007). Innovative Technologies for Education and Learning: Education and Knowledge-Oriented Applications of Blogs, Wikis, Podcasts, and More. *International Journal of Information and Communication Technology Education, 3*(3), 70-89. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 1325521271).

[8]. Jackson, L., von Eye, A., Biocca, F., Barbatsis, G., Zhao, Y., & Fitzgerald, H. (2006, May). Does home internet use influence the academic performance of low-income children? *Developmental Psychology, 42*(3), 429-435. Retrieved July 28, 2010, from PsycARTICLES database.

[9]. Krentler, K.A. & Willis-Flurry, L.A. (2005, July/August). Does Technology Enhance Actual Student Learning? The Case of Online Discussion Boards. *Journal of Education for Business, 80*(6), 316-321. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 900367601).

[10]. Lim, J., Kim, M., Chen, S.S., & Ryder, C.E. (2008, June). An Empirical Investigation of Student Achievement and Satisfaction in Different Learning Environments. *Journal of Instructional Psychology, 35*(2), 113-119. Retrieved July 28, 2010, from ProQuest Psychology Journals database. (Document ID: 1542878161).

[11]. Menchaca, M.P. & Bekele, T.A. (2008). Learner and instructor identified success factors in distance education. *Distance Education, 29*(3), 231-252. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 1618071831).

[12]. Nicholas, H. & Ng, W. (2009). Engaging Secondary School Students in Extended and Open Learning Supported by Online Technologies. *Journal of Research on Technology in Education, 41*(3), 305-328. Retrieved July 28, 2010, from ProQuest

Education Journals database. (Document ID: 1660271741).

[13]. Rogers, J.W. & Cox. J.R. (2008, Jan/Feb). Integrating a Single Tablet PC in Chemistry, Engineering, and Physics Courses. *Journal of College Science Teaching, 37*(3), 34-39.; Retrieved July 28, 2010, from ProQuest Education Journals database.

[14]. Saadé, R.G. & Kira, D. (2004, Winter). Effectiveness of an Interactive Application to Assist Learning: A Test Case. *Journal of Information Systems Education, 15*(4), 357-363. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 793505791).

[15]. Schrader, P.G. & McCreery, M. (2008, December). The acquisition of skill and expertise in massively multiplayer online games. *Educational Technology, Research and Development, 56*(5/6), 557-574. Retrieved July 28, 2010, from ProQuest Psychology Journals database. (Document ID: 1632036161).

[16]. Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006, Autumn). The Comparative Effectiveness Of Web-Based And Classroom Instruction: A Meta-Analysis. *Personnel Psychology, 59*(3), 623-664. Retrieved July 28, 2010, from ProQuest Psychology Journals database. (Document ID: 1138016761).

[17]. Tallent-Runnels, M. K., Thomas, J. A. Lan, W.Y., Cooper, S., et al. (2006, Spring). Teaching Courses Online: A Review of the Research. *Review of Educational Research, 76*(1), 93-125. Retrieved July 28, 2010, from ProQuest Psychology Journals database. (Document ID: 1132445571).

[18]. Tanner, J.R., Noser, T.C., & Totaro, M. W. (2009). Business Faculty and Undergraduate Students' Perceptions of Online Learning: A Comparative Study. *Journal of Information Systems Education, 20*(1), 29-40. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 1666755581).

[19]. Wang, S. & Reeves, T.C. (2007). The effects of a web-based learning environment on student motivation in a high school earth science course. *Educational Technology, Research and Development, 55*(2), 169-192. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 1289280211).

[20]. Zhao, J.J., Alexander, M.W., Perreault, H., Waldman, L., & Truell, A.D. (2009). Faculty and Student Use of Technologies, User Productivity, and User Preference in Distance Education. *Journal of Education for Business, 84*(4), 206-212. Retrieved July 28, 2010, from ProQuest Education Journals database. (Document ID: 1653117671).

6.0 ACKNOWLEDGMENT(S)

7.0 SERIOUS GAMES & VIRTUAL WORLDS TRACK

7.1 Leveraging Gaming Technology to Deliver Effective Training

Leveraging Gaming Technology to Deliver Effective Training

James D. Cimino
D2 TEAM-Sim
jcimino@d2teamsim.com

The best way to engage a soldier is to present them with training content consistent with their learning preference. Blended Interactive Multimedia Instruction (IMI) can be used to teach soldiers what they need to do, how to do each step, and utilize a COTS game engine to actually practices the skills learned. Blended IMI provides an enjoyable experience for the soldier, thereby increasing retention rates and motivation while decreasing the time to subject mastery. And now mobile devices have emerged as an exciting new platform, literally placing the training into the soldier's hands. In this paper, we will discuss how we leveraged commercial game engine technology, tightly integrated with the Blended IMI, to train soldiers on both laptops and mobile devices. We will provide a recent case study of how this training is being utilized, benefits and student/instructor feedback.

1.0 INTRODUCTION

Motivating soldiers to want to learn can be a difficult proposition. Even today, the majority of training is delivered by a classroom-based, instructor-led curriculum. Classroom-based courses typically result in one dimensional training, offering minimal opportunities to engage the soldier [6]. Even for simple training, one-dimensional curriculums usually consist of a long list of PowerPoint slides. Soldiers have a name for this type of training, "Death by PowerPoint." Blended Interactive Multimedia Instruction (IMI) solutions, incorporating different types of media and various levels of interaction, provide an engaging alternative to traditional classroom training. Blended IMI solutions have the ability to motivate soldiers who require repetitive training for collective, individual, and team performance tasks. This helps entice soldiers to learn tasks which could otherwise be considered tedious or boring [6]. Similarly, deployed soldiers can be engaged to remain proficient on individual and crew oriented tasks when supplied with engrossing tactical training applications on a handheld device. Handheld applications have the potential to be a powerful contributor in the process of ensuring that soldiers retain fundamental skills necessary for successful combat operations.

2.0 OPPORTUNITY

Have you ever watched a teenager with a new video game? They do not read any instructions; they simply load the disk into their game console and start to play. Youths learn without instructors. They learn through experience and both positive and negative feedback and consequences. Gaming technologies have the added advantage of letting the "player" be in control. First-person shooter games have been around for decades, and this new generation of soldier, "Generation X-Box," wants to participate in their training. They are more comfortable with a video game than sitting in a classroom viewing a PowerPoint slide deck or reading a technical manual.

With upwards of 70,000 new soldiers enlisting in the Army each year [9], it is imperative that every opportunity be taken to maximize access to training [3]. With continuing operational deployments, ready access to individual soldier tasks/collective training (i.e. crew drills) are equally important. Blended IMI allows for improvements in the quality of instruction in addition to increasing the efficiency of creating, deploying, and managing the instruction. With the ability to cover the widest array of material, blended IMI solutions can be completely web-deliverable. Blended IMI can offer the highest quality educational experience to

the greatest number of students. Transferring this training to tactical applications on a soldier's handheld device means better trained soldiers fully prepared for combat.

3.0 PROOF OF THE PROBLEM

Digitized course material (e.g. PDF files, PowerPoint slides, Word Documents) has been in use for decades. However, these are merely digitized versions of the existing course material which adds little to the effectiveness of the instruction. A blended approach utilizing multiple IMI levels and types, extracts value from these materials beyond what traditional classroom instruction can accomplish [5]. The most significant benefits of blended IMI solutions stem from their ability to transform the roles of instructors and students. This transformation allows for a reduction in the amount of class time needed, a decrease in the need for travel, and a positive return of investment for training dollars.

4.0 THE SOLUTION: A BLENDED IMI CURRICULUM

The use of a blended IMI curriculum provides benefits to four key areas of the modern educational process: student interactions, instructor interactions, accessibility and transportability, and return on investment.

4.1 Student Interaction

The best way to reach and engage a soldier is to present them with content consistent with their learning preference. In general, most people can be characterized as learning more effectively from one of three types of presentation: aural, visual, and physical kinesthetic. Auditory learners are more engaged by information presented in an aural format. Visual learners benefit the most from ocular content such as video and

text. Physical kinesthetic learners respond better to information disseminated during physical activity [2]. Blended solutions can provide simultaneous delivery of audio, visual, and physical kinesthetic content delivered on a single platform. An example of this is a video accompanied by voiceover, coupled with a basic game-based simulation exercise designed to allow the student to apply the knowledge transferred from watching the video.

Advances in the technology used to create today's multimedia instruction allow all of these formats to be seamlessly blended together to provide a consistent experience for the student. The result is an efficiently delivered package that has content catering to each style of learning. A significant benefit of blended IMI is the reduction in repetitive material that covers the same topics in different ways in order to engage different learning preferences. This reduces the time needed to train, freeing up time for additional courses or practical application exercises [7].

Blended IMI can provide a more enjoyable experience for the soldier, thereby increasing retention rates and motivation while decreasing the time to subject mastery [6]. Practical exercises can be reproduced through constructive games [10] that allow a warfighter to engage in training without fear of failure or poor performance. Without any performance anxiety the soldier is better able to utilize the training for the acquisition of knowledge and skills. These constructive games can be timed or objectively scored to provide feedback for the warfighter and a means of competition through which soldiers will challenge one another. Competition is an extremely effective motivator that is not only free, but entertaining. Training can now become something that soldiers are interested in doing on their own time.

Blended IMI empowers the warfighter to take a more prominent role in their own education by providing them the ability to

perform self remediation. Questions answered incorrectly can be linked to the area in the instructional content that contains the correct answer. This same principle can be applied to activities and constructive games, giving the soldiers the option to return to the instruction to review the procedure again before continuing their activity.

4.2 Instructor Interaction

The student is only half of the learning equation. The experience and insight of instructors cannot be replaced and is a valuable part of the learning process. Instructors can provide not only context to the learning materials, but also impart valuable real world experience. Moving to a blended curriculum transforms the role of the instructor to one in which they can be much more effective.

In the past, instructors had to spend a great deal of time doing little more than providing an audible version of instructional texts. This problem can be addressed by using blended curricula that utilizes a mix of media types. Creative and interesting take home and web-deliverable applications take the place of traditional homework, which usually consisted of reading printed materials. These enhanced pre-course or in-course homework assignments increase the amount of knowledge that students enter class with. Therefore, the instructor can spend their time in class answering questions, discussing advanced concepts, or relaying valuable personal experience [7].

Having students better prepared is not the only benefit to instructors that a blended IMI curriculum can bring. Through the use of IMI, instructors now have the ability to use a dashboard-like display to monitor the progress of all students in real-time. This can be used for instructional intervention when the instructor notices that a student is having issues with a particular concept or step of a process. Instructors can now act to improve critical decision making skills or

respond to infrequent, yet important, questions and scenarios. For example, if the entire class is running a basic constructive game of a certain maintenance procedure, only one of the students may have enacted the exact set of circumstances that would result in a rare safety issue. The instructor could stop the class and bring everyone's attention to that student's scenario to teach directly to this point. An instructor also has the ability to focus on individual remediation without interrupting the rest of the class.

5.0 ACCESSIBILITY AND TRANSPORTABILITY

Full-featured, blended IMI curricula can be created to meet the minimum system requirements for home computers as set forth by TRADOC's Army Training Support Center's Education and Training Support Directorate [1]. Not only will Army school houses be able to run this content, but most students will be able to access and run these courses from their residences or barracks, at home or abroad. The ability to train anywhere, anytime, on almost every computer made available to students is one of the most important benefits to using blended IMI. Being able to include this training as a tactical application on the soldier's handheld device means he can take effective training material with him wherever he is assigned or whatever the mission.

With a blended IMI curriculum, web-deliverable content allows for updates to training materials to be rapidly deployed. The modular nature of most blended training solutions also allows for course designers to upload content. This can provide additional context or relevance to soldiers, further increasing engagement and retention [4]. Required changes to course content can also be made without involving outside developers.

Mobile devices are emerging as an exciting new platform for the delivery of training.

Many warfighters already have a mobile device of some sort and are familiar with installing and using applications developed for their mobile devices. Training created for mobile devices can be used as a supplement to existing PC-based training or for sustainment purposes once a student has completed a course. It is important to make sure that continuity is maintained between mobile and PC based versions. This ensures that students are presented with a consistent interface and presentation style. The importance of maintaining consistency when updating content or moving to new platforms cannot be understated. Consistency between the versions also maintains a minimal learning curve for the student. Any confusion caused by a lack of consistency will be a detriment to the soldier's attitude towards the instruction.

Blended IMI curricula afford the opportunity for soldiers to continue learning and developing skills outside the classroom through the use of interactive leave behind materials. These materials are developed in conjunction with the blended IMI used in class so that the student can apply knowledge, practice skills, and deepen their understanding of key concepts [4]. Mobile versions of these leave behind materials allow students to use them for additional training or reference anywhere they are at any time. By literally placing the training into the soldier's hands enormous gains in the warfighter's initiative to train can be realized.

6.0 RETURN ON INVESTMENT

The replacement of any piece of an IT infrastructure can be expensive, time consuming, and disruptive to normal operations. A blended IMI solution should be designed with existing equipment in mind. The key lies in using the appropriate technologies to create multimedia presentations in order to ensure their compatibility with the widest range of computer hardware possible.

The web-deliverable nature of the content also means that, in cases where appropriate, classroom instruction can be distributed to multiple locations. This results in a decrease in travel and travel-related costs as well as an increase in the number of soldiers that a single instructor can manage in a class. The reduction in travel not only means a direct cost savings but also that soldiers will be required to spend less time away from their families.

The logistical price tag associated with training on any weapons system can be staggering. Examples of the overhead associated with these systems include storage, fuel, maintenance, and transportation costs. Blended IMI and handheld applications lower these costs by reducing to total number of systems required for actual "hands-on" training. Soldiers can practice anytime, anywhere, without needing the actual equipment. They can access instructors/subject matter experts anywhere in the world to get questions answered.

7.0 THE SOLUTION

An effective blended IMI solution encompasses the entire spectrum of student centered instruction from all four levels of IMI to the production of each individual multimedia component. Each course needs to be evaluated for the purposes of isolating individual teaching points, around which student centered interaction can occur. Instructor or SME (subject matter expert) input also needs to be incorporated to enhance the relevancy of developed IMI content. The result is IMI developed with instructors that can function independently or as a supplement to in class training.

Utilizing tools that allow for rapid development, and Commercial off-the-shelf

(COTS) software, the time and cost associated with course creation are reduced significantly. Effective blended IMI solutions should be designed to function on a wide array of hardware, resulting in a dramatic reduction in costs because of the ability to use existing computers and network infrastructure [1]. These solutions should empower Soldiers to access developed courseware from their home PC, laptop or hand-held device.

The next generation of training solutions should leverage commercial game engine technology, tightly integrated with the Blended IMI developed to teach the associated tactical training skills. The Blended IMI portion uses a "crawl – walk – run" methodology to tell the soldier what they need to do, show them how to do each step, and then utilizes the game engine to allow them to perform the skills learned. Blended IMI is used to show the soldier "what right looks like," and how to perform each step. During this portion of the training, learning checks are introduced in the form of multiple choice quiz questions. If a soldier gets a question wrong they have the opportunity to link back to the section of the training to review. All aspects of the training are tracked and recorded. Every step will be date/time coded, and every choice (correct or incorrect) is logged to aid in the After Action Review (AAR) process. With this innovative approach, it's easy to see who completed the training, skipped through sections of the training, or got specific questions wrong. If used for New Equipment Training, an instructor can use this AAR data to assess the class as a whole, or to "zero in" on individuals who need more attention. Soldiers who demonstrate higher levels of proficiency in the Blended IMI training and during game play can be "graduated" out onto the live equipment sooner, while the rest of the class continues to refresh via the Blended IMI.

Figure 2. Screenshot of Handheld Gaming Solution

Figure 1. Screenshot of PC-based Blended IMI Solution

The gaming portion of the training can be timed, and should mirror the performance standard that a soldier is expected to achieve in order to demonstrate proficiency in executing the given tasks. The game can be developed such that multiple soldiers can work together in a cooperative manner, much as they do in a real combat situation. The game portion of the solution should emulate the various roles associated with a given set of tasks, and should allow the soldier to pick their associated role.

As the soldier works through the various levels in the game, he or she should be able to suspend the game play and review the appropriate training video on what is

903

supposed to be done at that specific step. The soldier can then click on a button to return to the game, and pick up where they left off.

In order to improve the accessibility and sustainability of the training, a portable version of the game should be developed. The mobile version of the game should mirror the lock step sequencing of the pc-based game. A mobile version of the pc-based game allows a soldier to take the game with them, refresh their knowledge where ever they go, whenever they want.

8.0 PROOF OF THE SOLUTION: 3-6 ADA

My company, in conjunction with Raytheon, developed a solution for the 3-6 Air Defense Artillery School to create a prototype training application for the March Order and Emplace Crew Drill. The intent of the solution is to supplement training for the 3-6 ADA schoolhouse. Currently, the 3-6 Air Defense Artillery schoolhouse is faced with the challenge of having to train soldiers on various aspects of the Patriot Launch Station platform, but in some cases they have a physical shortage of equipment. Another issue facing the schoolhouse is once they get soldiers on the physical equipment, only two soldiers can participate in a crew drill at any one time, leaving the remaining soldiers to wait and observe their classmates. Finally, the schoolhouse was also seeking an alternative to "death by PowerPoint" for their Advanced Individual Training (AIT).

This blended IMI solution, complete with an interactive game, was developed to address the needs of the 3-6 ADA schoolhouse. The game, called "Launcher Dogs: March Order & Emplace," has been developed for both PC and mobile access. Four classes of 14-Tangos have participated in the initial pilot of the Blended IMI solution and played the associated "game," as have their instructors. This solution has been selected

to be part of the Phase 1 of Connecting Soldiers to Digital Applications (CSDA) pilot. The audience for this pilot will initially be limited to soldiers undergoing 14-T AIT training at Fort Sill. The game is being used as part of the training and sustainment initiatives for the 3-6 ADA schoolhouse. The game is being made available within the schoolhouse barracks, classrooms, and via the Apple iPhone mobile platform.

Utilizing Blended IMI training and the "Launcher Dogs: March Order & Emplace" game, AIT will require less time and resources to teach the Crew Drill, while simultaneously improving learning, training proficiency, interest and long term retention of skills: even in the absence of tactical equipment (Experiential Learning). It is anticipated that through these training efforts, the 3-6 ADA will:

- Decrease the time to learn a crew drill by 50%
- Reduce instructor contact hours by 50%
- Improve training proficiency by 25%
- Decrease caution and safety violations by 25%
- Decrease equipment damage due to new operator fault by 30%
- Significantly increase interest and training motivation
- Significantly improve retention enabling certification at Soldier's first unit
- Decrease maintenance costs from inexperienced use of tactical equipment

During actual tests of the training, soldiers repeated the game in an attempt to "better" their own time in comparison to their battle buddy. The results of this repetition directly translated to an improvement in performance and proficiency when the soldiers transitioned to the live equipment. Figures 3 and 4 depict the effectiveness of the training from both the student's and the instructor's perspectives.
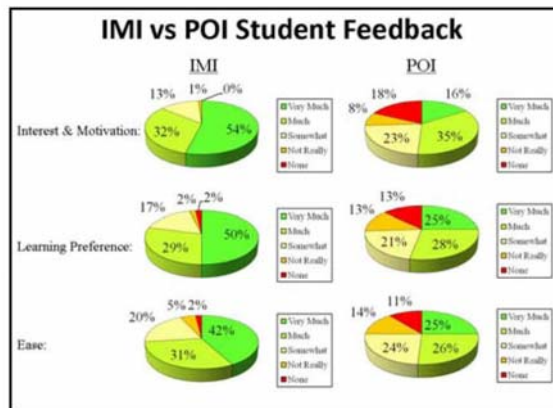
Figure 3. Student Feedback on IMI Effectiveness
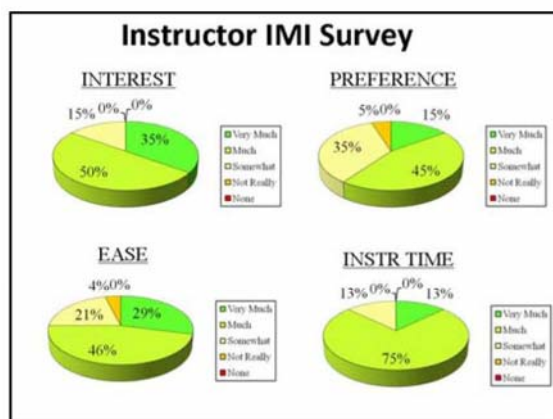
Figure 4. Instructor Feedback on the Effectiveness of Blended IMI

9.0 CONCLUSION

According to Dr. Roger Smith, CTO of PEO STRI, "Time-on-task is an important part of learning. The more time you spend rehearsing, exploring options, and studying outcomes, the better you will become at a skill. Games can add to that by encouraging soldiers to spend more time learning a skill" [9].

Games integrated into blended IMI solutions offer benefits to every facet of tactical training. One such benefit is an increase in training material quality resulting in soldiers

who are more engaged. This increased level of engagement results in higher retention levels and shorter time to subject mastery. Because of IMI's web-deliverable nature, soldiers can continue to gain these benefits while deployed. It is recognized that blended IMI solutions can also be leveraged to ensure all students are prepared before the first day of a course. However, sustainment materials can keep information fresh in a soldier's mind or be used as a reference after training has been completed. Instructors also benefit from the introduction of a blended IMI curriculum. With more prepared students they can focus their time on advanced concepts and relaying real world experience. Solutions can be produced that allow the instructors to update content for increased relevancy and accuracy. Both Games and blended IMI solutions can be designed for existing computer hardware and handheld systems in order to minimize the impact of adoption. All these benefits add up to cost and time savings to the DoD, as well as increased educational quality and accessibility for the warfighter.

10. REFERENCES

1. Army Training Support Center, Education and Training Support Directorate. Baseline Home Computer Configuration for Interactive Multimedia Instruction (IMI) Courseware. http://www.atsc.army.mil/itsd/imi/documents/bhcc_Mar09.htm. 9 March 2009

2. A. K., S. (2008). A conceptual analysis on the approaches to learning. *Kuram ve Uygulamada Eğitim Bilimleri*, 8(3), 707-720. Retrieved from PsycINFO database.

