CRM D0006277.A3/Sanitized
November 2005

Game-Based Experimentation for Research in Command and Control and Shared Situational Awareness

Peter P. Perla
Michael Markowitz
Christopher Weuve

CNA
4825 Mark Center Drive • Alexandria, Virginia 22311-1850

Approved for distribution: May 2002

Dr. Peter P. Perla
Director for Interactive Research Products
Office of the Senior Vice President for Research

This document represents the best opinion of CNA at the time of issue.
It does not necessarily represent the opinion of the Department of the Navy.

Cleared for public release. Distribution unlimited.
For copies of this document call: CNA Document Control and Distribution Section at 703-824-2123.

Contents

Summary

This paper documents the support CNA provided to the Warfare Analysis and Research Department of the Naval War College for an experiment conducted in the spring of 2002. This experiment took the form of a series of games played by teams from the U.S. Naval Academy, the U.S. Air Force Academy, and the Naval War College. The test bed was the internet-based game *SCUDHunt*, developed earlier by CNA and ThoughtLink Inc. for the Defense Advanced Research Projects Agency. In this simple yet elegant game, players take the roles of sensor asset managers and attempt to deploy their sensors to search a small, gridded map for hidden "SCUD" launchers. Each sensor has different characteristics of coverage and reliability. To play effectively, the players must work together, sharing information and developing their shared situational awareness in order to find the SCUDs and make accurate strike recommendations.

The Naval War College was interested in carrying out this experiment partly as a proof-of-concept about the value of using purpose-built games, and partly as a means of deriving insights into the effects of command styles and visualization techniques. Some of the incentives for this research included providing scientific advice to assist the designers of the Naval War College's Global War Game series.

The Naval War College has at least two broad concerns related to the Global series—the representation of modern command and control systems and techniques, and the role of shared situational awareness in the concepts underlying them. These concerns intersect strongly with an increased emphasis within the Department of Defense as a whole on how to develop new approaches and methods for Joint command and control that will better take advantage of new technology and concepts—like network-centric warfare and effects-based operations. Indeed, the key future-looking concept for the U.S. military—Joint Vision 2020—includes Joint command and control as one of the key Joint mission areas U.S. forces will have to perform.

Evolving thought in DoD's command-and-control community emphasizes that improved techniques for creating and maintaining shared situational awareness are key elements of any future system of Joint command and control. U.S. forces will achieve superiority over an adversary by achieving self-synchronization, partly because of increased shared situational awareness. This self-synchronization will enable U.S. forces to act at an increased operational tempo. Improving our understanding of the variables that affect shared situational awareness is critical to implementing these future concepts.

CNA's previous work for the Defense Advanced Research Projects Agency took some pioneering steps in exploring shared situational awareness using a game-based experimental technique. The Naval War College decided to use a similar approach for this experiment. The earlier CNA study focused on the effects of communications modes and shared-visualization tools on developing shared situational awareness. The Naval War College project focused on command styles and on the use of a "push" visualization mechanism. Measures of interest included a score associated with the shared situational awareness developed by the teams during the course of play, and a score for measuring the accuracy of the estimates of target locations made by the team's members.

The preliminary results of the analysis of the experimental data proved of only limited interest. Because of some technical and logistical difficulties associated with the execution of the experiment, it is difficult to state with confidence the reliability of the experimental outcomes. That said, the experimental data do seem to hint at a statistically significant improvement in both the shared situational awareness and accuracy of teams employing a "command by direction" style when compared to the same teams playing under "command by influence" or "command by plan" styles. These results indicate potential value in exploring this effect through additional follow-on experiments targeted specifically to explore this factor.

In addition to assisting them with the design and preliminary analysis of their experiment, the Naval War College tasked CNA to consider broader issues related to experimentation in this field.

There are three types of experiments:

1. Experiments designed to explore new ideas or phenomena
2. Experiments designed to test hypotheses
3. Experiments designed to demonstrate new concepts and their feasibility (or lack thereof).

Games provide a wealth of flexibility for exploring, testing, and demonstrating a host of variables and issues associated with decisionmaking. Unfortunately, a single iteration of a complex, multiplayer, large-scale operational wargame is expensive in time and money. Such games are poor vehicles for scientific experimentation, for hypothesis testing and "scientific proof."