3. Crawley, Jeff. "Gen. Dempsey visits Fort Sill to observe seamless training at schools." www.army.mil. 10 February 2010. http://www.army.mil/-news/2010/02/10/34473-gen-dempsey-

visits-fort-sill-to-observe-seamless-training-at-schools/

4. Lee, J. (2010). Design of Blended Training for Transfer into the Workplace. *British Journal of Educational Technology, 41*(2), 181-198. Retrieved from ERIC database.

5. TRADOC PAM 350-70-2 *Interactive Multimedia Instruction (IMI) Design and Development Guide,* 25 August 2006.

6. Ryan S.J.d. Baker, Sidney K. D'Mello, Ma.Mercedes T. Rodrigo, Arthur C. Graesser, Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments, International Journal of Human-Computer Studies, Volume 68, Issue 4, April 2010, Pages 223-241, ISSN 1071-5819, DOI: 10.1016/j.ijhcs.2009.12.003.

7. Ranky, P. (1996). Interactive multimedia for engineering education. *European Journal of Engineering Education, 21*(3), 273. Retrieved from Academic Search Premier database.

8. US Army Recruiting Command. (2010). *Support Army Recruiting.* Retrieved from http://www.2k.army.mil/faqs.htm

9. Atkinson-Bonasio, Alice (2008). Video Games in Military Training: An Interview with Roger Smith, The Escapist ,August 29, 2008, Retrieved from http://www.modelbenders.com/papers/RSmith_Escapist_080829.pdf

10. Smith, Roger, Page, Ernest H.,(1998). INTRODUCTION TO MILITARY TRAINING SIMULATION: A GUIDE FOR DISCRETE EVENT SIMULATIONISTS, Retrieved from http://www.modelbenders.com/papers/mil4des.html

10. ACKNOWLEDGMENTS

Leveraging Gaming Technology to Deliver Effective Training

James D. Cimino
D2 TEAM-Sim

D2 | TEAM-Sim

October 15, 2010

Outline/Agenda

- Introduction
- Opportunity
- Proof of the Problem
- Blended IMI
- Accessibility and Transportability
- Return on Investment
- Case Study: 3-6 ADA
- Conclusion
- Questions

2

Introduction

- Motivating soldiers to want to learn can be a difficult.
- Classroom-based courses offer minimal opportunities to engage the soldier
- Blended IMI solutions can motivate soldiers
  - Repetitive training for:
    - Collective
    - Individual
    - Team performance tasks

Opportunity

- Have you ever watched a teenager with a new video game?
- Gaming technologies let the "player" be in control.
- Upwards of 70,000 new soldiers enlist in the Army each year
- Blended IMI solutions can be completely web-deliverable

Proof of the Problem

- Digitized course material have been in use for decades
- "DEATH BY POWERPOINT"
- Adds <u>little</u> to the effectiveness of the instruction
- Limited interaction
- Not the way today's soldier wants to learn!

Blended IMI

- A mix of media types
- Transforms the roles of instructors and students.
- Extracts value beyond traditional classroom instruction
- Provides benefits to four key areas:
  - Student interactions
  - Instructor interactions
  - Accessibility and transportability
  - Return on investment

909

Student interactions

- Present content consistent with their learning preference
- Three types of presentation:
  - Aural
  - Visual
  - Physical kinesthetic
- Blended IMI empowers the warfighter to take a more prominent role in their own education.

Instructor interactions

- Students better prepared
- Instructors can spend more time:
  - Answering questions
  - Discussing advanced concepts
  - Relaying valuable personal experience
- Monitor student progress in real-time
  - Instructional intervention
  - Focus on individual remediation

Accessibility & Transportability

- The ability to train anywhere, anytime, on almost every computer
- Updates to training materials can be rapidly deployed
- Mobile devices are emerging as an exciting new platform for the delivery of trainingAvoid wordiness
  - Important for training continuity between mobile and PC based versions be maintained

Return on Investment

- Blended IMI solution should be designed with existing equipment in mind
- Reduction in total hours of training
- Reduction in total number of instructors
- Decreases in travel and travel-related costs
- Lower logistical costs by reducing to total number of systems required for actual "hands-on" training

Case Study: 3-6 ADA

A blended solution incorporating all 4 levels of IMI with an After Action Review at the completion of each module.

Two Man Crew Drill Prototype

- Reference Material
    - Videos of Two Man Crew Drills provided by Raytheon
    - ARTEP-44-635-Drill Documentation

- Crawl – Walk – Run
    - Conforms to ABCS Style Guide for PEO C3T
    - Mirrors ARTEP-44-635-Drill manual
    - Provides after-action review
    - Simulation utilized to engage student

Two Man Crew Drill Prototype

Challenges

- Limited access to SME
- Limited reference material
- Prototypes designed around materials provided
- Limited access to funding

Achievements

- Over 2 hours of IMI
  - Video
  - Flash
  - 3D Models and Character animation
  - Interactive
  - Link-back to video for refresher
  - After-action Review
- Framework from which additional modules can be developed/deployed quickly and cost-effectively

Two Man Crew Drill Prototype

Fort Bliss 12 March 2009

- Raytheon Montana St. facility
- 4 NCO's from 3-6 ADA
- Intent is to review IMI training, and to determine
- if of sufficient fidelity and accuracy to present to
- student test group.

Achievements

- Received "Go this station" on our IMI training from Instructors
- NCO's provided useful feedback and criticisms.
  - Expressed a uniform belief that what we have is an extremely valuable step that they believe soldiers will gravitate to.

![MODSIM WORLD Conference & Expo]
Two Man Crew Drill Prototype

Fort Bliss 13 March 2009

- Raytheon Montana St. facility
- 12 Soldiers from 3-6 ADA
  - No prior "hands on" experience with the Patriot Launch Station hardware.
  - Morning Session
    - 8 Soldiers to be put through our IMI instruction
    - 4 Soldiers to attend AIT conference training
  - Afternoon Session
    - Take solders out on equipment in Abernathy Park
      - IMI soldiers on 1st Launch Station
      - AIT Conference soldiers on 2nd Launch Station
    - Have soldiers demonstrate what they learned

![MODSIM WORLD Conference & Expo]
Two Man Crew Drill Prototype
Fort Bliss 13 March 2009
Results

- IMI test group took to the Computer-based training "like ducks to water"
  - Needed minimal instruction
  - Wanted to run the training portions repeatedly
    - Competition amongst soldiers to get the "Best Time"
- Soldiers provided useful feedback and criticisms.
  - Enjoyed the IMI Training
  - Felt they actually learned something
- At Abernathy Park
  - IMI test group was able to tell their instructors what steps they needed to perform
  - Control Group needed to be told by their instructors what steps to perform

914

![MODSIM WORLD Conference & Expo]

Two Man Crew Drill Prototype

Data Points:

- 75% soldiers completed Sim "mission".
  - 75% soldiers were able to complete the Sim mission on their 3rd attempt.
  - 25% soldiers completed training multiple times.
- 75% soldiers scored 75% or higher on the initial assessment.
- 100% soldiers showed improvements in time and assessment scores.

![MODSIM WORLD Conference & Expo]

Basic IMI Screen
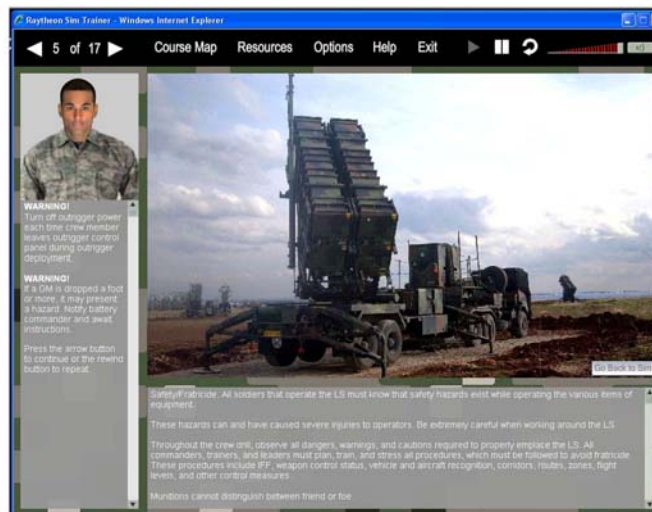
Interactive Game
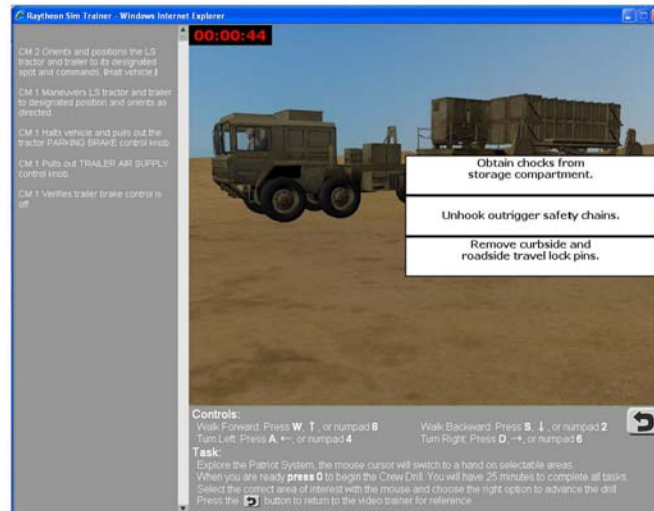
Incorrect Selection

916
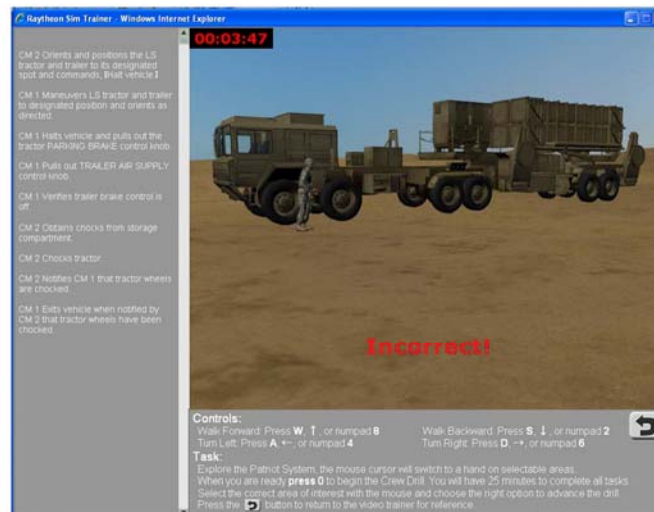
Video Link-Back for Remediation
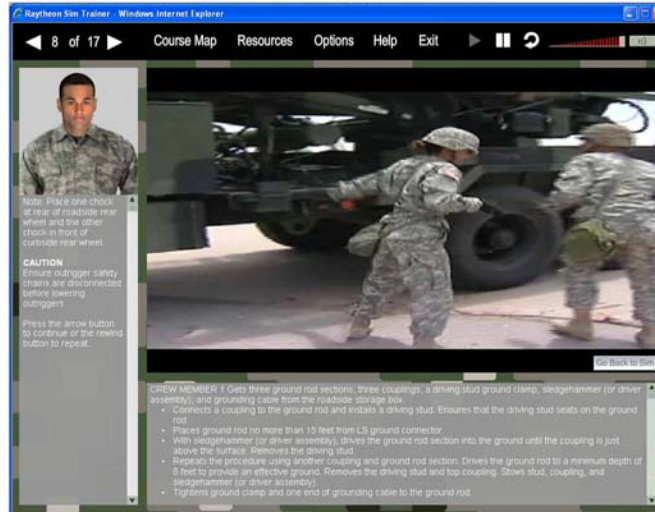
IMI vs POI Student Feedback
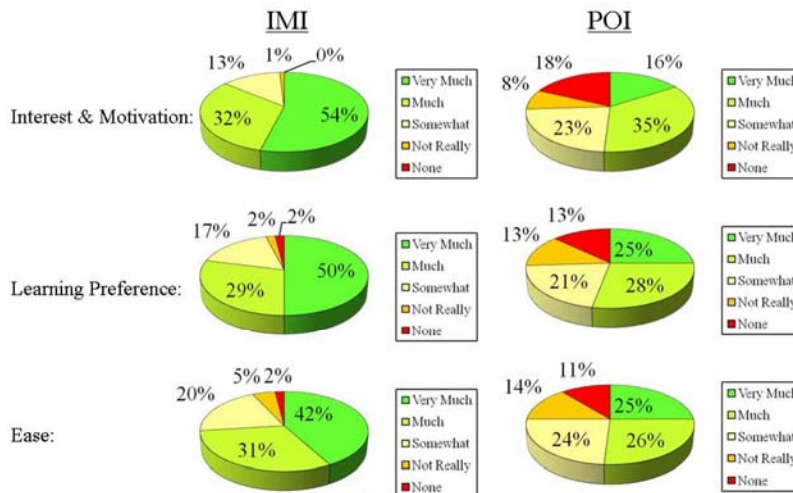
Expected Results

- Decrease the time to learn a crew drill by 50%
- Reduce instructor contact hours by 50%
- Improve training proficiency by 25%
- Decrease caution and safety violations by 25%

- Decrease equipment damage due to new operator fault by 30%
- Significantly increase interest and training motivation
- Significantly improve retention enabling certification at Soldier's first unit
- Decrease maintenance costs from inexperienced use of tactical equipment

Conclusion

- The more time you spend rehearsing, exploring options, and studying outcomes, the better you will become at a skill.
- Games & Blended IMI allows soldiers to continue learning and developing skills outside the classroom
- Increased engagement = higher retention levels + shorter time to subject mastery.
- Need to design for existing computer & handhelds to minimize adoption impact

Questions/Comments
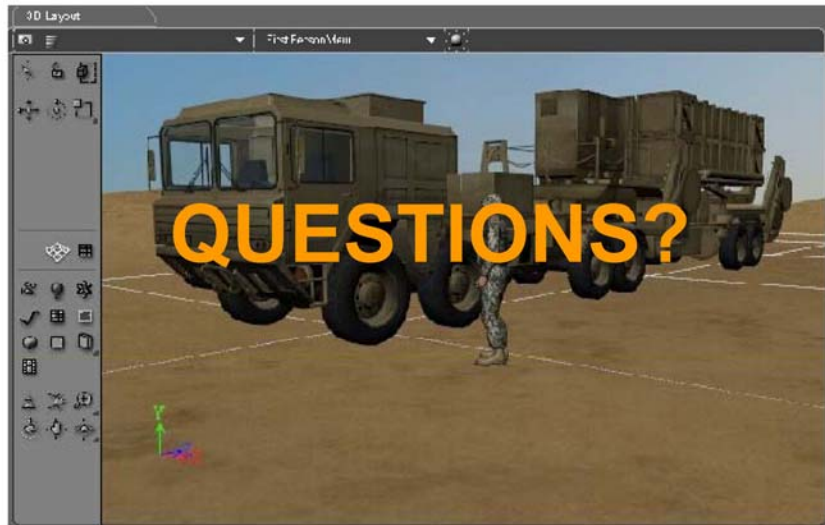
7.2   Optimizing Decision Preparedness by Adapting Scenario Complexity and Automating Scenario Generation

Robb Dunne, Sae Schatz, Stephen M. Fiore, Denise Nicholson
University of Central Florida, Institute for Simulation and Training
rdunne@ist.ucf.edu, sschatz@ist.ucf.edu, sfiore@ist.ucf.edu
dnichols@ist.ucf.edu
Jennifer Fowlkes
Chi Systems, Inc.
jfowlkes@chisystems.com,

Abstract.   Klein's recognition-primed decision (RPD) framework proposes that experts make decisions by recognizing similarities between current decision situations and previous decision experiences. Unfortunately, military personnel are often presented with situations that they have not experienced before. Scenario-based training (SBT) can help mitigate this gap. However, SBT remains a challenging and inefficient training approach. To address these limitations, the authors present an innovative formulation of scenario complexity that contributes to the larger research goal of developing an automated scenario generation system. This system will enable trainees to effectively advance through a variety of increasingly complex decision situations and experiences. By adapting scenario complexities and automating generation, trainees will be provided with a greater variety of appropriately calibrated training events, thus broadening their repositories of experience. Preliminary results from empirical testing (*N*=24) of the proof-of-concept formula are presented, and future avenues of scenario complexity research are also discussed.

1.0 INTRODUCTION

Decision-making in the military has evolved significantly since the eras of the phalanx and the Napoleonic regiment. Their strict, hierarchical command and control is inappropriate for the asymmetrical conflicts warfighters face now. Modern warfighters, often acting in small, distributed teams, are expected to make numerous, rapid decisions based on ambiguous information, all the while avoiding conflicts with rules of engagement, missions orders, and commander's intent. Unfortunately, lacking personal experience to draw upon, junior military personnel are often ill-prepared to make such complex decisions, and the outcomes of poor decisions may be disastrous: incorrect actions engaged, unnecessary risks taken, missions jeopardized, or casualties received.

Military personnel faced with unfamiliar situations are at a dangerous disadvantage if they lack the necessary decision-making skills. Unique situations are inherently risky and dangerous, fertile ground for poor decision-making [1]. Such situations demand increased attention, while draining vital cognitive resources [2], and they may engender anxiety that can detrimentally influence decision-making [3].

However, the more familiar a situation, the less risk and danger involved, and the better the decision-making. According to Klein's recognition primed decision (RPD) framework, decisions are made based upon decision-makers' available "pool" of internalized previously experienced situations [4]. Simply, the more experiences and situations individuals can draw from, the more likely they are to successfully navigate through multiple decision points. In regards to simulation, however, more experiences do not necessarily translate to a perception of scenario fidelity. Multiple experiences within simulation that are misaligned to the trainee's level of experience may not be perceived as accurately reflecting the complexity to be found in actuality. In their experiments, Bradley and Shapiro [5] found that at extreme levels of complexity, when cognitive capacity was taxed, everything became more real to participants. The challenge for simulation then, is presenting the optimum level of complexity to engender a sense of fidelity.

Just as decision-making in the military has evolved, so have its methods of training. Today, scenario-based training (SBT), defined as the purposeful instantiation of

simulated events to create desired psychological states [6], is a widely accepted instructional approach. One of the strengths of SBT is the presentation of varied situations that allow trainees to experience real-world problems prior to engagement. Systematically presenting training along a trajectory similar to Bloom's revised taxonomy [7], proceeding from declarative knowledge to higher-order levels of abstraction and creation, grounds the content and delivery in well-documented learning theory. Through such training, personnel can efficiently and effectively learn to integrate multiple skills, cope with realistic distracters, practice their higher order cognitive skills, and exercise naturalistic decision-making [8].

1.1 SBT Technologies

Although SBT can be delivered in a variety of ways, the remainder of this paper refers to scenario-based military training delivered through computer-generated virtual environments.

Currently, in the practice of SBT, scenarios are chosen for implementation by a trainer based on training objectives, a description of the scenario, the trainees' level of experience, and timeline requirements. Although training manuals may present sequences of training, they typically lack explicit recommendations for the content or sequencing of training scenarios. Thus, without a clear progression of scenarios, trainers rely on their own judgment when determining the scenarios' tasks and the order of presentation of scenarios within a set. Consequently, a trainer's sequencing may not align with trainees' levels of experience or performance, and mismatching trainees with training events can result in diminished—or even negative—decision preparedness.

Recent efforts to develop a SBT system that adapts to trainees' levels of experience have been undertaken by the authors, who are attempting to create *automated* methods that more effectively advance trainees through a

variety of increasingly complex training scenarios.

To achieve this goal, the investigators needed to operationalize the notion of *scenario complexity*. That is, the authors needed to objectively define the subjective idea of scenario difficulty. This objective formulation is necessary in order to develop the software algorithms that perform the automated instructional adaptation.

In the following section the authors present a brief definition of scenario complexity, identify and describe each of the characteristics used in the calculation of scenario complexity, and discuss the role of scenario complexity in the automation of scenario generation.

2.0 SCENARIO COMPLEXITY

To ensure trainees receive scenarios appropriate to their experience level, it is crucial to objectively define and instantiate *scenario complexity* so that computer-based training software can automatically assemble appropriate SBT sequences. Successful instantiation depends on taking the subjective and abstract and making it objective and concrete; that is, creating an *objective* computational metric of the *subjective* notion of difficulty.

The authors define *scenario complexity* as the objective quality of a scenario, which interacts with individual characteristics (such as trainees' expertise) to yield an individual's perception of the scenario's difficulty [9]. Most importantly, scenario complexity is calculated based upon three scenario elements that are *extrinsic from* rather than *intrinsic to* trainees: task complexity, task framework and cognitive context moderators. To be clear, the authors purposefully refrain from attempting to incorporate individual perceptions. Subjective interpretation of a task's difficulty or an individual's affective state in relation to a particular characteristic is un-actionable and cannot be calculated. It is for the purpose of operationalizing and incorporation into program software that objective calculation is pursued. For detailed

921

description of the formula see Dunne et al. [10].

2.1 Scenario Complexity Characteristics
The authors' computational definition of scenario complexity is as follows:

$$SC = (TC + TF) * CCM \qquad \text{Eq. (1)}$$

where SC = scenario complexity, TC = Task complexity, TF = Task framework, and CCM = Cognitive context moderators.

Each of the variables that comprise the total SC can be manipulated to increase or decrease the total SC. In addition, they can be altered to maintain the trainee in the same complexity range while at the same time presenting variety within the scenario. Each variable is calculated through individual functions that involve the following sub-variables.

2.1 Task Complexity
The task complexity component is subdivided into component complexity and coordinative complexity.

2.1.1 <u>Component Complexity</u>
The component complexity characteristic is composed of three sub- variables: the number of subtasks, required acts, and information cues.

First, component complexity considers the number of subtasks. Each training scenario is designed around at least one task with attendant learning objective(s). A task may stand alone, without a sub-task, but frequently, tasks include sub-tasks that must also be performed.

Second, each subtask requires one or more specific acts. Although a required act is principally a pattern of behavior, novice trainees internalize the conscious choice to engage in the behavior until it becomes automatic. Increasing the number of required acts presents greater opportunity for

transitioning conscious behavior to unconscious, automatized decisions.

Finally, each subtask may require monitoring of information cues. An information cue is a discrete source of information that must be monitored and/or processed from the environment. The trainee who is aware of these cues and chooses to monitor them will attain the desired performance more efficiently than a trainee who does not.