The original *SCUDHunt* game and experiment, as well as the current effort, demonstrate the potential for using a somewhat different sort of gaming environment to formulate and test hypotheses using rigorous scientific and statistical techniques. We can characterize these sorts of games as "distillations"—distinguishing them from simple "abstractions," like chess, and detailed "simulations," like Global.

Game-based experimentation is a scientifically rigorous approach to exploring fundamental command-and-control issues. The DoD command-and-control community should increase the use of distillation-style games as part of a program of experimentation and research related to shared situational awareness and command and control. To do so, DoD should look for opportunities to create game-based research efforts—laboratories, if you will—that the paper describes in more detail. Such laboratories are less specific facilities than they are assemblages of critical components and expertise—the games to serve as experimental test beds; the game designers to create the games; and the analysts and scientists to formulate the problems, design the experiments, and analyze the data. Such laboratories could be "virtual" organizations, bringing together subject-matter experts from across the United States and other nations as well. A "virtual, distributed laboratory for game-based experimentation" can help advance our understanding of command and control, information operations, network-centric warfare, and other critical concepts that, in many ways, remain buzzwords rather than realities.

Introduction

The Naval War College (NWC)—through its Warfare Analysis and Research Department (WARD)—is pursuing a program of research into designing militarily relevant and scientifically valid experiments to investigate shared situational awareness (SSA) in the evolving environment of U.S. Joint command and control (C2). One aspect of this research builds on the foundation of reproducible results obtained from CNA's prior work in the field of measuring SSA in a wargaming environment.[1]

Background: Shared situational awareness

Some of the incentives for the NWC's SSA research originated in the results of initial attempts to explore concepts associated with network-centric warfare (NCW) (or the broader concept of network-centric operations, or NCO) and effects-based operations during Global War Game 2000. During that game, some observers concluded "that modern information technology, and current concepts for its use did not free the Global 2000 participants from certain traditional constraints, at least not to the extent envisioned in theoretical discussions of network-centric operations."[2] The game attempted to implement a netted C2 system similar in concept to that envisioned for the future. In actual operation during the game, however, the players faced serious challenges in using the system to build SSA and to take practical advantage of the information superiority theoretically available to them. The game emphasized the development of a common

---

1. This research is documented in: CRM D0002722.A2, *Gaming and Shared Situation Awareness*, by Peter P. Perla *et al.* (November 2000) and CRM D0002895.A1, *Defining and Measuring Shared Situational Awareness,* by Albert A. Nofi. (November 2000).

2. Kenneth Watman, "Global 2000," *Naval War College Review, Spring 2001, Vol LIV, No. 2,* 75-88, pp. 86-87.

operating (or operational) picture (COP), but game play revealed that "a common operational picture does not in itself enable the parts of a force or staff to regulate themselves; some shaping and filtering of the data in that common operating picture is still required."[3]

Although several factors contributed to the mechanical and conceptual difficulties players encountered in Global 2000, two of the more significant appear to center around (1) the way military organizations build SSA, and (2) the way current wargaming systems and processes represent the workings of a network-centric military C2 system.

Unpublished observations of Global 2001 indicate that the NWC made progress in solving some of the problems associated with the bandwidth available to support the information flow in Global 2000. Nevertheless, additional issues arose as players continued to wrestle with applying some of the principles of NCO in the wargaming environment. Not surprisingly, no set of pre-game standard operating procedures survived contact with the players. Players developed workarounds to deal with perceived problems and those workarounds sometimes became the de facto procedure, even after the problems that had necessitated them had been fixed.

The promise of self-synchronization as a way to speed up operations beyond the enemy's ability to respond effectively seemed to be less than the panacea many had foreseen. In some cases, observers characterized the self-synchronization they saw during play in terms of everyone's agreeing that a particular situation was someone else's problem to solve. In at least one instance, the fast cycle time of information flow and the lack of certainty about the authoritativeness of information almost led to an inadvertent act of war, because bad information was passed along so quickly.

One of the key elements of network-centric operations is the ability of networked forces to share information, synchronize their operations on the basis of that information, and so act on the information quickly and precisely. Global 2001 indicated that trust—in the quality of information, in the authoritativeness of the sources providing it,

3. Watman, p. 87.

and in the projected actions of other elements of the force in response to it—is a fundamental issue we need to explore further.

There has been only limited research into the factor of trust or other elements affecting the extremely difficult task of building operationally effective SSA in the enormously complex environment of large-scale military operations. Much of the existing work on SSA has concentrated on tactical situations (particularly focused on aircrew and ground controllers) and small-scale teams.[4] One challenge is to extrapolate, or scale up, what has been learned from such small-scale experiments into the larger military environment. In particular, conducting experiments focused on key questions of interest in the Joint environment, such as the effects of alternative command styles, could provide new insights into the processes and effectiveness of building SSA in those environments.

In another article dealing with Global 2000, Captain Robert C. Rubel, USN, argued that the very approach to wargaming taken by the Global series may be in need of revision if it is to capture the key elements of NCW.[5] Rubel makes a case that the increasing emphasis on network-centric warfare requires rethinking and possibly rebuilding the basic framework of wargames at the level of Global (that is, at the operational and operational-strategic levels of war).

The Naval War College thus has at least two broad concerns related to wargaming, command and control, and shared situation awareness. These concerns intersect strongly with an increased emphasis within the Department of Defense (DoD) as a whole on how to develop new approaches and methods for Joint C2 that will better take advantage of new technology and concepts—like NCW and EBO. Indeed, the key future-looking concept for the U.S. military—Joint Vision 2020—includes Joint C2 as one of the key Joint mission areas U.S. forces will have to perform.[6] To help drive research and

4. See Nofi, *Defining SSA.*

5. Robert C. Rubel, CAPT, USN, "War-Gaming Network-Centric Warfare," *Naval War College Review, Spring 2001, Vol LIV, No. 2,* 61-74.

6. *Joint Vision 2020,* Director for Strategic Plans and Policy, J5, Strategy Division, U.S. Government Printing Office, Washington, DC, June 2000; available at http://www.dtic.mil/jv2020/jvpub2.htm

implementation of effective concepts in this area, the Joint Requirements and Oversight Council (JROC) initiated a study of current and future Joint Task Force command and control (JTF C2). The Joint C4ISR Decision Support Center (DSC) is the principal organization responsible for collecting and analyzing data, and preparing the papers and briefings associated with this project.[7] The current draft report of this effort identifies shared situation awareness as one of the 11 key operating characteristics of current and future JTF C2 systems.

It is thus clear that evolving thought in DoD's C2 community emphasizes SSA as one of the key elements of a system of Joint C2 that will enable U.S. forces to achieve superiority over an adversary by achieving self-synchronization among Joint forces. This self-synchronization, in turn, will enable such forces to act at an increased operational tempo and virtually paralyze the adversary's ability to react. Improving our understanding of the variables that affect the creation and maintenance of SSA among operating forces in operational situations is critical to implementing these future concepts of Joint C2.

CNA's previous work for the Defense Advanced Research Projects Agency (DARPA) took some pioneering steps in exploring SSA using a game-based experimental technique.[8] The WARD decided to use a similar approach for its first experiment of 2002. Whereas the earlier CNA study focused on the effects of communications modes and shared visualization tools on developing SSA, the WARD project focused on command styles and on the use of a "push" visualization mechanism.

The experiment would also serve a second purpose. It would provide the Naval War College with a proof-of-concept about the value of using purpose-built games to explore important elements of the C2 problem set in a more scientifically rigorous manner than other approaches. The NWC asked CNA to work with the Warfare Analysis and Research Department to increase the WARD's understanding of the gaming environment (*SCUDHunt*) used in CNA's prior research

---

7. More information on this study and its current state of progress may be found at the DSC web site, http://www.dsc.osd.mil/

8. See Perla, *Gaming and SSA,* and Nofi, *Defining SSA.*

on shared situational awareness. We would help the WARD modify the game as required, assist in designing the experiments the WARD planned to carry out using *SCUDHunt*, and aid with the collection and statistical analysis of the data generated by the experiment. The WARD also asked CNA to think about the broader issues related to C2 in the new environment, and how to use games both to represent that new C2 environment and to help conduct further research important to our understanding of that environment.