2.1.2 <u>Coordinative Complexity</u>
Coordinate complexity is concerned with the integration of subtasks and associated acts, which may be necessary for successful task completion. These subtasks are integrated and involve synchronization of activities to achieve the common goal or objective [11].

Without coordination of these subtasks and acts, trainee performance will suffer. By manipulating the degree of integration, trainees are presented with increasing levels of scenario complexity requiring, in turn, a greater number of decision points.

2.2 Task Framework
Task framework accounts for the relation between task paths and the outcome associated with each, and it addresses which outcomes are possible in a given task [12].

The authors suggest it is the task framework characteristic where the interplay of decision preparedness, performance and complexity is most acute. Tasks such as those with a single goal and a single means or path to achieve that goal, are well-defined. Tasks with multiple goals with several possible means or paths to achieve the goals are ill-defined tasks. Deciding if a particular means or path will achieve the desired goal requires a resolution of existing ambiguity; a calculation of potentiality for each path's success. Ambiguity and complexity make it difficult for decision-makers to determine what the possible outcomes might be, let alone the value they assign to them [13]. Increasing ambiguity and complexity is

therefore conjectured to be highly influential in advancing decision preparedness.

2.3 Cognitive Context Moderators

Cognitive context moderators address factors that often increase stress and distraction present in the scenario and are defined as external stimuli that affect the operator by increasing load and reducing cognitive resources for the task, thus causing less complex decisions to appear more complex.

These moderators can influence the resolution or quality of the evidence available for supporting judgments. According to Macmillan and Creelman [14] when attempting a forced-choice judgment to identify, for example "known or unknown", performance will be superior during a clear daylight encounter compared to night or under hazy conditions because the available evidence is superior.

It must be reiterated that these characteristics are built into the scenario; it remains the decision of the trainee to engage in these acts, monitor cues and satisfy scenario criteria. Through presentation of increasing levels of complexity, the scenarios contribute to the growing pool of experiences from which the trainee will draw.

The authors suggest automated generation of scenarios that take trainees on an efficient and effective trajectory towards optimized decision preparedness can be accomplished by this operationalization of scenario complexity. To ensure that adaptive generation results in positive outcomes, systematic implementation of this training framework must be grounded in decision-making and learning theory.

The following section describes two major theories adaptive scenario generation draws upon to increase decision preparedness.

3.0 OPTIMIZING DECISION PREPAREDNESS

The ability to make timely, appropriate, and effective decisions is an essential competence for warfighters [15]. Buch and Diehl [15] found that increases in the quality of decision-making have largely been by-products of in-field experience. However, they concluded that judgment and decision-making capability can be improved through training, incorporating situation-specific exercises and increasing the variety of variables as training progressed.

With an objective value for each scenario's complexity, variables can be manipulated to increase or decrease variety and complexity. This manipulation must be calibrated to trainee performance, aligning the level of complexity associated with a scenario to the instructional needs of trainees. However, in order to align the scenario to an individual's training needs—making the scenario that is "just right" —the simulation must employ a systematic instructional methodology [16].

The following section describes how the instructional methodology, based on Bloom's taxonomy, is supported through the operationalization of scenario complexity for SBT. Also discussed is the role of automated scenario generation and how it is utilizes Klein's recognition-primed decision-making framework and, by extension, Klein and Baxter's cognitive transformation theory [17] to improve decision preparedness. The section also describes the well-documented instructional efficacy of Vygotsky's zone of proximal development [16], used to enable proper alignment and sequencing of scenarios.

3.1 Recognition-primed Decision-making (RPD)

Recent decision-making theories have focused on decisions that are made in complex situations with high stakes, uncertainty and time pressure [13]. Naturalistic decision-making theory (NDM) attempts to explain such decisions by

923

suggesting that a decision-maker continually assesses a situation in order to recognize familiar characteristics and make judgments [19]. NDM has been observed across domains, such as firefighting and the military [20], where the decision-maker initially assesses the situation, looks for familiar patterns or prototypes, determines which goals make sense, identifies the relevant cues to expect, and determines what action should will be most appropriate.

Under the umbrella of NDM research, the RPD model has gained significant influence over the past 10 to 15 years. This model is based on the supposition that, in complex situations, humans usually make decisions based on the recognition of similarities between the current decision situation and previous decision experiences. Simply, Klein proposes that decisions are made based on the recall of the consequences of previous decisions made in similar situations.

Cognitive studies have shown that over 95% of human decisions conform to the RPD model in time-stressed situations [21]—the very type of situation frequently encountered by military personnel.

An extension of RPD is cognitive transformation theory [CTA]. This theory states that as novices progress towards expertise they develop "knowledge shields" that serve to protect their established concepts. These shields can negatively affect knowledge and skills acquisition and, in the operational theater, may lead to situations where default decisions are made based on biased judgment rather than experience [17].

However, Klein and Baxter reference Waller, Hunt, and Knapp [22] who found that varied and sufficient exposure to virtual training environments provides trainees with the needed practice to construct valid mental models and alleviate the obstacles created by such knowledge shields [17].

3.2 Zone of Proximal Development (ZPD)

The theory of the zone of proximal development (ZPD) is the scientific basis for why training (in this case, scenario) difficulty must be appropriately matched to trainees' current skill levels. The ZPD is the range within which learning and training tasks are neither too hard nor too easy. According to Vygotsky, development will only occur when a trainee is confronted by a task that lies within the zone, because if a task is too easy then no development will happen (although gains in fluency and accuracy may occur simply through repetition) [23] and if a task is too difficult to complete successfully, no cognitive development will occur [24] and motivation may suffer.

3.3 Implementation

For illustrative purposes let us say during SBT, a trainee performs exceptionally well. They have attained the goals and successfully navigated decision points. Another trainee has not performed so well. They have not attained the goals, and their decisions were inadequate. Following the principle of ZPD the next scenario in sequence should be neither too difficult nor too easy. Due to such differing performance, under today's SBT approach, it cannot be expected that one scenario will address the learning needs of both trainees. However, if the next scenario in sequence could increase in complexity for the high performer while either maintaining the current complexity level, with variation, or remediating at a lower complexity range for the other trainee, then the training should be more effective and efficient. Achieving this is the authors' goal.

Scenario complexity ranges derived from the computational formula developed by the authors, allow software to determine, with or without manual input from the trainer, the appropriate sequence of experiences for both of these trainees. If, according to ZPD, negative training exists outside both the upper and lower bounds, then scenario complexity defines the upper and lower

ranges of the complexity of the appropriate sequenced scenario.

Aligning scenario sequencing to trainee performance, therefore, ensures that trainees are not presented with decision points and situations which are too hard or too easy, but remain in the area of ideal learning, while presenting varied situations and experiences increasing decision preparedness.

To establish empirical basis for this theory of scenario complexity an initial study is being conducted. The focus of this study, as well as its design, methodology, and preliminary findings are presented in the following section.

4.0 EXPERIMENTATION
A pilot study was conducted to ascertain if the researchers' theory, equations, and hypotheses are usable and conceptually sound. Incorporating lessons learned from this pilot study, any reformulation to the original equation that is suggested by the results will precede further research. The end-state study will include construction of dynamic scenario-based training of varying levels of complexity designed by military personnel.

4.1 Design and Methodology
This pilot study involved a single sample group (N=24) comprised of undergraduate students recruited from a large southeastern university. Each participant completed four surveys, which asked participants to indicate their perception of the complexity levels of various situations. These situations' complexity levels were calculated *a priori* by the authors, using the described scenario complexity formula.

The first series of 6 questions asked the participants to indicate on a scale of 1-20 how simple or complex they thought it would be to drive a car: (Q_1) in an empty parking lot, (Q_2) to a familiar destination in light traffic, (Q_3) to an unfamiliar destination in a familiar city, (Q_4) to an unfamiliar destination in an unfamiliar city using a map, (Q_5) to a familiar

destination in heavy traffic and severe thunderstorm, and (Q_6) to an unfamiliar destination using a map, in light traffic and mild rain.

The second series of 6 questions asked the participants to indicate how simple or complex they thought it would be to: juggle three balls of different sizes (Q_7), juggle two balls and walk at the same time (Q_8), juggle two balls (Q_9), juggle three balls of the same size and walk (Q_{10}), juggle three balls of the same size while reciting the ABC's (Q_{11}), and juggle two balls of different sizes and walk while reciting the ABC's (Q_{12}).

The third series of 6 questions asked the participants to indicate how simple or complex they thought it would be to drive a car: in an empty parking lot while talking on the phone (Q_{13}), to a familiar destination in light traffic while talking on the phone (Q_{14}), to an unfamiliar destination while talking on the phone (Q_{15}), to an unfamiliar destination with a map while talking on the phone (Q_{16}), to a familiar destination in heavy traffic and severe thunderstorm while talking on the phone (Q_{17}) and, to an unfamiliar destination with a map in light traffic and mild raid while talking on the phone (Q_{18}).

The fourth series of questions set the participants into a single scenario comprised of 7 different situations and were asked to indicate how simple or complex they thought each situation would be to: drive to their friends house (Q_{19}), drive to their friends new house where they've never been before (Q_{20}), surf (Q_{21}), play a ping-pong toss game (Q_{22}), play the toss game with a fan aimed out over the game (Q_{23}), choose from a large menu of pizza toppings under a time constraint (Q_{24}) and to pick one of two sodas (Q_{25}).

4.2 Hypotheses
The authors hypothesized that the scenario complexity computation calculated by the authors would be similar to the results identified by the participants. In other words, the authors expected that the formula would yield relatively comparable results, regardless of who assessed the described scenarios.

4.3 Results

An independent one-sample *t*-test was conducted on the participant (*n*=24) responses to evaluate whether their means were significantly different from the test values established *a priori*.

Z-scores were also derived to ascertain the degrees of agreement among participants' values. *Z*-scores between -1.249 and 1.249 indicate relative agreement as to a situation's difficulty. *Z*-scores < -1.25 or > 1.25 represent significant disagreement.

The following table shows the z-scores for each item. The total mean and standard deviation are noted below.

Calculated Z-scores*

Item #	Response Mean	A Priori Expected Score	t-test value (mean/expected)	Participants' Z-Scores
1	1.45	1	$p = .06$	-1.75
2	2.91	5.6	$p = .00$	-1.46
3	7.62	6.4	$p = .08$	-.54
4	13	10.4	$p = .00$	.52
5	11.08	20	$p = .00$	.14
6	13.2	15.2	$p = .03$	.56
7	15.5	13.5	$p = .06$	1.01
8	12.2	2.8	$p = .00$	.36
9	6.37	1	$p = .00$	.78
10	16.08	12.1	$p = .00$	1.13
11	15.54	13.5	$p = .00$	1.02
12	14.33	19.3	$p = .04$	.78
13	2.2	1	$p = .00$	-1.6
14	4.12	5	$p = .00$	-1.22
15	11.29	5.7	$p = .05$	.19
16	14.75	13.3	$p = .00$	.87
17	14.79	20	$p = .1$	.87
18	16.58	13.3	$p = .00$	1.22
19	1.7	6	$p = .00$	-1.69
20	13	8.6	$p = .00$	.52
21	13.66	20	$p = .00$	.65
22	8.79	2	$p = .00$	-.31
23	13.2	6.6	$p = .00$	.56
24	4.95	3.3	$p = .11$	-1.06
25	3.29	1	$p = .00$	-1.39

Table 1: Calculated Z-scores, means, and *t*-test results; *N = 24, 95% confidence.

In the graph below the *a priori* complexity levels and participants' responses are illustrated. The light line represents the calculated levels while the dark line represents the mean levels of the aggregated participant responses.

Expected levels and participant response means
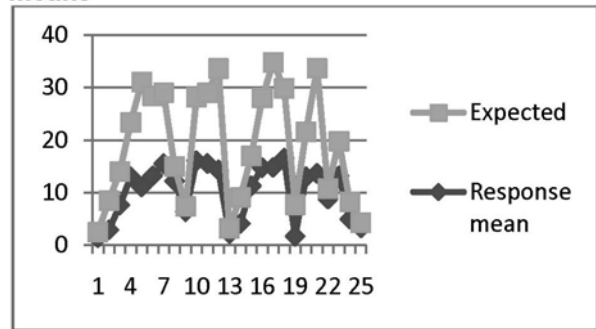
Figure 1: Expected and participant complexity levels

4.4 Conclusion

Two salient conclusions can be drawn from the results of this study. First, *Z*-scores indicate significant disagreement at both ends of the complexity spectrum, suggesting that the low- and high-end complexity scenarios have a wider range of responses than do the middle-ground values. Second, there was a significant level of agreement of perceived complexity relating to those situations which occupy the middle-ground.

5.0 DISCUSSION

These results indicate areas of both promise and challenge. As a proof of concept, the authors believe their formalization of scenario complexity is headed in the correct direction; however, there are adjustments and considerations which must be accounted for in future iterations. Attempting to quantify subjective perceptions is rife with inconsistency. Two participants may give two different values for the same situation even though they both perceive the situation as being very simple. Further, levels of experience may have, as Klein suggests, played a discriminating role. Participants with a larger repository of experiences to draw from may have perceived the described

926

situations with a lower level of difficulty compared to those with fewer experiences. This argues not only for the support of RPD but also for the necessity to adapt training and complexity levels to the trainee within their zone of proximal development or risk de-motivation and inefficient training.

Additionally, disparity between participant responses and the *a priori* calculated values may point to participants' over-confidence and/or a lack of understanding of the characteristics of the situation. That is, while the scenario complexity formula takes into account such characteristics as number of task outcomes and cognitive context moderators, participants may not be able to readily identify such attributes. Similarly, participants may identify such factors, but may not consider them impactful. For instance, in the item that asked participants for an evaluation of their ability to utilize a cell-phone while in traffic during a thunder storm, over-confidence may lead some to inaccurately assess the challenge of performing under such conditions.

Lessons from this pilot study suggest further refinement of both the formulation of scenario complexity as proposed, and of the research design. In order to investigate validity and reliability across the entire spectrum of complexity, future research requires recalculation of the weight given to each characteristic, and reformulation of the task-framework equation in particular, in addition to controlling for levels of experience and including more detailed instructions for participants.

Branching outward from this avenue of investigation, promising areas of research include, but are certainly not limited to, investigating the role of cognitive load on scenario complexity. That is, does objective calculation of complexity adequately address both a novice and expert trainee's different requirements of cognitive load? Second, in respect to integration of team members, how do multiple agents impact an individual's performance in dynamic, adaptive scenarios?

Third, to what degree is perception of scenario fidelity affected by increasing scenario complexity following Vygotsky's ZPD? Finally, what role do personal characteristics such as metacognition and self-efficacy play in trainees' performance?

6.0 REFERENCES

[1] Shapiro, S. (2010). Decision Making Under Pressure. *Futurist, 44*(1), 42.

[2] Dobbins, I., & Han, S. (2009). Rules Versus Evidence in Memory and Non-Memory Decision-Making. *Military Psychology, 21,* 113-122.

[3] Lerner, J. S., Small, D. A., & Loewenstein, G. (2004). Heart Strings and Purse Strings: Carryover Effects of Emotions on Economic Decisions. *Psychological Science, 15,* 337–341.

[4] Klein, G. A. (1989). Recognition-primed Decisions. In *Advances in Man–Machine Systems Research*, vol. 5, W. B. Rouse, Ed. Greenwich, CT: JAI Press, 47–92.

[5] Bradley, S., & Shapiro, M. (2004). Parsing Reality: The Interactive Effects of Complex Syntax and Time Pressure on Cognitive Processing of Television Scenarios. *Media Psychology, 6*(4), 307-333.

[6] Martin, G., Schatz, S., Bowers, C.A., Hughes, C. E., Fowlkes, J., & Nicholson, D. (2009). Automatic Scenario Generation Through Procedural Modeling for Scenario-based Training. *Proceedings of the 53rd Annual Conference of the Human Factors and Ergonomics Society, Santa Monica, CA.,* 1949-1953.

[7] Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice 41*(4), 212.

[8] Cannon-Bowers, J. A., & Salas, E. (1998). Individual and Team Decision Making Under Stress: Theoretical Underpinnings. In *Making Decisions Under Stress: Implications for Individual and Team Training*, Cannon-Bowers, J. A., & Salas, E., Eds. Washington, DC: APA Press, 17–38.

[79 Lum, H., Fiore, S. M., Rosen, M. R., & Salas, E., (2008). Complexity in Collaboration: Developing an Understanding of Macrocognition in Teams Through Examination of Task Complexity. *Proceedings of 52nd Annual Meeting of the Human Factors and Ergonomics Society, New York, NY.,* 1425-1429.

[10] Dunne, R., Schatz, S., Fiore, S., Martin, G., & Nicholson, D. (in press). Scenario-based Training: Scenario Complexity. *Proceedings of the 54th Annual Conference of the Human Factors and Ergonomics Society*, San Francisco, CA.

[11] Baker, D., Day, R., & Salas, E. (2006). Teamwork as an Essential Component of High-Reliability Organizations. *Health Services Research, 41*[4P2], 1576-1598.

[12] Campbell, D. (1991). Goal Levels, Complex Tasks, and Strategy Development: A review and Analysis. *Human Performance, 4*[1], 1-31.

[13] Knighton, R. (2004). The Psychology of Risk and its Role in Military Decision-Making. *Defence Studies, 4*[3], 309-334.

[14] Macmillan, N., & Creelman, C. (1991). *Detection theory: A user's guide.* New York, NY US: Cambridge University Press. Retrieved from PsycINFO database.

[15] Buch, G., & Diehl, A. (1984). An Investigation of the Effectiveness of Pilot Judgment Training. *Human Factors, 26*(5), 557-564.

[16] Oser, R. L., Cannon-Bowers, J. A., Dwyer, D. J., & Miller, H., (1997). An Event Based Approach for Training: Enhancing the Utility of Joint Service Simulations. Paper presented at the 65th Military Operations Research Society Symposium, Quantico, VA.

[17] Klein, G., & Baxter, H. C. (2009). Cognitive Transformation Theory: Contrasting Cognitive and Behavioral learning. In D. Schmorrow, J. Cohn & D. Nicholson [Eds.], The PSI handbook of virtual environments for training and education: Developments for the military and beyond. Vol. 1: Learning, requirements, and metrics. (50-65). Westport, CT: Praeger Security International.

[18] Vygotsky. L. S. (1986). *Thought and Language.* Cambridge. MA: MIT Press.

[19] Chapman, T., Nettelbeck, T., Welsh, M., & Mills, V. (2006). Investigating the Construct Validity Associated with Microworld Research: A Comparison of Performance Under Different Management Structures Across Expert and Non-expert Naturalistic Decision-making Groups. *Australian Journal of Psychology, 58*(1), 40-47.

[20] Klein, G., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid Decision-making on the Fire Ground. *Proceedings of the 30th Annual Human Factors Society Meeting,* Dayton, OH.

[21] Klein, G. A. (1998). Recognition-primed Decision Making. In *Sources of Power: How People Make Decisions.* Cambridge,MA: MIT Press, 5–30.

[22] Waller, D., Hunt, E., & Knapp, D. (1998). The Transfer of Spatial Knowledge in Virtual Environment training. *Presense: Teleoperators and Virtual Environments, 7*(2), 129-143.

[23] Vygotsky. L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes.* Cole, M., V. John-Steiner, S, Scribner, and E. Souberman. eds, Cambridge, MA: Harvard University Press.

[24] Louis, G. (2009). Using Glasser's Choice Theory to Understand Vygotsky. *International Journal of Reality Therapy, 28*(2), 20-23.

8.0 ACKNOWLEDGMENTS

7.3 Leveraging M&S in Soft Skills Training for the DoD

Leveraging M&S in Soft Skills Training for the DoD

James D. Cimino
D2 TEAM-Sim
jcimino@d2teamsim.com

Soft skills, also called "people skills," are typically hard to observe, quantify and measure. These skills have to do with how we relate to each other: communicating, listening, engaging in dialogue, giving feedback, cooperating as a team member, solving problems and resolving conflicts. Most of the soft skills training is scenario based, utilizing written or video-based scenarios, with limited or no branching, as well as quantitative feedback. This paper will outline a game-based approach to configurable, scenario-based, soft skills training. The paper will discuss the application of realistic visual behavior cues (e.g. body language, vocal inflection, facial expressions) and how these can benefit the learner. Using the concept of a "virtual vignette," this paper will discuss a prototype system intended to teach suicide prevention and provide qualitative feedback to the learner. The paper will also explore other soft-skills training applications for this technology.

1.0 INTRODUCTION

By definition, soft skills are more than just tangible facts. According to Konopka and Dupre, while it is difficult, it is not impossible to do "soft skill" development in non-traditional, non-face-to-face settings. In fact, there may actually be real advantages to doing it "from a distance" [2]. This paper will present an approach to using gaming technology to teach soft skills, with a focus on teaching suicide prevention, and looking at other potential applications.

The US Army [7] has provided a number of resources to teach its leadership and soldiers on how to recognize and properly deal with a potential suicide. The mission of the Army Suicide Prevention website [6] is to improve readiness through the development and enhancement of the Army Suicide Prevention Program policies. These policies are designed to minimize suicide behavior, thereby preserving mission effectiveness through individual readiness for Soldiers, their Families, and Department of the Army civilians. A major component of the training related to this Suicide Prevention Program focuses on "soft skills."

2.0 SOFT SKILLS

The US Office of Personnel Management [1] has identified an overlapping set of core competencies for executive development.

They are Leading Change, Leading People, being Results Driven, having Business Acumen, and Building Coalitions. Within these competencies are subsets of soft skills that could be considered relevant for Suicide Prevention "soft skills" training:

Leading Change
- Flexibility
  - Is open to change and new information; rapidly adapts to new information, changing conditions, or unexpected obstacles.

Leading People
- Conflict Management
  - Encourages creative tension and differences of opinion. Anticipates and takes steps to prevent counter-productive confrontations. Manages and resolves conflicts and disagreements in a constructive manner.

Results Driven
- Decisiveness
  - Makes well-informed, effective, and timely decisions, even when data is limited or solutions produce unpleasant consequences; perceives the impact and implications of decisions.
- Problem Solving
  - Identifies and analyzes problems; weighs relevance and accuracy of information; generates and evaluates alternative solutions; makes recommendations.

Building Coalitions
- Influencing/Negotiating

○ Persuades others; builds consensus through give and take; gains cooperation from others to obtain information and accomplish goals.

Every one of these soft skills could be utilized when dealing with a soldier who is a potential suicide risk.

3.0 CHALLENGES

According to Lauren Smith [3] there are several challenges to developing effective soft skills training. A common mistake is to limit interactivity. Web-based training that consists mostly of text, graphs and simple images tend to fall short when it comes to the transfer and application of the knowledge. Material in a format that encourages participants to consider how to apply a skill in a variety of situations and contexts will not help participants actually learn the skill.

Another challenge is to make abstract soft skills concepts more concrete and tangible by providing the appropriate steps, definitions, illustrations, and relevant examples of how to apply them.

It is important to allow participants to evaluate their current behavior and to gauge their proficiency. The challenge is to provide relevant scenarios that illustrate both poor and excellent use of the associated soft skill. Students may have trouble retaining general concepts, but they will remember the scenarios if they can relate to them.

Keeping the students interested is another challenge to delivering effective soft skills training. Students are more inclined to practice and learn the information presented in a training program when they are engaged. This used to mean utilizing a combination of different types of interactive multimedia such as audio, video, graphics, animations and games into the content. Content that simply emulates the standard "Death by PowerPoint" will not engage today's young learners.

To make soft skills training effective, one would need to provide expert feedback to participants throughout in order to make them aware of their progress. Immediate feedback is important to learning and should be incorporated into all web-based exercises where possible.

4.0 ROLE PLAYING

According to Charles Green [11], soft skills training comes in three forms: role plays, video replays and case discussions. The concept of using role playing techniques to teach soft skills is not a new. According to Green, there is no substitute for realistic "muscle memory" activity when it comes to learning soft skills. Role playing allows one to present more complex, hypothetical scenarios [11].

An example of this is a recent program [9] designed to be used by trained speakers. The goal was to offer doctors a look at how their treatment decisions affect long-term outcomes for patients with type 2 diabetes through 3-D video animations. In this software solution, a doctor would role-play with a virtual patient. These virtual patients were created such that the instructing physician could alter the characteristics of a selected patient in terms of their weight, age and A1C levels.

FIGURE 1: Video-based Role Playing Application

Changing any of these characteristics presented the instructor with immediate visual feedback. The selected patient would get immediately thinner or heavier, younger or older, based upon the associated settings that were selected via an intuitive Graphical User Interface (GUI). The back-end software provided over 900 different scenarios based upon the combinations of settings that the instructor selected. Both videos and text were used to display the selected patient's "responses" to the Doctors examination. This tool provided an interactive approach to getting Doctors to engage in discussions about potential treatment plans for the presented patient. It also afforded the instructor a wide array of patient profiles.

The solution utilized live actors to represent each patient. Over a period of several weeks, make up artists made each patient "older" and "fatter." At each stage, the patient was photographed and video tapped both sitting on the examination table and responding to questions.

And while the end product was visually appealing, the implementation does have its drawbacks. Any desired additions or changes to the scenarios would require the actors, make-up artists and videographers to once again be brought together and create the desired footage. If an actor wasn't available, then this could negatively affect the product. These types of changes are costly in both time and dollars.

5.0 TECHNOLOGY TO THE RESCUE

According to Margaret Kaeter [10], today's soft skills programs are more like arcade games. Advances in both Commercial Off-The-Shelf (COTS) gaming technology, as well as improvements in 3D modeling and character animation, make these viable alternatives to using live actors. 3-D worlds similar to the popular Second Life also have potential for teaching soft skills, but are beyond the scope of this paper.

The goal is to merge the engaging nature of the video game and the experiential nature of this medium to teach soft skills. A blended Interactive Multimedia Instruction (IMI) approach to this training may be the right direction. The ideal solution might utilize a combination of media, such as slides and videos, to teach the basic concepts for the utilization of necessary soft skills, and gaming technology to provide the scenarios. The remainder of this paper will focus on the COTS gaming application for the visualization of these training scenarios.

6.0 VIRTUAL VIGNETTE

The concept for the Virtual Vignette is that of a role-playing game. The person playing the game (a trainee) is placed into a scenario where they must interact with the game character that is a potential suicide risk. At the start of the game, the trainee gets to select their character, and alter certain elements for the scenario.

FIGURE 2: Virtual Vignette Character Properties Screen

These elements might include the characters stress level, age, suicide risk, tours of duty, etc. All of these elements would have corresponding questions, responses, actions, animations and consequences. These items would also have a time element associated with them such that any interaction delays could escalate the consequences and vastly change the outcome of the game.

931

The virtual Vignette will also provide an After Action Review (AAR) so that the trainee can get feedback as to what they did right, or what they did wrong, and why. The AAR will also detail sequence errors within the Virtual Vignette.

6.1 Look And Feel

In order for game-based training to be effective, it must look real, feel real and sound real. Today's gaming technology brings together the confluence of technologies necessary to make an effective training scenario for suicide prevention. The first of these technologies is the realistic 3D modeling of characters and environments. The characters need to look and move like real people. The avatars in virtual worlds, such as Second Life, have the level of detail or realism that today's young gamers have come to demand. If it doesn't look real, it will not be as engaging.

The 3D characters and their environment also need to be relevant. If a person is attempting to teach soldiers about suicide prevention, then the characters need to be in the appropriate uniform or dress, and should also be germane to the theater of operation that the trainee finds themselves in. A soldier about to be deployed to Afghanistan should be presented characters in the appropriate desert camouflage, with terrain and quarters that are representative of the locale. Anything else will detract from the effectiveness of the training and weaken the overall experience.

FIGURE 3: Virtual Vignette Character Interaction Screen (Role Play)

The animations and movements of the 3D characters need special attention as well. A character that represents a potential suicide risk needs to display the appropriate body language and gestures such that the scenarios feel real. Subject Matter Experts (SMEs) must work closely with 3D graphic artists and animators to get all of the character's idiosyncrasies and body movement correct. If one looks at top selling video games, a lot of detail goes into the realism of the 3D character movements. If quality is lacking in this area, the focus will be taken away from the realism of the training. A person only needs to have watched a movie with poor acting to understand how a bad actor will take the focus away from the story.

Sound is often over-looked, but it is another key component to the realism of the virtual vignette. 3D Characters need to speak, and their vocal inflections need to convey the emotions appropriate for the moment. Background noise is also important for the realism of the scenario. If the characters are standing near a truck, a player should hear the engine idle. If a door opens, he or she should hear it open and then slam shut.

Details, even minute ones, are paramount to creating a realistic scenario. Examples include, but are not limited to, the lighting, items on the floor, dust blowing, etc. Though they are infinitely easier to render in 3D

932

barren, objects such as desks are not typically clean, flat surfaces in real life, and they shouldn't be in a virtual one. These details add to the realism and help sell the scenario. These items could even include tools that one might need the player to interact with such as a telephone (to call for help), a pencil and paper (to write notes), or even a gun held by the soldier considering suicide.

6.2 Consequences

A good game needs to have consequences both positive and negative. These outcomes are driven by the choices made by the trainee. The outcomes should include positive and negative extremes, with various results in between. In the case of a suicide prevention vignette, the obvious negative outcome is that the soldier harms himself or another character. The positive outcome would be that the soldier agrees to accompany the trainee to seek help. SME's would work with the game scenario designers to detail alternative outcomes between these extremes.

6.3 Interaction

The trainee playing the game needs to have some degree of freedom to move about the virtual environment. Keep in mind that not everyone that will use this training will be a "gamer," so there should be limits in terms of keyboard controls or keys to invoke specialized moves.

In order for a game to be engaging, it needs to challenge the player. A simple way to achieve this is to vary the scenario slightly each time. In a role playing game such as this, the order in which the trainee asks the questions of the 3D character may impact the outcome. Based upon reactions from the avatar, certain questions could become unavailable, or new questions added.

The starting positions of the characters can also impact the overall game play and scenario randomness. For example, if the scenario is set in a barracks, the character can be placed either close to the trainee or further away. As the player starts to approach the avatar, a collision boundary could detect the proximity and trigger an adverse reaction from the character.

The role play interaction provides an excellent mechanism to focus on behavioral indicators and queues, helping to make the concepts more concrete. It will also help the participants gauge their own skill levels and provide them with ways they can improve their soft skills.

6.4 Programming

The complexity of the Virtual Vignette should not be understated. There is a lot of complicated programming and logic behind a system of this nature. However, once the base structure has been built, the potential for soft skills training is virtually limitless.

A core component for the Virtual Vignette is the scenario repository. This repository will contain all of the questions, responses, expert assessments, environments, sounds, character files and associated animations. The technical aspects of how all these components are constructed and integrated is beyond the scope of this paper. However, it should be noted that the Virtual Vignette will have a modular architecture, work within a standard web browser, and be SCORM conformant.

In lieu of live coaches, computer-generated feedback on responses to each scenario will serve as a self-check for the participants in the absence of an instructor.

7. ALTERNATIVE APPLICATIONS

While the initial application for the Virtual Vignette has been targeted to teach suicide

prevention, it has the potential to expand to other training initiatives for either soft or hard skills. Cultural awareness and sensitivity would be a logical application for this technology, where the avatar could be a village elder or even a suspected insurgent. This framework could be applied to the Army Combat Lifesaver Course, where the player would need to approach, assess, treat and successfully evacuate a wounded comrade.

8. CONCLUSION

Soft skills training will continue to be important for employees in all types of professions. The military is no exception. The type of computer-based training that has been discussed in this paper alleviates the discomfort of role-playing exercises. Students can go through training at their own pace, and take responsibility for their own development. Inevitably, this leads to better overall soft skills training results. Games that leverage 3D modeling and simulation are excellent for delivering realistic role play scenarios.

Regardless of the technology, it is important to recognize that solid content is at the core of quality soft skills training. Any effective training for soft skills must use a sound development approach and identify learning objectives. Ultimately, soft skills training is only as good as its content.

9. REFERENCES

1. U.S. Office of Personnel Management http://www.opm.gov/ses/recruitment/ecq.asp

2. "Soft Skills" From a Distance (2001) 17th Annual Conference on Distance Teaching and Learning
Konopka, Michael A, D.A.C.C.E, Professor, Sustaining Base Leadership and Management, Army Management Staff College
Dupre, Mary-Blaie, D.P.A., Professor, Sustaining Base Leadership and Management ,Army Management Staff College

3. Teaching the Intangibles, Smith, Lauren, Training and Development Magazine October 2007, 23 -25,

4. Business Solutions: Learning by Pretending, MICHAEL TOTTY Staff Reporter of THE WALL STREET JOURNAL, *May 23, 2005; Page R10*

5. "Learning Simulations: Experiential Education," CLO Special Report, October 2004.

6. Suicide Prevention Army Deputy Chief of Staff, www.preventsuicide.army.mil/

7 .US Army Medical Department Army Behavioral Health http://www.behavioralhealth.army.mil/sprevention/index.html

8, Radcliff, Je Bentley, "Why Soft Skills Simulation Makes A Hard Case for Sales Training", 2005, CompeteNet, http://www.competenet.com/downloads/SimulationWP-F1.pdf

9. Medical Marketing & Media, 2009, Digital Marketing, http://media.mmm-online.com/documents/11/21_mmm_bon09_2633.pdf

10. Kaeter, Margaret, January 28,2000, "Teaching soft skills the hard way", TechRepublic, http://articles.techrepublic.com.com/5100-10878_11-5032189.html

11. Green, Charles H., February 1, 2010, "Acquiring Soft Skills: You Gotta practice the Scales", Trust Matters,

http://trustedadvisor.com/trustmatters/732/Acquiring-Soft-Skills-You-Gotta-Practice-the-Scales

10. ACKNOWLEDGMENTS

Leveraging M&S in Soft Skills Training for the DoD

James D. Cimino
D2 TEAM-Sim

D2 TEAM-Sim

October 15, 2010

Outline/Agenda

- Introduction
- Soft Skills
- Challenges
- Role Playing
- Technology to the Rescue
- Virtual Vignette
- Alternative Applications
- Conclusion

2

Introduction

- Soft skills: more than just tangible facts
- Difficult to do "soft skill" development in non-traditional, non-face-to-face settings
- Utilize gaming technology to teach soft skills
- A major component of the Suicide Prevention Program training focuses on "soft skills."

Soft Skills

- Core competencies
  - Leading Change   – being Results Driven
  - Leading People   – Building Coalitions

- Subsets of soft skills relevant for Suicide Prevention training
  - Flexibility
  - Conflict Management
  - Decisiveness
  - Problem Solving
  - Influencing/Negotiating

Challenges

- Not to limit user interactivity
- How do we make abstract concepts concrete and tangible
- Provide relevant scenarios
- Keeping the students interested
- Provide expert and immediate feedback

Role Playing

- Three forms:
  - Role plays
  - Video replays
  - Case discussions
- Role playing allows for more complex and hypothetical scenarios
- No substitute for realistic "muscle memory" activity

Technology to the Rescue

- Today's soft skills programs are more like arcade games
- Viable alternatives to using live actors
  - COTS gaming technology
  - 3D modeling and character animation,
- Goal: merge the engaging nature of video games and the experiential nature of this medium
- Use a blended IMI approach

Virtual Vignette

- Easy
  - to populate with scenarios/modify existing scenarios
  - to update
  - to re-use
  - to deploy
- Engage
  - interactive 3D game-like environment to role-play
- Evaluate
  - After-action Review recaps what was done correctly, incorrectly or out-of-sequence.
- Educate
  - Repository of relevant videos and supporting materials (i.e. documents, presentations, and images)

940

941
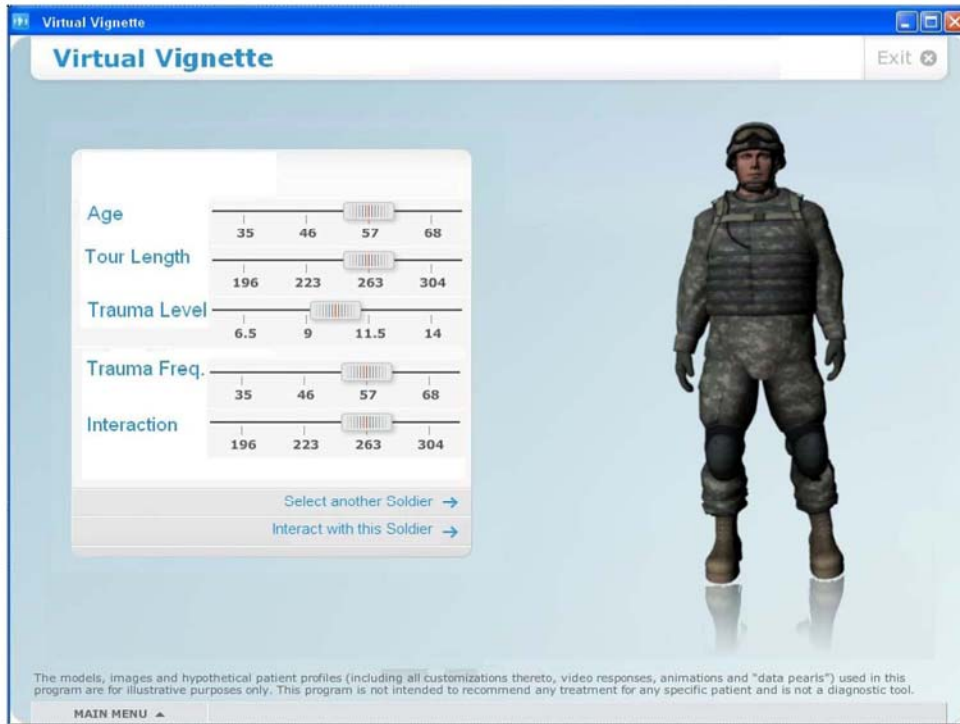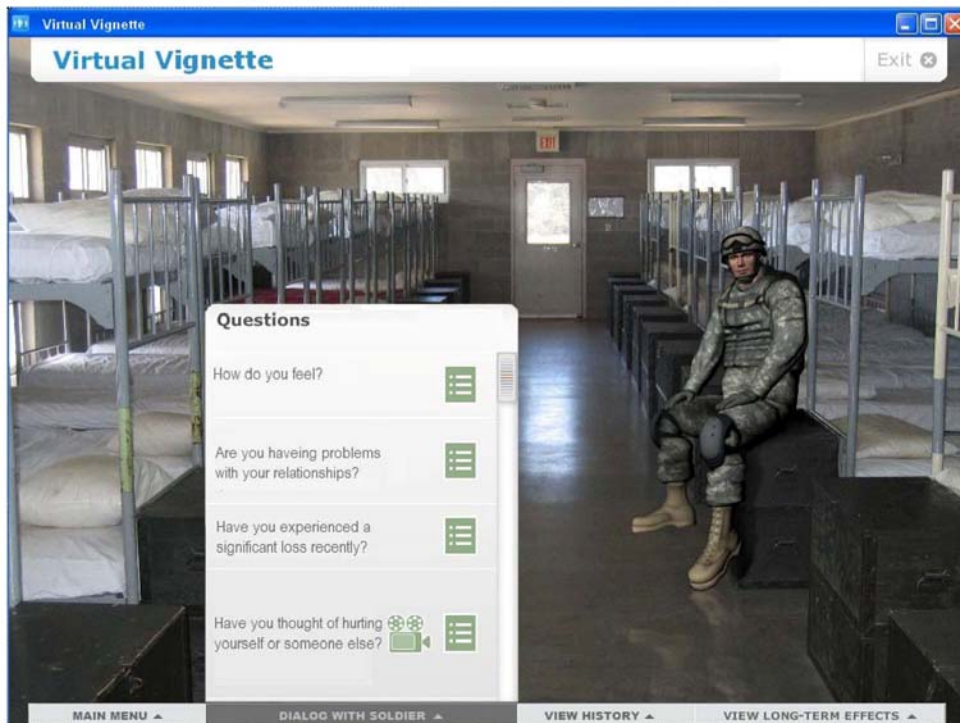
Virtual Vignette

- **Look And Feel**
  - 3D characters and environment need to be relevant
  - Animations and movements
  - Realistic sound
- **Consequences**
  - both positive and negative
  - detail alternative outcomes between extremes
- **Interaction**
  - needs to challenge the player
  - not everyone is a gamer

Alternative Applications

- Cultural awareness and sensitivity
- Army Combat Lifesaver Course
- Leadership Training
- Sexual Harassment Prevention
- Only limited by our imagination

Conclusion

- Soft skills training will continue to be important
- Games that leverage 3D modeling and simulation are excellent for realistic role play
- Computer-based training alleviates the discomfort of role-playing exercises
- Students can go through training at their own pace
  - Take responsibility for their own development.
- Training is only as good as its content

Questions/Comments

943
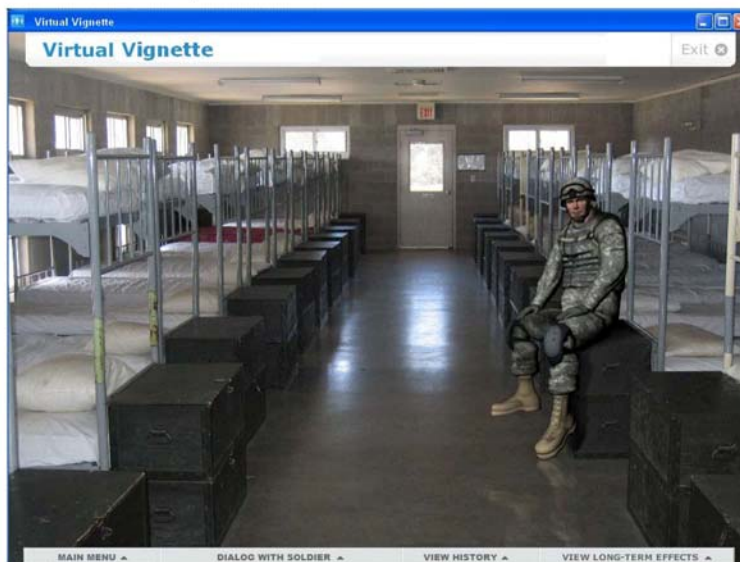
7.4　Defining and Leveraging Game Qualities for Serious Games

Defining and Leveraging Game Qualities for Serious Games

Michael W. Martin & Yuzhong Shen
Old Dominion University
Department of Modeling, Simulation, and Visualization Engineering
mmart081@odu.edu yshen@odu.edu

Abstract. Serious games can and should leverage the unique qualities of video games to effectively deliver educational experiences for the learners. However, leveraging these qualities is incumbent upon understanding what these unique 'game' qualities are, and how they can facilitate the learning process. This paper presents an examination of the meaning of the term 'game', as it applies to both serious games and digital entertainment games. Through the examination of counter examples, we derive three game characteristics; games are self contained, provide a variety of meaningful choices, and are intrinsically compelling. We also discuss the theoretical educational foundations which support the application of these 'game qualities' to educational endeavors. This paper concludes with a presentation of results achieved through the application of these qualities and the applicable educational theories to teach learners about the periodic table of elements via a serious game developed by the authors.

1.0　INTRODUCTION

The term "serious games" is somewhat open-ended, and even people who work with serious games have a hard time agreeing upon its exact meaning. This work is not presented with the intent of settling the debate on the meaning of the term, but simply as an effort to add to the greater body of discussion. Additionally, this work is presented because the authors believe that this approach to looking at serious games can be useful, not only in the academic sense of defining the term, but in actual development of serious games, as well.

To that end, this paper will first present a discussion of both common definitions of "games", and "serious games", followed by an effort to identify some of their salient characteristics to clarify the definitions. Then, the paper will discuss the development and deployment of a serious game that was created by the authors, based on these concepts.

2.0　BODY

2.1　Definitions for Game and Video Game

Mirriam-Webster's online dictionary provides several definitions of the term "Game" which are relevant to the discussion:

"3a(1) : a physical or mental competition conducted according to rules with the participants in direct opposition to each other…3a(5) : the manner of playing in a contest… 3c(2) : any activity undertaken or regarded as a contest involving rivalry, strategy, or struggle" [1].