Outline

This paper reports on CNA's work on this quick-response project. There are four main sections.

In the first section, we discuss some of the ideas behind our notions of *game-based experimentation*. This section relates three different broad concepts of gaming to three broad classes of experimental objectives.

The second section particularizes this discussion. It proposes some ideas about how game-based experimentation can make an important contribution to research in command and control. We focus on the integration of theoretical understanding into practical models, the province of the wargame designer.

The third section of the paper deals with the specific WARD experiment using *SCUDHunt*. Because the WARD is preparing its own full report on this project, this section concentrates on the data and analysis of the experiment, detailing only enough of the background to provide context for that analysis.

The fourth and final section of the paper looks to the future. It recommends an effort to document what we have learned so far about game-based experimentation. It also discusses the need for the development of game-based laboratories for future scientific research, particularly in C2.

Game-based experimentation

Games in general, and wargames in particular, focus on the decision-making processes of human players.[9] Games can be tailor-made to explore specific kinds of decisions under specific sets of assumptions. As such, games can be laboratories—or at the very least, laboratory equipment, test tubes and beakers, as it were—for conducting experiments designed to study human decision processes.

Without trying to write a thesis on the philosophy of science, we do think it important to talk a little about this notion of experimentation if we are to understand the role and value of gaming in it. There are lots of ways of thinking about experimentation and experiments, but one way we find most useful for our purposes distinguishes three types of experiments:

1. Experiments designed to explore new ideas or phenomena
2. Experiments designed to test hypotheses
3. Experiments designed to demonstrate new concepts and their feasibility (or lack thereof).

The range of complexity and scope of coverage available through the use of games provide a wealth of flexibility for exploring, testing, and demonstrating a host of variables and issues associated with decision-making. Traditionally, gaming—wargaming in particular—has been associated with historical approaches to analysis more than with scientific ones. Traditional wargames helped to explore new ideas or demonstrate new concepts. A single iteration of a complex, multi-player, large-scale operational wargame is expensive in time and money to produce, and virtually impossible to replicate. Such games

9. See Perla, Peter. *The Art of Wargaming*. U.S. Naval Institute Press. Annapolis, MD: 1990.

seemed poorly adapted as vehicles for scientific experimentation, for hypothesis testing and "scientific proof."

The original *SCUDHunt* game and experiment demonstrated the potential for using a somewhat different sort of gaming environment to formulate and test hypotheses using rigorous scientific and statistical techniques. *SCUDHunt* is a wargame, but not in the sense that Global is a wargame. Indeed, *SCUDHunt* and Global seem almost at opposite ends of a wide spectrum of gaming environments. We can characterize that spectrum in many ways, but, for our purposes, we propose to look at it in terms of three "-tions:" abstractions, distillations, and simulations.

Abstractions

We characterize abstractions as pure strategy or decision games. They are usually based on geometrical and conceptual environments or on highly stylized representations of the real world. The decisions the players make are also usually highly abstracted when compared to real-world decisions. The usual examples of abstract games include the classic games of checkers and chess. A more representational sort of game, one moving from the realm of abstraction to that of distillation, is the popular commercial game, Hasbro's *Stratego*.

Distillations

When we talk about games as distillations, we are using terminology sometimes associated with recent efforts to explore agent-based approaches to studying combat, such as ISAAC and EINSTEIN.[10] Games meet our definition of distillations when they reduce real-world problems and entities into simplified representations focused on a few prominent elements of that real-world environment. One way of distinguishing distillations from abstractions is that real-world

10. See, for example, CNA Research Memorandum 97-61.10, *Irreducible Semi-Autonomous Adaptive Combat (ISAAC): An Artificial-Life Approach to Land Combat*, August 1997. Andrew Ilachinski's ISAAC and EINSTEIN applications are extensively documented at http://www.cna.org/isaac/

language and concepts can be used to describe situations, actions, and outcomes in a distillation without a lot of mental gymnastics.