There are a number of more in depth works on the subject of games, and they all tend towards these same basic definitional components. In the 1961 book Man, Play and Game, Roger Caillois, described games as being activities that are fun, distinct, uncertain, non-productive, rule-driven, and fictitious [2]. Clark Abt, in 1970, described games as an "activity among two or more independent decision-makers seeking to achieve their objectives in some limiting context" [3]. More recently, Katie Salen and Eric Zimmerman describe "artificial conflict, defined by rules, that results in a quantifiable outcome" [4]. While each definition has its own nuances, in general, they fairly closely resemble the dictionary definitions found above. Those common recurrent components found in these definitions can be summed up as participants, goals, rules, and challenges. The participants have goals within the game, which they try to achieve via a set of

rules. The rules define the participants' interactions, and in the application of these rules, the participants try to overcome challenges in order to achieve their goals.

The definitions apply to the general term game. However, the focus of this paper will be on computer based games. This is simply the stipulation that the game in question takes part on a computer. Michael Zyda, Director of the University of Southern California Viterbi School of Engineering's GamePipe Laboratory builds a similar dictionary based definition proposing that a video game is:

*"A mental contest, played **with a computer** according to certain rules for amusement, recreation, or winning a stake."* [5] (emphasis added)

2.2 Definitions for Serious Game and Definitional Pitfalls

Having established a basic definition for "game", the next step is to determine the meaning of "serious games".

In understanding the nature of a Serious Game, as defined in this paper, it is helpful to consider that game players learn something in every game. If this were not true, then a player's performance on a game would be the same the first they played as the last. However, this is not so, and players improve their performance by learning the mechanics which govern the game.

Most of the time, what the players learn is useless in the real world. Players of Nintendo's famous game Super Mario Bros. [6] learn that mushrooms are evil, but that the player can jump on them to kill them. They also learn that they can jump on turtles and use their shells to kill other enemies, like the mushrooms. Learning these aspects of the game helps the players perform better, but this knowledge usually transfers very poorly to the real world.

In a Serious Game, the knowledge that the user learns within the game is transferable

to the real world. The game mechanics and content have a sufficient degree of fidelity with real life mechanics and subject. When the player learns something in the game, that knowledge is transferable, within reason, to the real world as well as the game.

A number of established serious game authorities define serious games in terms of their intent. One of the most succinct definitions for serious games is found on the Michigan State University Serious Game Design Program webpage:

"Serious games are games with purpose beyond just providing entertainment." [7]

Put another way, "Serious Games" are game played for a serious intent or reason. Zyda proposes a similar definition, though with more detail as to the nature of the purpose:

"Serious game: a mental contest, played with a computer in accordance with specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives." [5]

Building upon the definitions above, a Serious Game might be "an activity consisting of participants, goals, rules, and challenges with a purpose beyond entertainment."

These definitions for "Games" and "Serious Games" are useful for grounding the baseline understanding of these concepts. However, as analytic propositions, they may also be seen as being overly inclusive. They encompass the concepts in widely applicable terms, and as a result, may also be valid for things which are not games. For example, a game of football has participants, goals, rules, and challenges. However, arguably the chore of mowing the lawn has all of these components as well. This does not necessarily invalidate the definition as much as it highlights that this definition does not provide sufficient

945

information to distinguish games from non-games.

Normally, a traditional game maker will not suffer from such a definitional dilemma. Arguably, the one and only value metric used for entertainment games is their "fun" or entertainment value. Serious games add a competing metric: the ulterior purpose beyond entertainment. This additional aspect can confound the value assessment.

With these competing values, the field of serious games can become quite confusing. Organizations of all sizes and types attempt to leverage viable serious games to achieve their goals. Ambiguity in what constitutes a game allows the term to be applied to a variety of efforts which are perhaps better identified as non-games.

As with the "lawn mowing" example above, the serious game definition might encompass the use of a budgeting program. Budgeting program have participants, goals, rules, challenges, and a purpose beyond entertainment. Such a definition might be technically correct, but ultimately it is unhelpful in trying to determine how to best make use of serious games. Clearly, if someone expected a game, and was given a budget to balance, they might be dissatisfied. Such confusion could contribute to disillusionment on the part of serious game customers, users, and developers.

2.3 Characterization of Serious Games

To better understand and apply the term "serious games" it is helpful to identify what aspects of games make them unique from non-game activities. There are two serious game counter examples which help provide additional clarity through contrast. The first counter example is generic computer based training, and the second is computerized sand box training.

2.3.1 Computer Based Training

There are a number of computer based learning or training application which have

seized upon the "Serious Games" trend. They apply game-like facades to traditional computer education activities and declared them to be games. This misappropriation of the term "Serious Game" is compounded by the fact that, as with the examples above, many traditional computer based learning activities fulfill the definition of game provided above.

A quiz or test, common to most educational systems, is a prime example of an activity that can satisfy the majority of the game attributes listed above. A quiz has rules (fill in the blank, multiple choice, etc.), it has goals (to score the highest possible score) and conflict (the difficulty of the questions). It is a difficult to argue that quizzes or tests are fun, but some educational applications decorate test-like experiences with "fun" veneers, such as cute graphics and silly sound effects and label them serious games. Two examples of this are Grammar Gorillas [8], and Snork's Long Division [9]. Grammar Gorillas has the player select specific parts from a sentence. Snork's Long Division has the player simply perform long division. Both are self declared serious games, but clearly fall short of what would normally be considered a game.

In examining why these activities do not seem like game, one finds that at their core, these programs don't provide the learner with any meaningful choices. The user simply answers the question correctly or fails the question. While the students choose how to answer the question, these choices are not meaningful in that there is only one correct answer, with no viable alternative. Raph Koster identified this pitfall in A Theory of Fun and Learning[10]. Using Tic-Tac-Toe as an example, he illustrates that the game ceases to be a game when the players learn that they have no choices. At that point, the game becomes a simple drill in the rote application of logic.

Further, by not providing meaningful alternatives, these software examples limit

946

the potential for creating a psychosocial moratorium, as described by James Paul Gee. The psychosocial moratorium, a phrase which Gee adapts from psychologist Eric Erikson, is "a learning space in which the learner can take risks where real world consequences are lowered," [11]. The lack of choice collapses the exploratory learning space, and deprives the player of the opportunity to reflect upon the information being presented and applied. David Shaffer argues that games have the potential to allow a more "authentic" method of learning than traditional schooling techniques because games set the stage for learners to think not only about what the right answer is, but also how they know an answer is right and what is the process by which they arrive at that answer [12]. Quizzes, which limit the "players" interaction with their world (the quiz) to a single correct answer, with all other options being incorrect, do not engender the same degree of introspection that a wider array of viable choices might.

In order to provide that sense of freedom that fosters exploration and, as Gee and Shaffer suggest, learning, games need to adhere, to an extent, to game designer Sid Meier's dictum that games are a "series of interesting decisions" [13]. Meier is renowned for his complex and involved strategy games, like the Civilization and Tycoon series, where each decision can have profound consequences. However, games can provide the player with meaningful choices without having to resort to such elaborate depth. The term meaningful, in this sense, can refer not to a profound consequence for the available choices, but rather, to having any consequence at all.

Some activities, like the above mentioned quizzes, offer no meaningful choices – the player's only option is to answer the question correctly. At the other extreme, some games offer a wide array of choices that effectively have no consequence. For example, many games allow the player to visually customize their character. A player

can spend hours adjusting the eye color, cheek bone height, hair style, etcetera, but ultimately none of that has any effect on the game play. In contrast, a meaningful choice allows the player to select between viable alternatives with concrete consequences, without any clear optimal choice, thereby allowing the player to freely explore the conceptual space created by the game.

In order to differentiate games from quizzes, the first necessary characteristic of games (and serious games) is that they provide the user with an array of meaningful choices.

2.3.2 Sandbox Experiences
The next counter example of non-games is the open sandbox experience. These types of programs are commonly used in military and police training. They are programs that create worlds in which typically large groups of individuals engage in educational or learning scenarios.

Such programs have been in use for several decades. One of the first was SIMNET, which was developed and deployed in the early to mid 1980's [14]. For many years, such programs have used specialized hardware and software, which was inexpensive in comparison to the resource cost of conducting the training using real world equipment and locations.

Recently, however, the commercial entertainment software industry has proven that high quality experiences can be delivered on commercially available hardware and software. And these experiences can be delivered for even less than the cost of simulation using specialized hardware and software.

One recent development in this field is the US Army's adoption of a program called Virtual Battlespace 2 (VBS2) [15]. VBS2 is developed by Bohemia Interactive, which was previously a commercial entertainment company. Bohemia developed a great number of games including a tactical virtual reality shooting game called Operation Flashpoint. Unlike earlier training

simulations using specialized software and hardware, VBS2 is based on the Operation Flashpoint game, and is designed to run on common desktops and laptops. Because the program shares software technology with Bohemia Interactive's new game, Operation Flashpoint II, there is a natural tendency to call VBS2 a serious game.

However, VBS2 is not a game in the traditional sense. It is designed to host large numbers of networked users in a virtual environment. In game terms, it would be considered a large multiplayer game. Unlike games, however, this experience is not a closed system. The multiplayer sessions don't have scores, or objectives defined within the game. They lack the framework associated with even freeform multiplayer games. Instead, VBS2 sessions are designed to be administered by teams of instructors, who then take on the role of giving the players objectives, assessing their performance and providing them with feedback. In commercial terms, this would be considered an open world experience, similar to Second Life. Players can interact, but there is not game structure to support the interactions.

In commercial games, a player can play the game entirely by himself or herself. This concept even applies to multiplayer games. A player sitting down with the commercial version of the game Operation Flashpoint can play the game by themselves, even when engaging in multiplayer sessions. They can join a multiplayer session with no prior coordination, play the game, and then quit when they desire. Most importantly, the game provides internal feedback loops, via scores and other performance measures, to let the player know how well they performed. Granted, in games which are exclusively multiplayer in nature, this assumes a robust network structure with available sessions, but given that assumption, there is no overhead to playing the game other than the player and the game itself. VBS2 and other serious games

like it lack this fundamental game characteristic.

If VBS2 did have those qualities, then users could train at their own pace, learning the materials as appropriate to their individual skills. Targeting training at the individual level could greatly increase the user engagement and ultimately the effectiveness of the training. No longer would quick learners be held up by the slow members of traditional classes or training groups, nor would the slow learners be dragged along faster than they can assimilate the material. If the instructional framework were properly embedded, then the serious game would be a self encapsulated experience, just as commercial video games are.

Therefore, the second quality that a serious game should have is that it is a self encapsulated experience. Given that this discussion has been centered on video games, it is worth noting that this quality would not exclude serious game that are not played on a computer. Because of their automation advantages, computers facilitate this characteristic. However, it is possible to have board or card and paper games, or even athletic games that are playable solo as well as with other people. Granted, these games are rare, and the difficulty in creating them is much higher than traditional board games, but it is not impossible to conceive of.

2.4 The Issue of Fun
A few of the definitions of games, such those of Caillois and Zyda, also refer to entertainment or fun. Many of the definitions omit such concepts, perhaps due to their highly subjective nature. However, as a basic metric of value for games, it is undeniable that these are fundamental aspects of the concept.

Customers pay money to play games not because the games provide some sort of reward, but because the game experience, itself, is rewarding. The intrinsic value of

948

the game experience outweighs any extrinsic benefit bestowed by playing the game. The games, without any consideration to outside benefit, are intrinsically compelling.

"Intrinsically Compelling" encompasses the concepts of "fun", but it also makes room for other aspects, such as the satisfaction of overcoming a challenge, or earning a reward, which may not be entirely fun. For example, many Massively Multiplayer Online Role-Playing Games (MMORPGs), like World of Warcraft [16], routinely include what player communities often refer to as "grinding". World of Warcraft is a game in which players try to improve their character by accomplishing various tasks and thereby gaining experience points. Grinding is a low risk way to gain experience points. It is a repetitive act, which is widely described in negative terms, often involving tediously killing large numbers of weaker enemies which pose little threat. Killing these weak enemies might be a boring and repetitive task, but it gains points which grant the player some form of in-game reward, such as a more powerful character. It is a significant component of this game genre which is not normally deemed as fun. Yet it is intrinsically compelling.

Regardless, however, of what creates that intrinsic compulsion, be it fun or the feeling of achievement, or some other factor, this intrinsic compulsion and educational value do not necessarily compete. Raph Koester proposes that fun is, in fact, the brain's reaction to learning [10]. Keeping in mind that paradigm that players are always learning when they play, it follows that the two are closely interrelated.

Even from a practical stand point, a player who feels intrinsically compelled by a serious game is more likely to engage with the game, and therefore, assuming the serious game is designed well, more likely to achieve the desired "non-entertainment" purpose.

Thus, the third quality of a serious game is that it should be intrinsically compelling.

3.0 DISCUSSION

Based on these concepts, the authors have been developing a Serious Game entitled Elemental Solitaire™. The effort to develop the game began simply enough with the idea to build a combination of popular classic card game mechanics, such as solitaire, and the periodic table.

Because of the nature of this project, it was more likely to fall prey to the "Computer Based Training" pitfall than the "Sandbox Experience" pitfall. The mechanics of the game lent themselves well to creating an encapsulated game-play experience. However, avoiding the "quiz-like" danger required more effort.
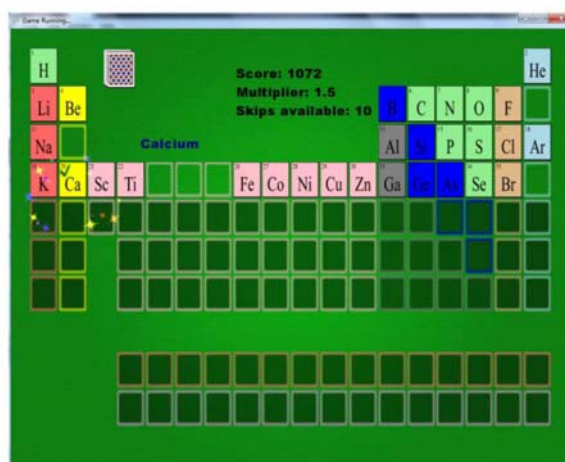

Fig 1: Elemental Solitaire Game Screen

In early iterations of the game, the simplest designs precluded any such meaningful choice. As the players were given elements, they either placed them correctly, or they were penalized for failing to do so. Though the program had a graphic interface and the elements had an appearance of playing cards, there was no conceptual space to explore. The program simply presented the user with quiz questions disguised in a graphical form. Though different in execution, this program, in spirit, resembled many of the aforementioned online quizzes. In order to add more

decision space to explore, three items were added to provide the user with more meaningful choices.

First, players were given ten 'skips', so if they wanted to delay placing an element, they had the ability to do so ten times, without penalty. This gives the player a small degree of control in choosing whether to place an element or not. It also adds a measure of strategic depth, as players must ration their skip choices, and forcing them to weigh the risk of skipping a present element versus the need to be able to skip an element later on.

Second, when placing an element, the players are given a countdown timer. As the time passes, hints as to the correct element position on the table are automatically given, including the family color, the row, and ultimately the actual element position. The balance is that the score the player receives for placing a card decreases as more hints are provided, until, ultimately, no points are awarded if the card's correct position is shown to the player. A player can also choose to capitalize on this, and if they are willing to take a lower reward, can even manually advance the timer to display the next hint, without having to wait. Again, this mechanic allows the player decide how to balance risk and reward. As risk diminishes, so does the reward.

Lastly, as mentioned above, an abstract scoring mechanic was added to the game. This system rewards the player with a specified amount of points for each element correctly placed, and deducts an amount for incorrect placements. The amounted award for correct placement is inversely proportional to the time taken to place the element. Additionally, not only do players get rewarded for correct answers, but they also increase a score multiplier, which links their decisions on how they choose to answer questions with the rewards they can receive on future correct answers. As they score better, their multiplier increases,

allowing them to score even higher on subsequent correct placements. This factor also adds significance to the decision the player must make in balancing the risk and reward of placing the elements.

These three mechanics of the game combine to create a decision space for the player to explore, and extend the game space beyond simply entering a right or wrong answer. These "choice creating factors" do not possess the depth nor the scale of the types of decision often made in a highly strategic game, such as Sid Meier's Civilization games, but they do create a small space for the learner to explore. These factors set the conditions for the psychosocial moratorium, and presumably improve the facilitation of learning. With an array of possible actions, the player can play how they like, and, in the words of game designer Chris Crawford, imprint their own personality on the game [17]. Fig. 2 shows a screenshot from a game in play, with a hint showing the family and row of the element.
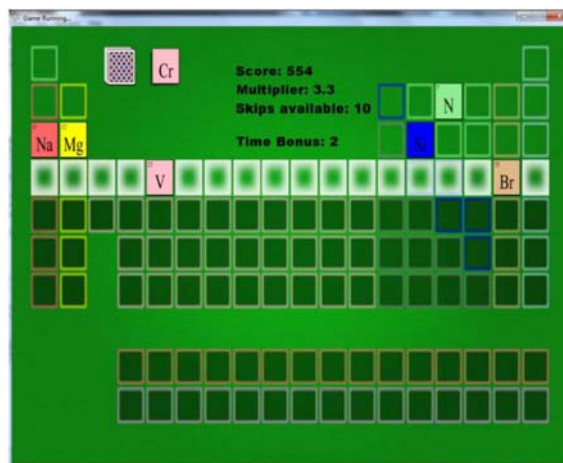

Fig 2. Game Screen with Hints

Additional discussion of the development of Elemental Solitaire™ can be found in the paper *Differentiating Between Serious Games and Computer Aided Instruction* [18].

950

4.0 CONCLUSION

While the traditional approach to understanding something is to examine what it is, this paper presents a characterization of Serious Games based on what they are not. From this examination of counter-examples, we have derived three qualities which a Serious Game should possess:

1. They provide meaningful choices to the user.
2. They are self encapsulated experiences.
3. They are intrinsically compelling.

Elemental Soliatire™ is an example of how these characteristics can be used to guide the development of a Serious Game. This program is going to be used as a test platform to assess the effectiveness of these and other game design principles in developing Serious Games.

5.0 REFERENCES

1. Merriam-Webster Inc. *Game - Definition and More from the Free Merriam-Webster Dictionary*. Mirriam-Webster Online 2010 [cited 2010 15 July]; Available from: http://www.merriam-webster.com/dictionary/game.
2. Caillois, R., *Man, Play and Games*. 1961, New York: Free Press of Glencow.
3. Abt, C.C., *Serious games*. 1970, New York,: Viking Press. xvi, 176 p.
4. Salen, K. and E. Zimmerman, *Rules of Play: Game Design Fundamentals*. 2003, Cambridge: MIT Press.
5. Zyda, M., *From Visual Simulation to Virtual Reality to Games*. Computer, 2005(September 2005): p. 8.
6. Miyamoto, S., T. Tezuka, and H. Yamauchi, *Super Mario Bros.* 1985, Nintendo Co., Ltd.
7. Michigan State University. *MSU Serious Games Design Program*. [cited 2010 July 15]; Available from: http://seriousgames.msu.edu/.
8. *funbrain.com Grammar Gorillas*. 2010 [cited 2010 20 Feb]; Available from: http://www.funbrain.com/grammar/index.html.
9. *Snork's Long Division Game - KidsNumbers.com*. 2010 [cited 2010 20 Feb]; Available from: http://www.kidsnumbers.com/long-division.php.
10. Koster, R., *A Theory of Fun for Game Design*. 2005, Scottsdale, AZ: Paraglyph Press. x, 244 p.
11. Gee, J.P., *What video games have to teach us about learning and literacy*. Rev. and updated ed. 2007, New York: Palgrave Macmillan. 249 p.
12. Shaffer, D., *In Praise of Epistemology*, in *The Design and Use of Simulation Computer Games in Education*, B. Shelton and D.A. Wiley, Editors. 2007, Sense Publishers: Rotterdam. p. 7 - 27.
13. Camargo, C. (2006) *Interesting Complexity: Sid Meier and the Secrets of Game Design*. Crossroads: The ACM Student Magazine.
14. Lenoir, T. and H. Lowood (2003) *Theaters of War: The Military-Entertainment Complex*.
15. Bohemia Interactive Simulations. *BISimulations*. 2010 [cited 2010 23 July]; Available from: http://www.bisimulations.com/.
16. Blizzard Entertainment Inc. *World of Warcraft Community Site*. 2010 [cited 2010 27 July]; Available from: http://www.worldofwarcraft.com/index.xml.
17. Crawford, C., *The Art of Computer Game Design*. 1984, Berkeley, Calif.: Osborne/McGraw-Hill. xiv, 113 p.
18. Martin, M.W. and Y. Shen, *Differentiating Between Serious Games and Computer Aided Instruction*, in *Virginia Modeling, Analysis and Simulation Center 2010 Student Capstone Conference*. 2010: Suffolk, VA.

7.5 GIS Data Based Automatic High-Fidelity 3D Road Network Modeling

GIS Data Based Automatic High-Fidelity 3D Road Network Modeling

Jie Wang and Yuzhong Shen
Department of Modeling, Simulation, and Visualization Engineering
Old Dominion University
jwang016@odu.edu, yshen@odu.edu

Abstract. 3D road models are widely used in many computer applications such as racing games and driving simulations. However, almost all high-fidelity 3D road models were generated manually by professional artists at the expense of intensive labor. There are very few existing methods for automatically generating 3D high-fidelity road networks, especially those existing in the real world. This paper presents a novel approach that can automatically produce 3D high-fidelity road network models from real 2D road GIS data that mainly contain road centerline information. The proposed method first builds parametric representations of the road centerlines through segmentation and fitting. A basic set of civil engineering rules (e.g., cross slope, superelevation, grade) for road design are then selected in order to generate realistic road surfaces in compliance with these rules. While the proposed method applies to any types of roads, this paper mainly addresses automatic generation of complex traffic interchanges and intersections, which are the most sophisticated elements in the road networks.

1.0 INTRODUCTION

Road networks are critical infrastructures of human civilization and probably the most important means of transportation. With advances in computing technologies, 2D and 3D road models have been employed in many applications, such as computer games and virtual environment construction. Roads are complex 3D structures and traditional road models were generated by professional artists manually using modeling software tools such as Maya and 3ds Max. This approach requires both highly specialized and sophisticated skills and massive manual labor. Procedural modeling based automatic road generation methods [1-6] create road models using specially designed computer algorithms or procedures and they can dramatically reduce or eliminate the amount of manual editing needed for road modeling. However, most existing procedural road modeling methods aimed at the visual effects of the generated roads, not the geometric or architectural fidelity that mostly determines the driving experience.

Geographic information systems (GIS) are computer-based systems widely used to store, manipulate, display, and analyze geographic information in many fields. With the rapid development and widespread use of GIS techniques, vast GIS data are captured through various digital data collection methods such as remote sensing via cameras, digital scanners and LIDAR.

As one kind of them, road GIS data which record the information of the real road network in the best possible way have also been used in many applications and facilitate our lives greatly, e.g., the automotive navigation system. Since real road GIS data contain road network information which is indispensable for some applications, especially for transportation, homeland security and defense applications, it will significantly reduce both the time and lab cost if 3D road network can be modeled from road GIS data directly and automatically. However, there are very few existing methods for GIS based automatic 3D high-fidelity road networks generation. Most GIS based modeling work and software focus on buildings [7, 8], vegetation and rural landscape visualization [9-11] rather than roads.

Therefore, a method that can automatically produce 3D high-fidelity road network models from real road GIS data will greatly benefit numerous applications involving road networks. This paper addresses this problem by proposing a novel method which is used in an ongoing project to automatically generate 3D high-fidelity road network models from existing road GIS data in compliance with a set of selected civil engineering rules. The proposed method consists of several steps, including road GIS data preprocessing, road representation parameterization, civil engineering rules based road surface

modeling, and intersection and interchange generation.

This paper is organized as follows: section 2 describes each step of the proposed method in detail; section 3 discusses the application fields and advantages of the proposed method, and finally this paper is concluded in section 4.

2.0 BODY

In this section, the whole road network modeling method is described using a road network which contains several road segments, road intersections and a traffic interchange as shown in Figure 1(a) as an example.

2.1 Road GIS Data Preprocessing

2.1.1 GIS Data Import

Although various data formats have been developed for GIS applications, such as GML (Geographic Markup Language) and TIGER (reference file format), shapefile format is the most widely used format and it is utilized in this research. A typical shapefile consists of a main file, an index file, a dBase file and a projection file. Among them, the first three files define the geometry and attributes in a shapefile. Three kinds of shape types are used to represent geometric shape features, which are point, polyline and polygon [12]. The main file contains the primary reference data with one record per shape feature; the index file stores the position and content length for each record in the main file; the dBase file contains the feature attribute for each record in the main file. A library named shapelib [13] is used in this project to read data from shapefile.

2.1.2 Road Network Topology Extraction

Since in existing road GIS data, roads are represented as 2D (latitude and longitude) centerlines in the form of polylines, i.e., connected line segments that consist of consecutive but discrete road centerline points as shown in Figure 1(b), road

network topology or connectivity information is not explicitly represented, that is, it is difficult to determine if two polylines are connected without explicitly comparing the points that compose the polylines. To expedite the automatic road generation process and to facilitate road navigation, an explicit representation of the road network or topology is necessary. In this research, a method proposed in [14] was employed to extract the topology information of the road network from raw road GIS data. The output of this method is a road network represented as a graph composed by road intersections (nodes) and road polylines (links) as shown in Figure 1(c).

2.1.3 Road Network Simplification

The raw road GIS data contain redundant representations for road links that have multiple names. That is, the same road with two or more different names is stored as two or more independent roads in the raw road GIS data. While this redundancy might be useful for other purposes, the same physical road should have only one 3D representation in this project. Hence road links with the same physical positions but different names are combined into one road link with multiple names, eliminating the redundancy of the network representation.

2.1.4 Road Classification

In order to obtain more information about road network, road links are classified into different categories according to their names which usually can be obtained from the road GIS data. Some keywords are identified to do the classification and totally six road categories are defined in this research, which are highway, local, ramp, bridge, tunnel, and unknown. For instance, keywords used to indentify local roads are a set including "LN", "ST", "RD", "DR", "AVE", "BLVD" and "PKWY". Keywords used for the road classification in this project are listed as follows:

- Highway:"I-" (e.g. I-64), "HWY".
- Local: " LN", " ST", " RD", " DR", "AVE", "BLVD", "PKWY".
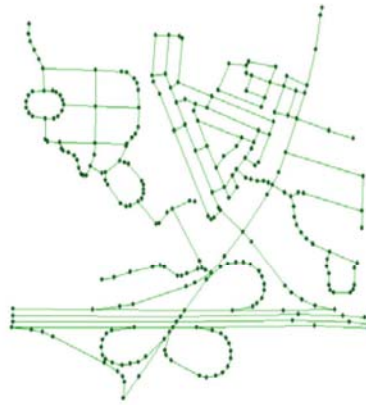
953

Figure 1 A road network example which contains several road segments, road intersections and a traffic interchange. (a) Satellite photo around VA-403 Norfolk, VA from USGS. (b) Road centerline and discrete road centerline points from road GIS data shown in ArcGIS for the same area. (c) Road network after topology extraction with extracted nodes indicated by red points.

- Ramp: "RAMP", 0-9+A-Z (e.g. 14B).
- Bridge: "BRIDGE".
- Tunnel: "TUNNEL".
- Unknown: "NULL".

2.2 Parametric Representation

2.2.1 Parameterization
As mentioned before, the road GIS data only contain discrete points of the road network in the form of polylines. This discrete representation has several drawbacks, with the most serious one being not supporting arbitrary resolution of representations. The proposed method converts the original discrete representation of the road network into parametric representations that have the advantages of supporting arbitrary resolution of levels of details (LOD) and reduced memory usage. Two standard parametric forms are defined as Straight Lines and Circular Curves.

- Straight Line (Line): A straight line connecting two points and specified by its ID, start point, end point, start side vector and end side vector.
- Circular Curve (Curve): Part of a circle and specified by its ID, position of center point, radius, start point, end point, start side vector, end side vector, start angle, end

angle and direction (clockwise / anticlockwise).

Among these parameters, side vectors indicate the extension direction of road surface. This representation of road centerline data has several advantages. First of all, it is relatively simple and easy to understand and implement. More importantly, it is well-suited to apply civil engineering principles, especially for superelevation generation of curve road for which the center point and radius are required.

2.2.2 Segmenting and Fitting
In order to divide the road links into segments that can be represented in standard segment forms, three types of critical points are identified to segment the road polylines: acute turn, s-turn and turn start/end [15] based on their geometric features. After the critical points are identified, road polyline links are partitioned into a set of segments that are groups of discrete points ready for segment fitting to obtain their appropriate analytic representations. Then least square methods [15] are employed to fit these road segments into straight lines and circular curves optimally. Based on the parametric

954

representation of road network, different levels of details can be employed to generate the road model.

2.3 Civil Engineering Rule Based Road Surface Modeling

Since the road surface is the most significant feature of a road and its quality dramatically affects our driving experience, principles and rules on road surface design are of ultimate importance. To model this most critical component of roads realistically, namely, road surface, a basic set of civil engineering rules for road design are selected, including design speed, cross slope, superelevation, grade, etc. The road surfaces will be generated in compliance with these civil engineering rules. Among all of them, normal cross slope for most road surfaces and superelevation for some curved road surfaces are two major factors directly determining the main shape of the road surfaces. Besides these, method for pavement modeling and rendering is also discussed in this section.

2.3.1 Normal Cross Slope

Sloping on roadway cross section is employed to meet the drainage needs and direct water off the traveled way to facilitate road users and reduce accident potential. According to American Association of State Highway and Transportation Officials (AASHTO) standards [16], a plane model with a peak in the middle and a 2% cross slope downward toward both edges is preferred in this research for normal road surface modeling as shown in Figure 2.

2.3.2 Superelevation

The superelevation of a curved road segment is used to balance the centrifugal force that moves the vehicle traveling on this curve outward with gravity and side friction according to the laws of mechanics. Based on civil engineering rules [16], the superelevation rate can be determined based on the side friction factor, road radius and vehicle speed and in this research, the standard superelevation rates suggested by [16] are used for different range of curve

radius, road types and surface conditions. The centerline of the roadbed is used as the axis of rotation for superelevation and an illustration of a superelevation transition from normal crown to full superelevation can be found in Figure 2.
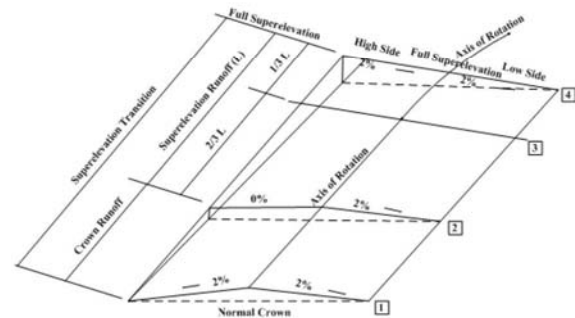

Figure 2 An illustration of a normal crown to full superelevation transition

2.3.3 Pavement

Road surfaces contain different types of pavements such as asphalt and concrete. Several methods can be used to render the road surface (pavements), which are texture mapping, programmable pixel shaders, and programmable vertex shaders. Besides visual simulation of the road surfaces, the geometry of the road surfaces can be further modified to reflect the variations on the road surface using programmable vertex shaders by adjusting the vertex positions. For instance, subtle variations of the road surface can be achieved by applying Perlin Noise and significant changes of the road conditions can be obtained through addition of holes to the road surface. We have generated a series of mathematical models for modeling different kinds of craters on planetary surface [17]. These craters can be further modified to simulate the wear and tear of road surface.

2.4 Intersection and Interchange

2.4.1 Intersection

The method discussed in [18] for junction synthesis was adopted, expanded and finally integrated into this research to
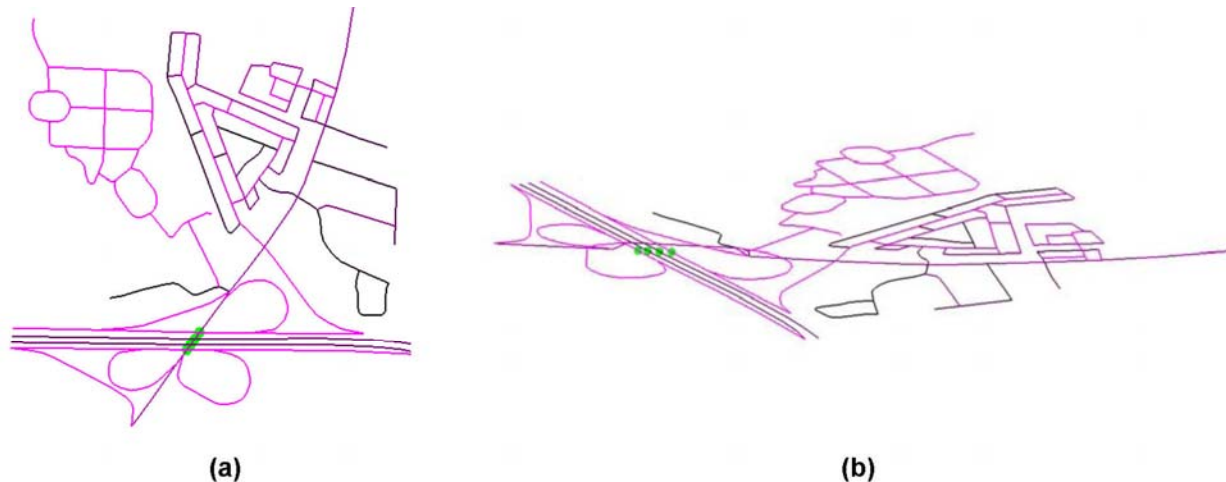
955

Figure 3 3D examined road network with generated interchange (overlapped positions are indicated by green points). (a) Top view. (b) Side view.

generate the road intersection. Compared with the original method that just considered synthesizing junctions from connected straight line road segments, our expanded method produces junctions from both straight and curved road segments with the help of side vectors. The resulting road intersection is represented by several parameters, including the position of the intersection center (node), the end point of the centerline and two boundary points for each connected segment. This process produces smooth transitions between road segments.

2.4.2 Interchange

Road interchanges are special road intersections that combine ramps and grade separations at the junction of two or more highways in order to reduce or eliminate traffic conflicts, improve driving safety and increase traffic capacity. Traffic interchange modeling is the most critical and challenging component of road network modeling due to its complexity and data deficiency that existing road GIS data are 2D and do not contain any height information (vertical position). Hence in order to generate 3D models of the interchanges, the overlapped positions in the interchanges are identified firstly and appropriate elevations are assigned to road links so that road links do not intersect or collide with each other. In

detail, since in the road GIS data, road intersections generally only occur at the end points of road links (polylines), if two road polylines contain the same point that is not the endpoint of either polyline or both polylines, this point is the location where these two road polylines overlap. And then, after overlapped position identification, the elevations of the overlapped road links are estimated based on mathematical formulations combining with observation of real traffic interchanges. Next the concept of "elevation level" is used to roughly represent the different height of road links and greater level values correspond to higher elevations. Finally after determining the elevation level for each overlapped road point, absolute elevations will be calculated based on the terrain elevation and level height, and linear interpolation is used to compute the elevations for road points located between two overlapped positions. A primary result of the interchange generation is shown in Figure 3 and reasonable relative elevation relationship is got via our proposed method.

2.5 Implementation

The proposed methods are implemented on the Microsoft XNA platform [19]. Road network models were created based on GIS data using the proposed methods and rendered with various shaders. At the same

956

time, if needed, bridges and tunnels can also be generated with just some small modifications of common road segment generation program. Furthermore, considering about the environmental terrain modeling, digital terrain elevation and satellite image can also be combined seamlessly in this system to enhance the final effects. Figure 4 shows the generated 3D model for the exampled road network.

3.0 DISCUSSION

Although civil engineering principles are emphasized in the proposed work, it is worth mentioning that we do not aim at generating a product that can be used for real road design and construction. Instead the intended audience and users of this work are professionals in the modeling and simulation industry, computer and video game industry, and other computer graphics applications that require realistic roadway models. The purpose is to provide rapid and efficient 3D road modeling for such applications that have higher requirements on high-fidelity road models, such as racing games and driving simulations. As such, not all civil engineering rules on road design will be utilized in the proposed work. It is also important to note that existing road GIS data does not provide complete information that is needed to generate 3D road models from the GIS data, for example, without elevation (height) information, it is difficult to determine the exact vertical locations of road network. In addition, some existing roads actually do not conform to the design standards (especially some old roads in urban areas). Considering all these factors, although the 3D road models to be produced by the method proposed in this paper are still reasonable approximations of real roadways and may not have the exactly the same structure of the real existing ones, they will have enough fidelity and resolution that are required by high-end modeling and simulation applications.
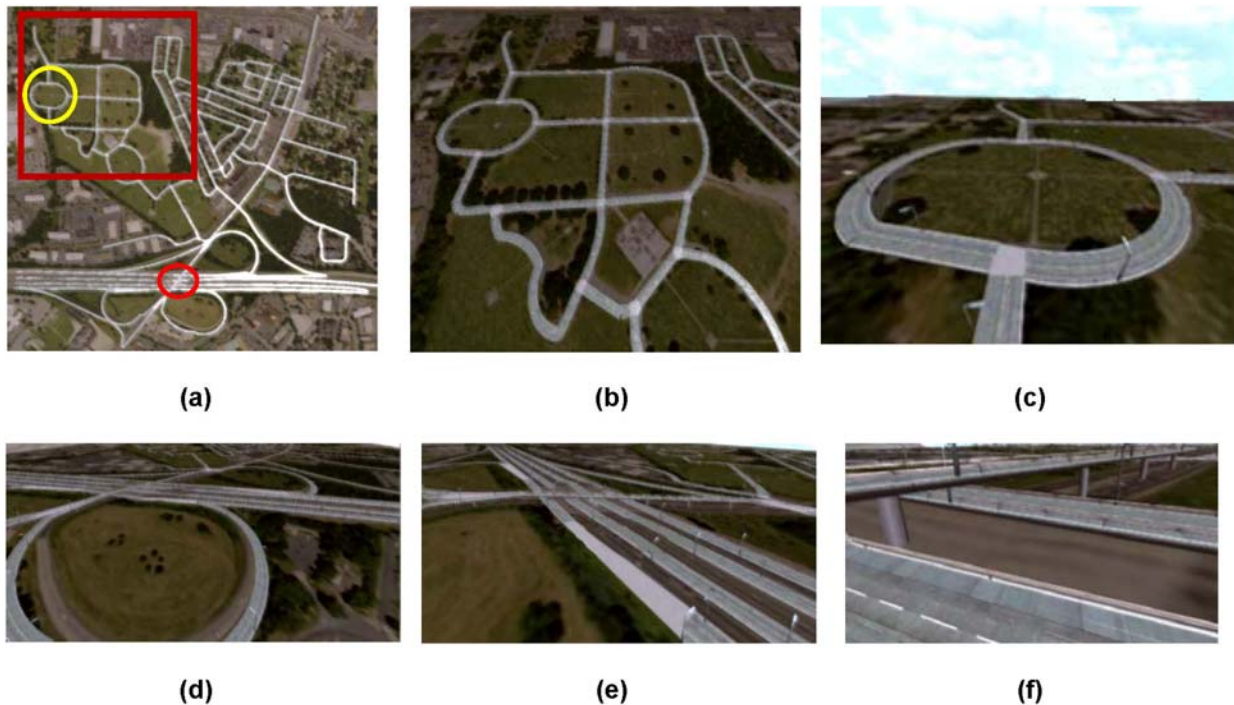

(a)　　　　　　　　(b)　　　　　　　　(c)

(d)　　　　　　　　(e)　　　　　　　　(f)

Figure 4 Experimental results (a) 3D road network generated for the exampled road network. (b) Top view of the area indicated by the red rectangle in (a). (c) Side view of the area indicated by the yellow circle in (a). (d)(e)(f) Different views of the interchange part of the generated 3D road network model.

957

4.0 CONCLUSION

In conclusion, this paper worked on the automatic generation of 3D high-fidelity road network models from real road GIS data. The proposed method can apply to the modeling of the whole road network that is critical for applications that have stringent requirements on high fidelity road network, such as driving and transportation simulation. Also with minor modification, the proposed method can be extended for other areas, such as generation of subway system based on 2D subway maps.

5.0 REFERENCES

[1] Y. I. H. Parish and P. Muller, "Procedural modeling of cities," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*: ACM, 2001.

[2] J. Sun, X. Yu, G. Baciu, and M. Green, "Template-based generation of road networks for virtual city modeling," in *Proceedings of the ACM symposium on Virtual reality software and technology* Hong Kong, China: ACM, 2002.

[3] T. Lechner, B. A. Watson, U. Wilensky, and M. Felsen, "Procedural City Modeling," in *1st Midwestern Graphics Conference*, 2003.

[4] D. F. Evans, "High-fidelity Roadway Modeling for Real-time Driving Simulation." vol. Master: The University of Iowa, 1995.

[5] P. Willemsen, J. K. Kearney, and H. Wang, "Ribbon Networks for Modeling Navigable Paths of Autonomous Agents in Virtual Environments," *IEEE Transactions on Visualization and Computer Graphics,* vol. 12, pp. 331-342, May-June 2006.

[6] G. Thomas and S. Donikian, "Modelling virtual cities dedicated to behavioural animation," *Computer Graphics Forum,* vol. 19, pp. 71-80, 2000.

[7] CommunityViz, "CommunityViz," in *http://placeways.com/*.

[8] W. Lin, "Constructing a GIS-based 3D urban model using LiDAR and aerial photographs," in *Geography*. vol. MASTER OF SCIENCE, 2005, p. 175.

[9] A. Werner, O. Deussen, J. Döllner, H. C. Hege, P. Paar, and J. Rekittke, "Lenné3D–Walking through landscape plans," *Trends in real-time landscape visualization and participation. Berlin: Wichmann,* pp. 48–59, 2005.

[10] S. Ernst, "Inside Biosphere3D," in *The free digital globe from a developer prospect.*

[11] E.-o. Software, "Vue 8 - Digital Nature," in *http://www.e-onsoftware.com/*.

[12] Wikipedia, "Shapefile," in *http://en.wikipedia.org/wiki/Shapefile*.

[13] "Shapelib," in *http://shapelib.maptools.org/*.

[14] S. K. Jakkula, Y. Shen, and J. Sokolowski, "Extraction of Road Network Topology for Transportation and GIS Applications," in *MODSIM World Conference & Expo* Virginia Beach, VA, 2007.

[15] J. Wang, O. Unal, M. Cetin, Y. Shen, and Y. Papelis, "High-Fidelity Roadway Modeling and Simulation," in *MODSIM World Conference & Expo 2009* Virginia Beach, VA, 2009.

[16] AASHTO, *A Policy on Geometric Design of Highways and Streets*. Washington, D.C, 2004.

[17] J. Wang and Y. Shen, "Synthesizing High-resolution Terrains with Craters," in *2008 Modeling and Simulation Capstone Conference* Norfolk, VA, 2008.

[18] J. Sun, G. Baciu, X. Yu, and M. Green, "IMAGE-BASED TEMPLATE GENERATION OF ROAD NETWORKS FOR VIRTUAL MAPS," *International Journal of Image and Graphics,* vol. 4, pp. 701-720, 2004.

[19] Microsoft, "XNA," in *http://www.xna.com/*.