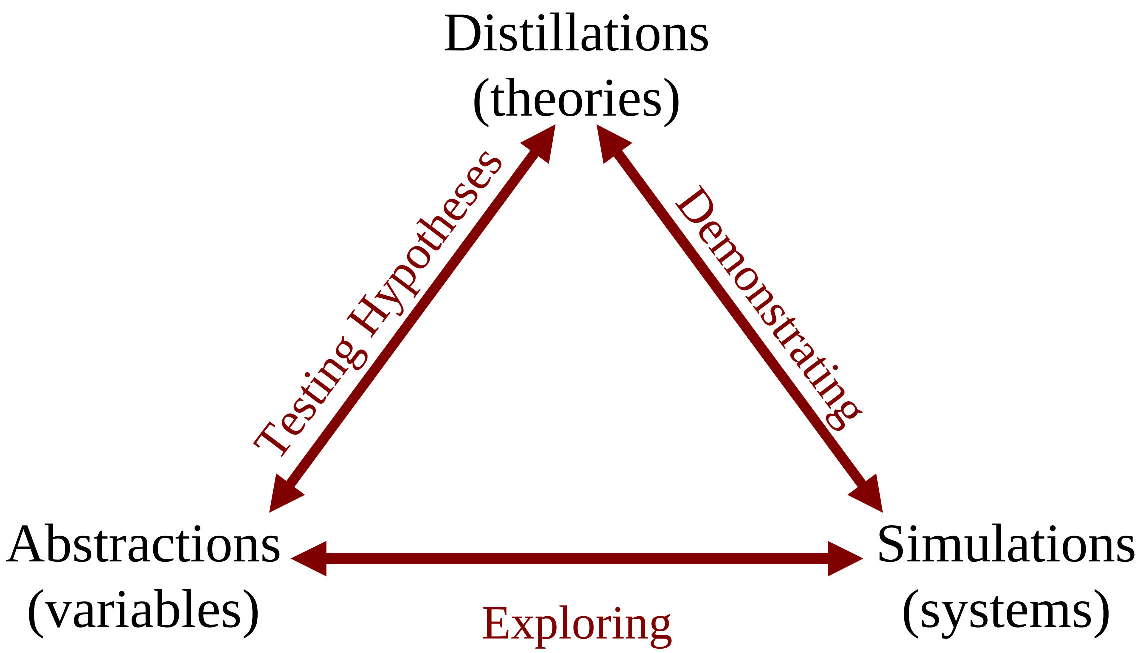
SCUDHunt is an example of what we would call a distillation. The actions of the players, the scenario, the available systems, and the results are all easily expressed and understood as if the game were the real world. Many commercial games have military themes and are targeted at a mass market. Such games must have relatively simple rules (to be playable by a diverse audience), and so they can be classed as distillations. One good example of such a commercial game is Hasbro's *Battle Cry*. The even more classic Parkers Brothers game of *Risk!* lies along the continuum between abstraction and distillation. Introductory games designed more for the hobby-oriented market than a mass market (games such as The Avalon Hill Game Company's *Tactics II* or *War at Sea*) may have somewhat more complicated rules and mechanisms, and so lie along the next continuum, that between distillations and simulations.

Simulations

This last category, simulations, represents the upper end of the game spectrum. Because they are still games, simulations remain reductions of the real world. But a simulation attempts to reduce that reality to a much more detailed model. Representational forms are endowed with far more, and more richly represented, characteristics of their real-world counterparts than found in distillations. In addition, simulations tend to represent more intangible elements of actual operations, such as information flow, command and control restrictions, operational decision-making, and logistical constraints and effects. Simulations tend to be much more quantitatively based and modeled than other forms of games. Traditional DoD warfare gaming systems such as the Enhanced Naval Warfare Gaming System, JANUS, or SIMNET are at the upper level of simulations. Commercial board wargames and computer wargames focused on the limited hobby market (Avalon Hill's *Advanced Squad Leader,* Microsoft's *Close Combat: A Bridge Too Far,* Jane's *Fleet Command*) push the envelope from distillation to simulation.

Games and experimentation

As is the case with any attempt to quantize a continuum, this notion of wargames as abstractions, distillations, and simulations glosses over the fact that most games will share elements of at least two of these groups. Indeed, the three classes of games we have defined above clearly flow into each other. As a result, though each class is perhaps most useful for specific elements of experimentation, there are some strong overlaps, particularly if you look at the adjacent pairs of types.



Exploring, testing, and demonstrating

Abstractions and simulations are useful for the exploring type of experimentation. Abstractions can help you explore the existence and relative importance of fundamental variables and their relationships as you begin to articulate the basic principles of a field of research. At the other end, simulations can help you explore how well systems or other practical implementations of principles you have derived from theory may actually work in the real world and lead you to develop new ideas to explore. The series of wargames played at the Naval War College during the 1920s and 1930s helped the Navy explore many of the operational and support concepts that proved critical during World War II.

Abstractions and distillations are useful for testing hypotheses. Again, abstractions tend to focus on fundamental variables. Distillations, on the other hand, are most usefully applied to test theories about relationships among those fundamental variables. The current experiment conducted by the WARD is studying how the ability of teams to create SSA can be affected by things like their C2 structure, their modes of communication, and their techniques of sharing their visualization of the operating environment. This experiment—and other research using *SCUDHunt* and related test beds—seeks both to explore variables and to test theories.[11]

Finally, distillations and simulations can be very useful for demonstrating relationships and concepts derived from theory or invented from whole cloth. In the military environment, simulations in the form of wargames or field exercises have played a major role in such demonstrations for decades, a prime example being the fleet battle problems carried out by the U.S. Navy during the period between World Wars I and II. The use of the SIMNET system to "recreate" the

---

11. See Perla, *Gaming and SSA*, and Anthony Dekker, *C4ISR Architectures, Social Network Analysis and the FINC Methodology: An Experiment in Military Organizational Structure (Revised)* http://www.dsto.defence.gov.au/corporate/reports/DSTO-GD-0313.pdf

actions at 73 Easting during the Gulf War is a more recent example.[12] A CNA project for the Navy-Marine Corps Intelligence Training Center (NMITC) demonstrated how the use of a commercial computer wargame could allow instructors to demonstrate the effects and relationships between intelligence preparation of the battlefield and subsequent tactical operations over that battlefield, providing a useful practicum to supplement standard classroom lectures and case studies.[13]

Games, particularly wargames or wargame-like games, are useful test beds for a wide range of scientific, technical, and military experimentation of all three categories: exploration, hypothesis testing, and demonstration. The next section looks in a little more detail at using such games for experimentation in C2 research.

Experimentation in C2 research

If we were developing a new science of command and control from a complete state of ignorance, we might expect to employ the three classes of experiments described in the previous section—to explore, to test, and to demonstrate—in that sequence. We might play around with the physical phenomena of interest, trying to learn enough to begin asking specific questions and proposing some specific hypotheses. Then we could formulate those hypotheses rigorously and conduct formal tests of them, using experiments tailored to answer those specific questions. The answers we get then may suggest some practical applications of our new knowledge. We could design some new apparatus or process or way of operating. We might then conduct an experiment to see whether that new technique actually produces the results we expected. And so we begin the process anew, exploring why the results we obtained took the form they did.

---

12. Jesse Orlansky and Colonel Jack Thorpe, eds., *73 Easting: Lessons Learned from Desert Storm via Advanced Distributed Simulation Technology, IDA D-1110* (Alexandria, VA: Institute for Defense Analyses, 1992).

13. CNA Research Memorandum D0002917.A2, *Integrating Wargaming into the NMITC Curriculum: Summary and Recommendations*, by William D. Brobst *et al.*, Dec 2000.

In the field of C2 research, or even in the sub-field of research into SSA, we are not starting from scratch. Much is known (or at least presumed) about some of the variables, theories, and systems of interest. Other aspects of C2 (particularly associated with the first C, command) seem virtually unexplored. Not surprisingly, in this sort of environment in flux, we see a range of experiments, of all three types, as researchers pursue different lines of inquiry.

The Naval War College's Warfare Analysis and Research Department has chosen to focus on some particular directions in regard to C2 research. In part, its efforts are intended to provide insights to the designers of Global Wargame 2003 about how they might best incorporate modern C2 ideas, particularly those associated with network-centric operations, into the structures, processes, and technologies used to build and support the game. Even within this narrower focus, the WARD faces the issues of multiple levels and goals of experimentation. We can see the same dilemma in the broader C2-research environment of DoD and other military establishments around the world.