GIS Data Based Automatic High-Fidelity 3D Road Network Modeling

Jie Wang and Yuzhong Shen

Department of Modeling, Simulation, and Visualization Engineering

Old Dominion University

MODSIM WORLD
Conference & Expo

Outline

- **Introduction**
- **Body**
  - Road GIS Data Preprocessing
  - Parametric Representation
  - Civil Engineering Rule Based Road Surface Modeling
  - Intersection and Interchange
  - Implementation
- **Discussion**
- **Conclusion**

9/2/2010

2

Introduction

- **Road is an essential feature of civilization.**
- **Road models are widely used.**
  - Computer games
  - Virtual environment construction
- **2D/3D road models were generated manually.**
  - Creator, 3DMax and Maya
  - Massive labor & skilled artists
- **Procedural modeling**
  - Virtual effects
  - Geometric or architectural fidelity

Introduction

- **GIS: Geographic Information Systems**
  - Road GIS data
- **3D road network modeled from road GIS data**
  - Directly & automatically
  - Reduce both time & labor cost.
  - Few existing methods
- **Ongoing project**
  - Generate 3D high-fidelity road network models.
  - Existing road GIS data
  - A set of selected civil engineering rules

Outline

- **Introduction**
- **Body**
  - Road GIS Data Preprocessing
  - Parametric Representation
  - Civil Engineering Rule Based Road Surface Modeling
  - Intersection and Interchange
  - Implementation
- **Discussion**
- **Conclusion**

Road GIS Data Preprocessing

- **Road GIS Data Import: Shapelib**
  - Road: polyline consisting of consecutive but discrete 2D centerline road points

- **Road Network Topology Extraction** (Jakkula, 2007)
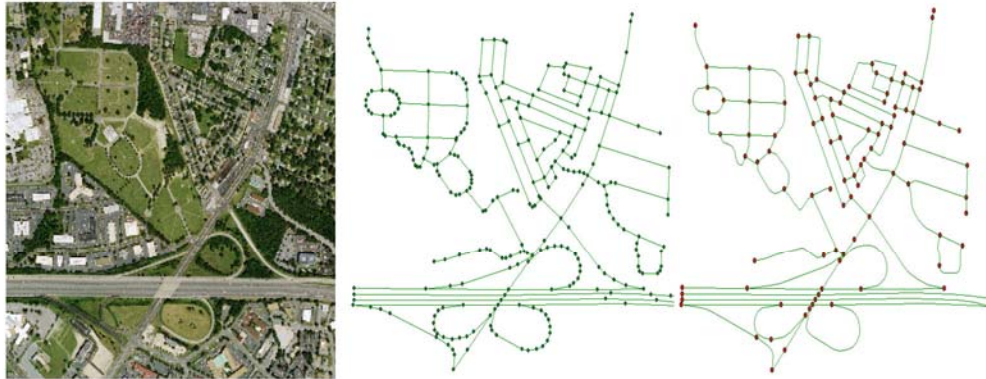
961
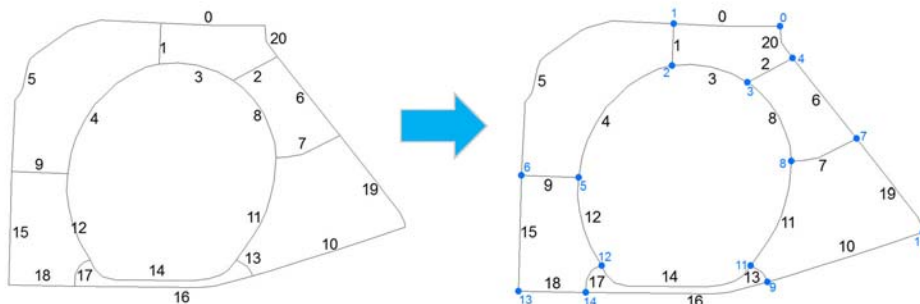
Road GIS Data Preprocessing

Figure 1. A road network example which contains several road segments, road intersections and a traffic interchange. (a) Satellite photo around VA-403 Norfolk, VA from USGS. (b) Road centerline and discrete road centerline points from road GIS data shown in ArcGIS for the same area. (c) Road network after topology extraction with extracted nodes indicated by red points.

Road GIS Data Preprocessing

- **Road Network Simplification**
  - Same physical positions but different names
- **Road Classification**
  - Highway:"I-" (e.g. I-64), "HWY".
  - Local: " LN", " ST", " RD", " DR", "AVE", "BLVD", "PKWY".
  - Ramp: "RAMP", 0-9+A-Z (e.g. 14B).
  - Bridge: "BRIDGE".
  - Tunnel: "TUNNEL".
  - Unknown: "NULL".

Parametric Representation

- **Parameterization:** converts original discrete road network into parametric representations.

- **Two Standard Forms (segment units)**

  - Straight Line (Line): A straight line connecting two points (its ID, start point, end point, start side vector and end side vector).

  - Circular Curve (Curve): Part of a circle (its ID, position of center point, radius, start point, end point, start side vector, end side vector, start angle, end angle and direction).
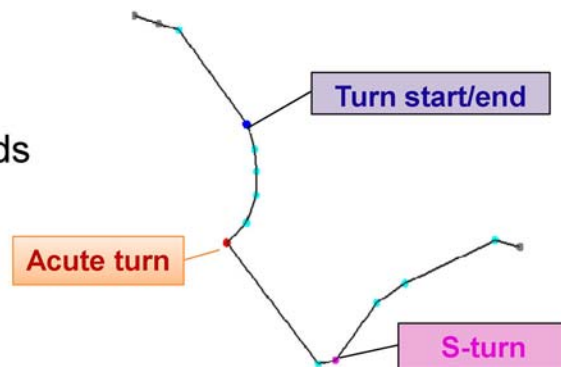
Parametric Representation

- **Segmenting**: divides road link into segments that can be represented in standard segment forms.
  - Three types of critical points are identified.
  - Acute turn, s-turn and turn start/end (Wang, 2009)

- **Fitting**
 - Least square methods
    (Wang, 2009)

Turn start/end

Acute turn

S-turn

Civil Engineering Rule Based Road Surface Modeling

- **Road Surface**
  - The most significant feature of a road
  - Its quality dramatically affects driving experience.

- **Standard**
  - American Association of State Highway and Transportation Officials (AASHTO) standard

- **A basic set of civil engineering rules**
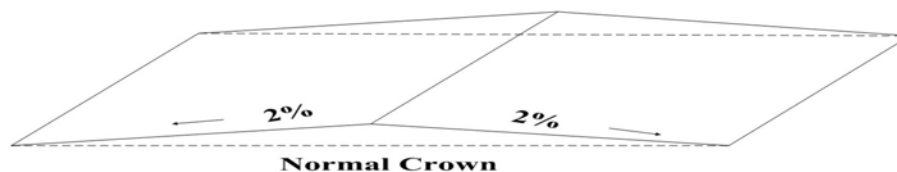 - Normal Cross Slope
  - Superelevation
  - Pavement

A POLICY ON

Geometric Design of Highways and Streets
2004
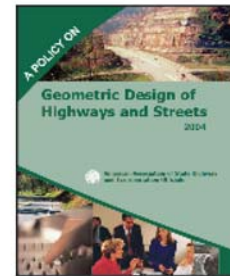
Normal Cross Slope

- **Slope**
  - Be employed to meet the drainage needs and directs water off the traveled way.

- **A plane model**
  - A peak in the middle and a 2% cross slope downward toward both edges is suggested and used.

2%   2%

Normal Crown

Superelevation

- Be used to balance the **centrifugal force** with **gravity** and **side friction** for **curve road segment**.

- **Superelevation rate**
  - The side friction factor, road radius and vehicle speed
  - Standard superelevation rates are used for different range of curve radius, road types and surface conditions.
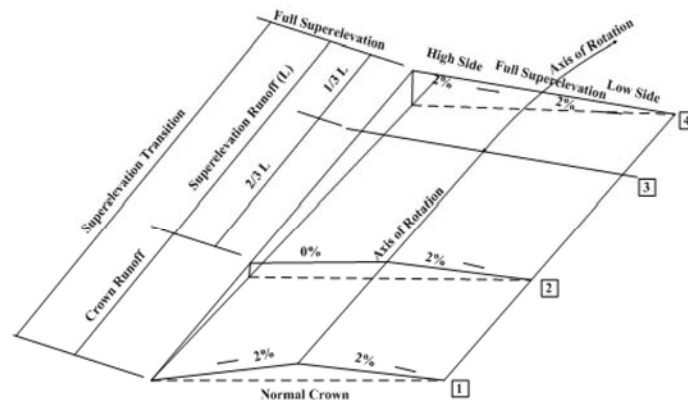
Superelevation

Figure 2. An illustration of a normal crown to full superelevation transition.

Pavement

- **Different types of pavements**: asphalt and concrete
  - Texture mapping
  - Pixel shaders & Vertex shaders
- **Variations on the road surface**
  - Subtle variations : Perlin Noise
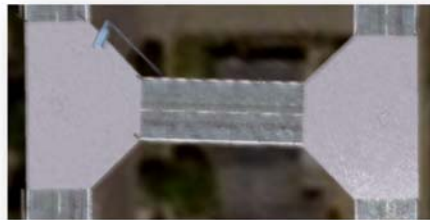 - Wear and tear : crater modeling (Wang, 2008)

9/2/2010   15

Intersection

- The method in (Sun, 2004) for junction synthesis
- Produced junctions from both straight and curved road segments.
- **The resulting road intersection**
  - The position of the intersection center (node)
  - The end point of the centerline
  - Two boundary points for each connected segment

9/2/2010   16

Interchange

- Special road intersection combining ramps and grade separations at the junction of two or more highways

- Reduce or eliminate traffic conflicts, improve driving safety and increase traffic capacity.

- The most critical and challenging component
  - Its complexity & data deficiency

9/2, 7

Interchange

- **Identify overlapped positions in the interchange**.

- **Determine elevation level for each overlapped road point**.
  - Elevation level
  - Mathematical formulations
  - Observation of real traffic interchanges

- **Calculate absolute elevation for each road point**.
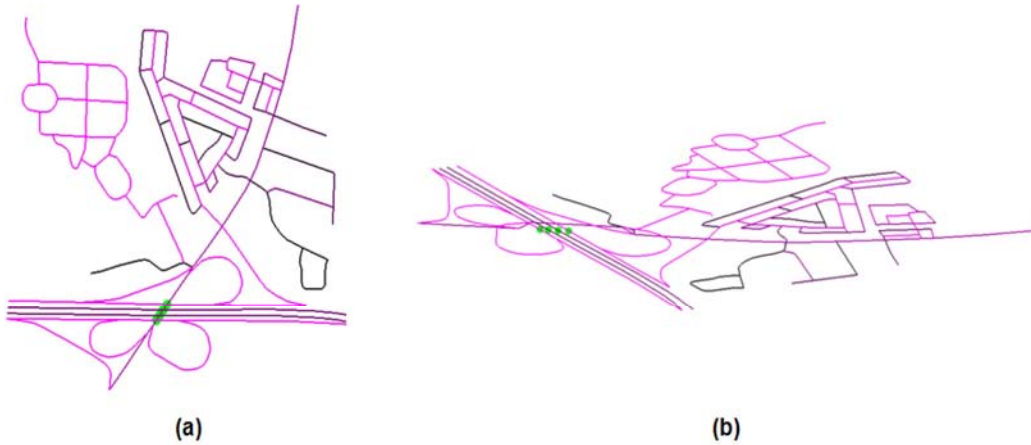 - Terrain elevation and level height
  - Linear interpolation

9/2/2010 18

967

Figure 3. 3D exampled road network with generated interchange (overlapped positions are indicated by green points). (a) Top view. (b) Side view.

Implementation

- Microsoft XNA platform
- Rendered with various shaders.
- Bridges and tunnels
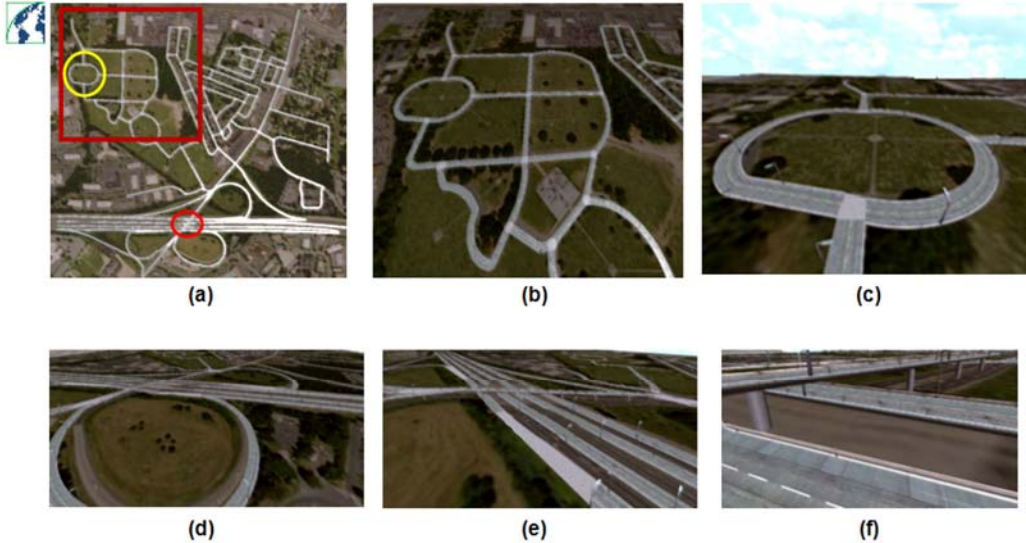- Digital terrain elevation and satellite image

968

Figure 4. Experimental results (a) 3D road network generated for the exampled road network. (b) Top view of the area indicated by the red rectangle in (a). (c) Side view of the area indicated by the yellow circle in (a). (d)(e)(f) Different views of the interchange part of the generated 3D road network model.

Outline

- **Introduction**
- **Body**
  - Road GIS Data Preprocessing
  - Parametric Representation
  - Civil Engineering Rule Based Road Surface Modeling
  - Intersection and Interchange
  - Implementation
- **Discussion**
- **Conclusion**

Discussion

- Do not aim at real road design and construction.
- Rapid and efficient 3D road modeling for computer graphics applications requiring realistic roadway models
  - Racing games and driving simulations
- Existing data do not provide sufficient information.
- Existing roads do not conform to the design standards.
- Reasonable approximations of real roadways
- May not have the exactly the same structure.

Outline

- **Introduction**
- **Body**
  - Road GIS Data Preprocessing
  - Parametric Representation
  - Civil Engineering Rule Based Road Surface Modeling
  - Intersection and Interchange
  - Implementation
- **Discussion**
- **Conclusion**

Conclusion

- The automatic generation of 3D high-fidelity road network models from real road GIS data

- Apply to the modeling of the whole road network.

- Be extended for other areas.
  - Subway system

Questions/Comments

☺ Jie Wang: jwang016@odu.edu

☺ Yuzhong Shen: yshen@odu.edu

Thank You !

8.0 PAPERS FROM MODSIM WORLD 2009 (PREVIOUSLY UNPUBLISHED)

8.1 Category Learning Research in the Interactive Online Environment *Second Life*

Category Learning Research in the Interactive Online Environment *Second Life*

Jan Andrews
Vassar College Program in Cognitive Science
andrewsj@vassar.edu;

Ken Livingston
Vassar College Program in Cognitive Science
livingst@vassar.edu;

Joshua Sturm
Vassar College Program in Cognitive Science
josturm@vassar.edu;

Daniel Bliss
Vassar College Program in Cognitive Science
daniel.p.bliss@gmail.com;

Daniel Hawthorne
Vassar College Program in Cognitive Science
freedijym@gmail.com;

Abstract. The interactive online environment *Second Life* allows users to create novel three-dimensional stimuli that can be manipulated in a meaningful yet controlled environment. These features suggest *Second Life*'s utility as a powerful tool for investigating how people learn concepts for unfamiliar objects. The first of two studies was designed to establish that cognitive processes elicited in this virtual world are comparable to those tapped in conventional settings by attempting to replicate the established finding that category learning systematically influences perceived similarity. From the perspective of an avatar, participants navigated a course of unfamiliar three-dimensional stimuli and were trained to classify them into two labeled categories based on two visual features. Participants then gave similarity ratings for pairs of stimuli and their responses were compared to those of control participants who did not learn the categories. Results indicated significant compression, whereby objects classified together were judged to be more similar by learning than control participants, thus supporting the validity of using *Second Life* as a laboratory for studying human cognition. A second study used *Second Life* to test the novel hypothesis that effects of learning on perceived similarity do not depend on the presence of verbal labels for categories. We presented the same stimuli but participants classified them by selecting between two complex visual patterns designed to be extremely difficult to label. While learning was more challenging in this condition, those who did learn without labels showed a compression effect identical to that found in the first study using verbal labels. Together these studies establish that at least some forms of human learning in *Second Life* parallel learning in the actual world and thus open the door to future studies that will make greater use of the enriched variety of objects and interactions possible in simulated environments compared to traditional experimental situations.

1. Introduction

The study of how people acquire and represent knowledge of category concepts is a broad area of research in cognitive science and psychology that includes a wide variety of issues and approaches. The human ability to group objects into categories and thereby treat them as equivalent for certain purposes is fundamental to human cognition, providing a foundation for memory, language, and reasoning.

The process by which new category concepts are acquired is also highly relevant to the study

of learning and therefore to the field of education. One method of studying this process is to teach adults (or children) unfamiliar categories and observe resulting changes in perceptual judgments of category instances. Using stimuli of various kinds, both physical objects and computer-generated images, our laboratory and others have demonstrated that certain effects generally occur (see, e.g., [1], [2]): namely, objects classified together are treated as more alike, an effect we call *compression*, and/or objects classified differently are treated as more distinctive, an effect we call *expansion*.

The broad conception of category learning assumed by this approach (and the field more generally) treats potential category instances as consisting of a set of values of various features or dimensions. For example, a particular dog would have values on such dimensions as body size, furriness, color of fur, length of tail, and so forth. We propose that compression and expansion effects caused by learning a new category distinction essentially constitute a change in the way these dimensions are represented, a kind of warping of the psychological dimensional space, such that the learner becomes more sensitive to dimensions important to the category distinction and less sensitive to those irrelevant to the distinction. This results in the psychological clustering of items that are classified together, allowing objects that differ from each other in category-irrelevant ways to cohere with one another and contrast with objects clustered in other categories.

In a typical category learning experiment, participants are shown a series of items, usually from a set of artificial stimuli created by the experimenters, and trained to classify them into two categories by receiving feedback on their classification responses. Training is stopped when classification accuracy reaches a high level or a certain number of runs through the stimuli ("blocks") have occurred. Participants then judge a large number of pairs of stimuli and either rate their similarity or decide whether the objects are identical or not, to determine how alike or confusable objects are within or between categories. A control group of participants judges the same stimulus pairs in the same way without having received classification training.

To test for compression/expansion effects, it is necessary to use unfamiliar categories of objects, since there can be no "control" group for categories that are already known. In addition, for the trained group, only successful learners' data are included because the effects are hypothesized to arise from acquisition of the category concepts. Interestingly, compression and expansion appear to also require multidimensional objects and do not seem to occur for objects consisting of a single dimension of variation (see [3] for a famous counterexample, and [4] for evidence against this counterexample).

In order to ensure that the learning processes tapped in this experimental paradigm are relevant to learning in the real world, it would be very helpful to apply it in a more flexible, dynamic, and interactive environment than that of the standard research laboratory where static stimuli are displayed one at a time on a flat computer display and the participant enters responses via a keyboard press. The research reported in this paper represents an effort to recreate this category learning paradigm in the interactive online 3-D environment *Second Life*. While some scholars have discussed the potential utility of *Second Life* for scientific research (e.g., [5], [6]), we are not aware of its previous use for studying category learning. The purpose of the first study was to determine whether compression/expansion effects would also occur in a category learning task in *Second Life*, in order to establish continuity between the cognitive/perceptual processes being tapped in standard laboratory tasks and in *Second Life*. If successful, this study would support the use of *Second Life* for testing new hypotheses related to category learning.

2. EXPERIMENT 1

2.1 Method

2.1.1 Participants A total of 44 Vassar College undergraduates either volunteered as part of an introductory psychology class requirement or were paid for their participation.

2.1.2 Materials All objects, including stimuli used in the categorization experiment, were built using the *Second Life* (SL) build tools (version 1.19.1 of SL). Categorization stimuli were

inspired by those used in [7]. Each of the 32 different stimuli was constructed by first resizing and then linking four separate objects, or "prims," in the language of SL, to a central sphere: two rectangular, and two pyramidal (see Figure 1). The spherical center of each object was wrapped with the "horizontal stripes" texture. The four separate objects were then attached to the central sphere of each stimulus, with each of these protrusions possessing the default texture. The first protrusion was attached to the top of the sphere and consisted of an inverted red, pyramidal prim, linked to a red rectangular prim. This top two-part protrusion was set to glow at an intensity of 0.10, to enhance its saliency. The second protrusion was a white pyramidal prim attached to the lower, left-hand portion of the central sphere and was the same for all stimuli, as was the third protrusion, a white rectangular prim attached to the lower, right-hand portion of the central sphere.
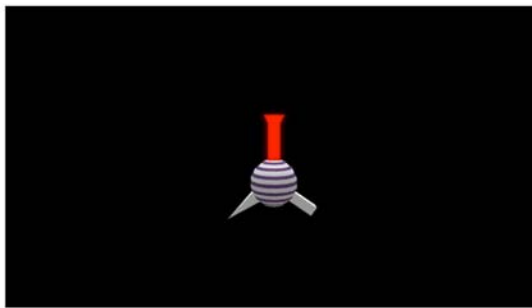

Figure 1.   Example of a stimulus used in Experiment 1

The stimuli varied on two dimensions: the height of the top protrusion and the width of the stripes covering the central sphere. There were eight values of each dimension. The top protrusion varied in increments of 0.5m between stimuli and ranged from 2.5m to 6.0m. Stripe width varied from 0.4 to 13.1 repetitions per object. In order to make the increments of this variation perceptually uniform, a set power function of 1.357 was applied. This implies that each increase in stripe width between stimuli corresponded to 2.5 just-noticeable-difference units.

The 32 stimuli were subdivided into 16 Gexes and 16 Zofs (nonsense labels selected for low associability). Members of the Gex category possessed the widest stripes and the longest top protrusions. Members of the Zof category possessed the narrowest stripes and the shortest top protrusions. The stimuli were displayed on a black rectangular background.

A mobile seat (the pink square in Figure 2) was also created and then programmed using the Linden Scripting Language (LSL) to respond to the click of a participant's mouse. The participant navigated this world of Gexes and Zofs as an androgynous avatar.
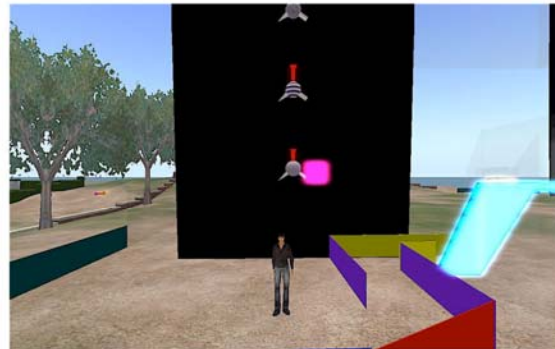

Figure 2.   The participant avatar and experimental setting used in Experiment 1

2.1.3 Procedure Participants were randomly assigned to the learning or control condition. They were tested individually and entered SL using a Macintosh computer running OSX 10.4.11. Positioned in front of the screen, participants were told that they would first use the arrow keys on the keyboard to navigate the avatar through a short orientation path, in order to become familiar with the skills needed to complete the task. Once through the path, the participants were asked to click on the virtual seat in front of the avatar to fix the avatar's view in "mouse look" mode (first-person perspective).