At the top level of game design, the designers of Global need to conceptualize how the players will be grouped into organizations to play future Global War Games. They must define the processes through which the players will make decisions, communicate those decisions to other players and game controllers, and monitor the execution and outcomes of those decisions. The designers must also integrate available technologies to support both the organizations and the processes.

At the bottom level of research, the WARD is attempting to explore and test hypotheses about some of the fundamental variables and theories associated with future C2, such as SSA. But these basic investigations must move rapidly and effectively in the direction of providing practical insights and recommendations to the designers of the Global series.

There is, of course, a middle level of work that must be carried out, similar in kind to the notion of games as distillations, as described in the preceding section. This middle level of research must focus on building a simplified structure based on the top-level representation of an actual, working command and control system as used in the

Global War Game or other operational/strategic wargames. This middle-level game structure can then be used to expand the tests of effects of basic ideas derived from the low-level exploratory work. This level is essential in connecting the low-level emphasis on fundamental variables and the high-level emphasis on working C2 systems. The mechanism through which this middle-level game environment can achieve such linkages lies primarily in the development of simplified, yet still sufficiently rich, mathematical models that link fundamental variables like SSA to complex command and control systems or concepts like NCO. At this level, exploration and hypothesis testing go hand-in-hand, interweaving abstractions and distillations to help formulate ideas for a simplified approach to representing key elements of warfare in a more wargame-like environment such as Global.

We designed *SCUDHunt* originally to explore a limited set of variables and to test simple hypotheses about them. The basic elements of the game system focused on the way teams acquired, processed, and communicated information among the members of the team. We kept their decision-making environment deliberately simple. The emphasis was on tasking the players to build a picture of where the targets were hidden by sharing information about the results of searches carried out by the sensors controlled by each team member. The players might also use various means to cooperate in the placement of those sensors so that they could coordinate their search of the entire target area of interest. At the end of the game, the individual players were tasked to make recommendations about which possible target locations to strike to kill the SCUD launchers with the least collateral damage (that is, without attacking many locations that did not, in fact, contain a SCUD launcher).

This basic construct has proven to be gratifyingly adaptable. During 2001, the Management Information Systems Department and the Center for the Management of Information of the University of Arizona used *SCUDHunt* in an experiment exploring leadership, trust, and situational awareness for the Army Research Institute. More recently, the Army Research Institute teamed up with George Mason University to run some new *SCUDHunt* experiments to explore issues associated with training soldiers to work together and share information more effectively in distributed environments. This experiment compared the effects of training troops only in their own skill set relative to cross-training them in the skills and tasks of other team members. Anthony Dekker, of the C3 Research Centre of Australia's Defense Systems and Technology Office, adapted *SCUDHunt* for use as a test-bed in applying a methodology for evaluating C4ISR architectures.[14]

The WARD's current *SCUDHunt* research was designed to explore how different "command styles" (command by plan, command by direction, command by influence) and techniques for sharing the information and interpretations that compose SSA ("shared visualization" or "push visualization") may affect the extent of such SSA and the quality of decision-making that SSA may allow a team to make. This research also can give us insights into the underlying dynamics and processes of how human beings interact with information, decisions, and other people. Such insights are crucial if we are better to understand and model such interactions. And it is this process of modeling that is critical to making scientific progress.

14. Anthony H. Dekker. *C4ISR Architectures, Social Network Analysis and the FINC Methodology: An Experiment in Military Organizational Structure,* DSTO Report DSTO-GD-0313.

The NWC experiment

The specific experiment that the WARD conducted in the spring of 2002 focused on exploring and testing hypotheses about the effects of C2 methods on shared situational awareness of teams. The WARD is reporting in full on this experiment under separate cover. This section documents some of CNA's support for the experiment, particularly the statistical design and analysis of the main experimental data.

Experimental goals and structure

Based on the discussions and distillations of the planning conferences associated with this effort, the fundamental working hypothesis is that SSA is a function of command attributes and collaboration techniques. In particular, two key questions are:

- Does command style affect the development of SA in teams?
- Does collaboration method affect the development of SA in teams?