In the classification task, all 32 stimuli were arrayed vertically in a different random order against each of eight black backgrounds, one behind the other (see Figure 2). The same random orders were used across participants. Participants viewed the stimuli one-by-one while moving upward on the seat, moving past the objects displayed against the black background. The seat was programmed to stop at each stimulus location. After each stimulus was shown, participants indicated which category they thought the object belonged to by clicking either the "Gex" or the "Zof" button that

appeared in the upper right-hand corner of the screen. Immediately following the button press they received auditory feedback concerning the correctness of the response. Arrows directed participants to each successive set of stimuli (with a new black background and seat). Participants completed all eight blocks unless they achieved a total of at most one incorrect response in two consecutive blocks, at which point training was stopped immediately.

After the classification task, participants rode on a seat to the similarity task area. There they viewed the 32 stimuli in pairs, seeing the objects in a given pair one at a time for 3 seconds each. Participants rated the similarity between members of each pair on a scale from 1 to 9 (1 being least similar and 9 being most similar) by clicking one of nine buttons that appeared in the upper, right-hand corner of the screen. A total of 90 pairs of stimuli (30 Gex-Gex, 30 Zof-Zof, 30 Gex-Zof) were rated, and the pairs were presented in the same random order across participants. The 90 pairs were split into six blocks of 15 pairs; as above, blocks were separated onto different black backgrounds, and participants followed arrows to each successive block.

Control condition participants completed only the similarity judgment task, without any prior category learning.

2.2 Results

Four of the 24 participants in the learning condition failed to pass the learning criterion – they did not complete two consecutive blocks with a total of at most one incorrect response in the classification task – so their data were not included in the analysis.

A 2 (group: learning vs. control) by 2 (pair-type: members within the same category vs. members from separate categories) analysis of variance with repeated measures on the second factor was performed on the similarity ratings. This yielded a significant main effect of pair-type ($F_{(1,38)} = 501.724$, $p < .001$) and a significant interaction between group and pair-type ($F_{(1,38)} = 7.847$, $p = .008$). As shown in Figure 3, within-category pairs were judged to be more similar than between-category pairs by both groups, but the learning group judged within-category pairs to be more similar than did the control group and the between-category pairs to be less similar

than did the control group, corresponding to compression and expansion, respectively. However, planned one-tailed t-tests revealed that only the compression effect was significant ($t(38) = 1.748$, $p < .05$).
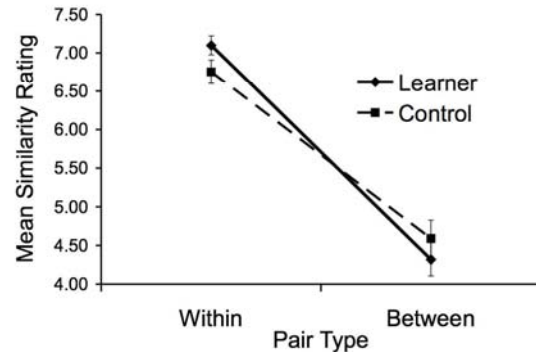

Figure 3. Results of Experiment 1

2.3 Discussion

Experiment 1's instantiation of the category learning paradigm described in the Introduction within the environment of *Second Life* produced a significant compression effect of the sort typically found in standard laboratory versions, suggesting that the category learning processes tapped are substantially the same. While our ultimate goal is to develop more innovative uses of *Second Life* for research on category learning, we next took advantage of some of the more straightforward features of the laboratory we had already constructed there to explore a long standing question about the role of verbal labels in category learning.

Normally when we learn new categories we simultaneously learn words for those categories. In fact, using a single word to refer to a set of objects that vary on several dimensions has been hypothesized to be central to the category learning process in humans (e.g., [8]). The single label provides an explicit feature common to all category instances and may support the learning of categories in important ways. To rigorously test this claim requires the study of category learning in the absence of verbal labels, and that poses a methodological challenge. We made creative use of *Second Life's* unique stimulus-building tools to meet this challenge.

In Experiment 2 we designed a novel response system to allow participants to learn the same categories of objects used in Experiment 1, but without any verbal labels. This allowed us to test two interesting questions: (1) Will category learning be more difficult under these conditions compared to those of Experiment 1? And, even if this is the case, (2) will those participants who do successfully learn the category distinction exhibit compression effects similar to those found in Experiment 1? That is, will the underlying mechanism of category learning be the same in the absence of labels?

3. EXPERIMENT 2

3.1 Method

3.1.1 Participants A total of 39 Vassar College undergraduates either volunteered as part of an introductory psychology class requirement or were paid for their participation.

3.1.2 Materials The stimuli and categories were identical to those used in Experiment 1. However, two visual patterns were used in place of the verbal labels "Gex" and "Zof." These patterns were colorful designs similar to tie-dyed fabric and slightly different from each other (see Figure 4). They were chosen because they are extremely difficult to describe in any simple way and thus do not easily lend themselves to description by a single label.
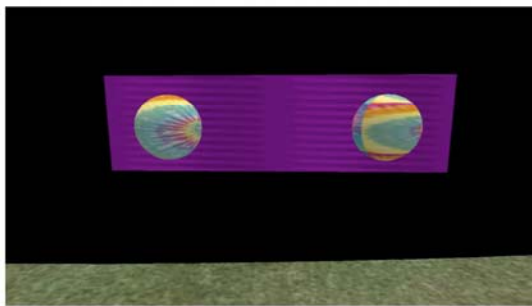

Figure 4. Nonverbal response buttons used in Experiment 2

3.1.3 Procedure The procedure was identical to that used in Experiment 1 except that the labels "Gex" and "Zof" were excluded from the classification judgments and feedback. Participants pressed one of the two buttons shown in Figure 4 to indicate which category they thought a stimulus belonged to and were simply told whether they were correct. The right-left positions of the two response buttons relative to each other were randomly varied to prevent participants from using the labels "right" and "left" for the two patterns.

3.2 Results

Nineteen of the 39 participants failed to meet the learning criterion used in Experiment 1, and their data were not included in analyses. A chi-square test of independence showed that learning success was related to the presence (Experiment 1) or absence (Experiment 2) of verbal labels ($\chi^2(1, N = 63) = 6.58$, $p = .01$).

In order to determine whether compression or expansion occurred, the data from this experiment were combined with the control group data from Experiment 1 and analyzed in the same way. The resulting 2 (group: nonverbal vs. control) by 2 (pair-type: within-category vs. between-category) analysis of variance with repeated measures on the second factor yielded a significant main effect of pair-type ($F(1,38) = 682.612$, $p < .001$) and a significant interaction between group and pair-type ($F(1,38) = 8.852$, $p = .005$). As shown in Figure 5, the pattern of the means is identical to that of Experiment 1 and, once again, planned one-tailed t-tests revealed that only the compression effect was statistically significant ($t(38) = 1.850$, $p < .05$).

Additional t-tests showed that the mean similarity ratings for the learning groups in Experiment 1 (with verbal labels) and Experiment 2 (without verbal labels) did not differ significantly for either the within-category pairs or the between-category pairs.

3.3 Discussion

The use of complex visual patterns in place of verbal labels in Experiment 2 made the category learning task, otherwise identical to that used in Experiment 1, much more difficult. This is consistent with results reported by Lupyan, Rakison, and McClelland [8] using a similar task. As they note, while it is impossible to be certain that participants were not surreptitiously using invented verbal labels, the fact that learning was significantly more difficult using nonverbal category responses suggests that explicit verbal labels do facilitate category learning. It is interesting that half of our participants in the

nonverbal condition were nonetheless able to learn the categories to a very high level of accuracy.
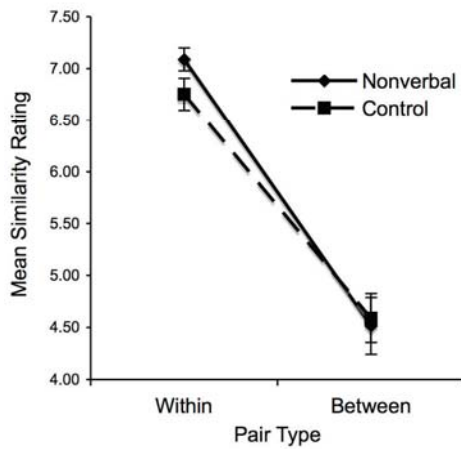

Figure 5. Results of Experiment 2

However, our main interest was in determining whether, for those participants who were able to master the category distinction, the compression result obtained in Experiment 1 would still occur. In fact, the compression effect for successful learners was virtually identical in the two experiments, suggesting strongly that while verbal labels may make category learning easier, they are completely unrelated to the compression/expansion processes associated with category learning. This is consistent with the idea that these processes are indeed fundamental to the formation of new category concepts.

4. CONCLUSION

These experiments on category learning in *Second Life* demonstrate both its continuity with standard laboratory research and its utility for testing new hypotheses. Our evidence that *Second Life* taps the same cognitive processes observed in laboratory research supports its validity for further research on learning and cognition.

Our next study, currently underway, makes much greater use of *Second Life*'s potential for creating dynamic, interactive stimulus features and engaging, goal-oriented tasks. We expect this will allow us to explore category learning processes in ways that are not possible in a standard laboratory situation but that are actually significantly more like the real world.

REFERENCES

1. Goldstone, R. (1994). "Influences of categorization on perceptual discrimination," *Journal of Experimental Psychology: General,* 123, pp. 178-200.

2. Livingston, K.; Andrews, J. (2005). "Evidence for an age-independent process in category learning," *Developmental Science,* 8(4), pp. 319-325.

3. Tajfel, H.; Wilkes, A. (1963). "Classification and quantitative judgement, " *British Journal of Psychology,* 54, pp. 101-114.

4. Andrews, J.; Livingston, K. (2000). "Accentuation of category differences: Revisiting a classic study." In L. Gleitman & A. Joshi (Eds.), Proceedings of the 22nd Annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum.

5. Boellstorff, T. (2008). Coming of Age in Second Life: An Anthropologist Explores the Virtually Human. Princeton: Princeton University Press.

6. Bainbridge, W. S. (2007). "The Scientific Research Potential of Virtual Worlds," *Science*, 317, pp. 472-476.

7. Cooke,T; Jakel, F.; Wallraven, C.; Bulthoff, H. (2007). "Multimodal similarity and categorization of novel three-dimensional objects," *Neuropsychologia*, 45, pp. 484-495.

8. Lupyan, G.; Rakison, D.; McClelland, J. (2007). "Language is not just for talking: Redundant labels facilitate learning of novel categories," *Psychological Science*, 18(12), pp. 1077-1083.

Appendix A — MODSIM World 2010 Conference & Expo Organization CP3 Board Members, Directors, Advisors, and Conference Committee

Members

Chairman
Josh Jackson, SAIC
Vice Chairman
William L. Younger, Jr. MYMIC, LLC
President
Thomas Reese, VMASC
Secretary/Treasurer
E. Dana Dickens, III, HRP

Directors

Jeff Bergenthal, Lockheed Martin
Wayne Buck, NATO
Don Combs, EVMS
Randy Garrett, Northrop Grumman
Fred Lewis, NTSA
Bill Waite, Aegis

Advisors

Dan Abahazy, USJFCOM/J7
Sharon Welch, NASA Langley

Conference Committee

Conference Chair
Robert "Bob" Armstrong, Booz | Allen | Hamilton
Program Chair
Jennifer McNamara, BreakAway, Ltd.
Deputy Program Chair
Dr. Thomas E. Pinelli, NASA Langley
Outreach Chair
Tammy Van Dame, Naval Sea Systems Command, Dam Neck
STEM Coordinator
Mark Clemente
Logistics chair
Andrew M. Sinclair, HRP
Conference Coordinator
Bonnie B. Stewart, Organization Management Group

Appendix B — MODSIM World 2010 Conference & Expo Track Chairs and Track Descriptions

Defense Track

Chair, Steve Husak, Steve Husak & Associates

Deputy Chair, Thomas J. Sides, Global Maritime Security Systems, Combat Directions Systems Activity, Dam Neck

The Hampton Roads, Virginia, area is home for many Department of Defense major commands. It is also home to the U.S. Joint Forces Command and NATO's Allied Command Transformation. Together, these organizations create a unique synergy that makes the Hampton Roads region a focal point for the development and use of modeling and simulation (M&S) to meet complex training requirements. The Defense Track provides academic, government and industry participants the opportunity to address M&S technologies relative to their ability to resolve or mitigate current and projected operational challenges. The Defense Track will specifically focus on the following five key aspects of M&S: (1) currently available products and training tools; (2) new and emerging technologies and related tools; (3) best practices and case studies of successful applications of M&S; (4) the use of M&S for decision-making, support and risk assessment; and (5) discussions and predictions about the future of M&S for use in defense. Other aspects of interest include the development and advancement of M&S for: defense-related training, assessment and evaluation; technology integration; protection of critical infrastructure and infrastructure resiliency; disaster (contingency) planning, implementation, response and recovery; interoperability; standards; and the development of policy and strategies for managing global, national and regional events.

Engineering & Science Track

Co-Chair, C. Matthew O'Connor, Combat Directions Systems Activity, Dam Neck

Co-Chair, Dr. Daniel P. Schrage, School of Aerospace Engineering, Georgia Tech

Deputy Chair, Kevin Stenstrom, Raytheon

Modeling and simulation-based engineering and science is rapidly becoming an essential scientific methodology for research (theoretical and experimental) and development; concept generation and consumer marketing; product design and manufacturing; systems development and integration; project management; systems engineering; logistics; new methods for verification, validation and uncertainty quantification; and integration and interoperability. Continuing advances in computational science and networking technologies have made modeling and simulation (M&S) a powerful and ubiquitous tool for engineers and scientists, and have made it possible to vastly extend the range, depth and applications of modeling and simulation—especially when the phenomena being investigated are not observable or measurements are impractical or too expensive.

The Engineering & Science Track provides a forum for individuals to meet, exchange and debate ideas, network, foster future research opportunities, and learn from national and international experts on a wide variety of engineering and science topics and issues. By exploring the challenges, applications, technologies and future directions, this track will challenge traditional perspectives, cultivate new ideas, and foster dialogue on how M&S can address the present and future challenges facing engineering and science.

Health & Medicine Track

Chair, Jennifer Arnold, Booz | Allen | Hamilton

Deputy Chair, Christine Shamloo, Booz | Allen | Hamilton

Deputy Chair, Rachel Spencer, SiTEL MedStar Health

Deputy Chair, Brian Levine, SAIC

The scale and complexity of the health care sector is almost as daunting as the unprecedented levels of change this sector is facing. It is becoming increasingly clear that modeling and simulation (M&S) can be instrumental in addressing the multi-faceted challenges health care is facing. The Health and Medicine Track recognizes that M&S tools, techniques and standards are playing an increasingly important role in improving patient safety. M&S is also changing the way medicine is taught and practiced, how health care professionals are trained, the development of solutions to health problems, how we respond to public health issues, and the conduct of basic and applied medical research in a variety of areas including genetics, neuroscience and population dynamics. Fundamental to the application of M&S is the underlying assumption that insight into the behavior of a system can be developed or enhanced from a model that adequately represents a selected subset of the health care system's attributes. This track provides a forum for individuals to meet, exchange and debate ideas, network, foster future research opportunities, and learn from national and international experts on a wide variety of health and medicine topics and disciplines. By exploring applications, technologies and future directions, this track will challenge traditional perspectives, cultivate new ideas, and foster dialogue on how M&S can address the present and future challenges health and medicine face.

Homeland Security & First Responders Track

Chair, Bruce Milligan, Booz | Allen | Hamilton

Deputy Chair, Jay Allen, OUSD (P&R), ADL Initiative

Deputy Chair, Tammy Van Dame, Combat Directions Systems Activity, Dam Neck

One key factor that both first responders and those involved with homeland security have in common is that the level of their professional training and their ability to quickly and effectively respond to threats to our national security, natural disasters and other catastrophic events, is something that could have a potential impact upon any citizen of this nation. Therefore, effective training using the most robust and best-developed methods and technologies is a critical

component to the personal safety of each and every American. The Homeland Security & First Responder Track focuses on computer models, simulations and serious games that relate to homeland security. Many such tools are currently in use, or are in development by, groups ranging from emergency preparedness personnel (including police, firefighters, public works, National Guard and others), to those involved in domestic counter-terrorism or counterintelligence, border security, maritime and aviation security, and many similar realms. Conference speakers will range from well-known authors and game designers to those who have been in the front lines of disaster response, law enforcement and homeland security.

Human Dimension Track

Chair, Dr. Kara Latorella, Crew Systems & Aviation Operations Branch, NASA Langley

Deputy Chair, Phil Jones, MYMIC, LLC

Modeling Human Behaviors & Interactions: Understanding the human dimension cuts a broad swath; from individual performance and thought to the behaviors of humans in complex human/machine systems, cultures and personal interactions. This track focuses on modeling and simulation as it relates to individual human behavior (information-seeking and monitoring, decision-making and controlling) and the influence of effect and environmental conditions on behavior, as well as the interaction of individuals− especially as participants in complex human/machine systems, and human social networks and organizations. This track will address the utility of models and simulations in terms of their application to solving design and analysis problems such as: characterizing individual differences; predicting human performance and errors; designing and evaluating human/system integration and human/automation interaction; team construction and performance; and predicting communication dynamics, or other emergent social behaviors. Innovative approaches to requirement definition, validation and verification, and data collection and analysis will also be presented.

K-20 STEM Education Track

Chair, Mark Clemente, Educator-in-Residence, National Institute of Aerospace/Virginia Beach City Public Schools

Deputy Chair, Dr. Vincent Charles Betro, STEM Outreach Coordinator, University of Tennessee, SimCenter at Chattanooga: National Center for Computational Engineering

Accepting the Challenge: The K-20 Science, Technology, Engineering and Math (STEM) Education Track will be a forum organized to draw national attention to the need to educate and train the next generation of engineers and scientists in the theory and practice of modeling and simulation-based (M&S) engineering and science and the myriad of associated technologies. Maintaining our competitive advantage and full utilization of M&S-based engineering and science requires fundamental changes in America's K-20 education system—especially in the teaching/learning of STEM—and the integration of M&S into the K-20 curriculum, as both content and as an instructional strategy. The forum's purpose is to create a dialogue that will lead to the development of: (1) a national plan for integrating M&S-based engineering and science

into K-20 education; (2) a literacy framework for M&S; and (3) a research agenda. This forum brings together nationally known speakers and panelists from academia, business, education, government and industry, and will include some of the nation's foremost thinkers and practitioners. This forum is designed for K-20 STEM educators and administrators; legislators; state and local school board members and superintendents; college faculty and administrators; engineers and scientists; state and federal policy makers; business, corporate and civic leaders; and members of workforce boards, consortia, partnerships and alliances.

Serious Games & Virtual Worlds Track

Chair, Dr. Benjamin Bell, CHI Systems, Inc

Deputy Chair, Dr. Winston "Wink" Bennett, Warfighter Readiness Research Division, Human Effectiveness Directorate, Air Force Research Laboratory

Committee Members: Paul Cummings, ICF International; Dr. Jerzy Jarmasz, Defence Research & Development Canada; Dr. Stephanie Lackey, University of Central Florida Institute for Simulation & Training; Dr. Sae Schatz, University of Central Florida Institute for Simulation & Training; and Lt. Joel Walker, USAF, Air Force Research Laboratory

Rapid advances in computer hardware and software continue to transform technologies and blur the distinctions between simulations and games. Paralleling this trend, aggressive progress in bandwidth, compression and animation technologies have moved virtual worlds from the shadows of technophilic oddity into the bright light of productive work. This cross-cutting track focuses on the related but distinctive areas of serious games and virtual worlds—disciplines that share a focus on simulation and interaction, but possess distinctive approaches, technologies and cultures. The Serious Games & Virtual Worlds Track will explore all dimensions of serious games and virtual worlds, including those relating to the broader MODSIM World 2010 tracks: Defense, Engineering & Science, Health & Medicine, Homeland Security & First Responders, The Human Dimension, and K-20 STEM Education. By exploring the applications, technologies and future directions of serious games and virtual worlds, this track intends to challenge traditional perspectives, cultivate new ideas, and foster dialogue focusing on how these capabilities can extend the reach of modeling and simulation.

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-03-2011	Conference Publication	

4. TITLE AND SUBTITLE

Selected Papers and Presentations Presented at MODSIM World 2010 Conference & Expo

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Pinelli, Thomas E. (Editor)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

736466.01.04.07.08

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER

L-20001

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)

NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

NASA/CP-2011-217069/Part 2

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 66 System Analysis and Operations Research
Availability: NASA CASI (443) 757-5802

13. SUPPLEMENTARY NOTES

Thomas E. Pinelli, NASA Langley Research Center, Strategic Relationships Office; thomas.e.pinelli@nasa.gov; Phone: (757) 864-2491

14. ABSTRACT

MODSIM World 2010 was held in Hampton, Virginia, October 13-15, 2010. The theme of the 2010 conference & expo was "21st Century Decision-Making: The Art of Modeling& Simulation". The conference program consisted of seven technical tracks - Defense, Engineering and Science, Health & Medicine, Homeland Security & First Responders, The Human Dimension, K-20 STEM Education, and Serious Games & Virtual Worlds. Selected papers and presentations from MODSIM World 2010 Conference & Expo are contained in this NASA Conference Publication (CP). Section 8.0 of this CP contains papers from MODSIM World 2009 Conference & Expo that were unavailable at the time of publication of NASA/CP-2010-216205 Selected Papers Presented at MODSIM World 2009 Conference and Expo, March 2010.

15. SUBJECT TERMS

Modeling and simulation; Agent-based modeling; Game theory; Discrete event simulation; Cognitive modeling; Visualization

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
U	U	U	UU	522	19b. TELEPHONE NUMBER *(Include area code)* (443) 757-5802