Both these questions are of interest for both collocated and virtual (or distributed) teams.

To explore these issues and answer these questions in a specific setting, the WARD chose to use a version of the *SCUDHunt* game, as well as the basic statistical design employed in CNA's earlier work for DARPA.[15] In both experiments, teams of four players played the game. Each player controlled one or more search assets of varying ability to detect SCUDs. False positive results were possible. Players made their moves by designating which squares of a 5 x 5 grid their assets would search on that turn, subject to certain restrictions on their placement.

15. Perla, *Gaming and SSA*.

Once all players had designated their search squares, the system responded with the results of those searches. These results took the form of symbols representing three possible results: nothing detected, evidence of vehicles (possibly SCUD launchers) detected, and strong evidence of a SCUD launcher detected. Once all detection information was complete, players could discuss among themselves the results and which sensors had provided which results. They were also tasked with providing strike recommendations based on their best estimates of where the SCUD launchers might be, given the current state of their information.

The players knew for certain that three SCUD launchers were in play, and their goal was to find all three of them. The game was to last for five or six game turns. At the end of each turn, players were to submit their strike recommendations. For each turn of the game, a "shared awareness score" was calculated in the manner of the DARPA study. This score was simply the ratio of the total number of target squares recommended by the players during the "strike plan" phase at the end of the turn and the total number of distinct squares in the overall set. This measure produces values ranging from 1 if no players recommend the same target squares, to 4 if all players recommend the same set of targets.

Although SSA is the primary element that we originally designed *SCUDHunt* to investigate, SSA alone is not the only measure of interest. The NWC also sought to learn about the effects of variables—and of SSA—on the ability of the players to make accurate decisions about the locations of the SCUDs. As a first-order measure of accuracy, we chose to use the fraction of strike recommendations made by the players that were actually targeted on squares containing SCUDs.

Variables studied

Specific variables to be explored in this experiment were characterized as command styles and collaboration mechanisms.

Command style was of three types:

- Command by direction: A fifth player, a commander, gave specific orders to each of the four sensor players for where to place their assets each turn.
- Command by plan: An overall plan was promulgated by the control group acting as a higher command authority, with branches and options for how the sensor players were to proceed with their search, leaving them with some flexibility in how they would implement the plan.
- Command by influence: An overall mission was defined (in simplest terms, to find the SCUD launchers) and the players were left free to coordinate among themselves about how best to carry out that mission—this command style was very like the basic free-play approach used in the DARPA experiment.

These alternatives were defined on the basis of concepts defined and described in van Creveld's seminal work *Command in War*.[16]

Collaboration mechanisms were of two types:

- Shared visualization (including text chat)
- "Push visualization" (also including text chat).

The visualization techniques differed in subtle ways. The shared visualization tool was part of the original *SCUDHunt* game system. It allowed all the players a visual representation of the *SCUDHunt* game board, showing the results of each search for each turn of the game (though the source of the particular search result was not indicated by the mechanism). The "push visualization" tool was a new mechanism developed and implemented by the WARD. This application allowed the players to enter graphic symbols into a display of the *SCUDHunt* game grid. Symbols like circles, squares, and triangles represented the extent to which a player believed that a given grid square contained a SCUD. There were four symbols, representing No SCUD,

16. Martin van Creveld. *Command in War*. Harvard University Press. Cambridge, Massachusetts, 1985.

Possible SCUD, Probable SCUD, and Confirmed SCUD. In addition, each player had an identifying color, so that players could distinguish which of them had placed which symbols on given squares. Note that this display was NOT generated automatically by the system. The players had to take positive action to provide this information to their team members. However, each player did receive the resulting display automatically, hence the characterization as a "push" system.

For experimental purposes, the two variables were combined to form six "factorial treatments" (that is, each treatment was composed of two factors, command style and visualization technique). This combination of treatments allowed the overall experiment to follow the same Latin Square statistical design as in the DARPA experiment, thus facilitating both the design of the experimental procedures and the subsequent data analysis.

The Latin Square is an experimental design that allows for efficient use of experimental test subjects when there is some concern about the possible effects of extraneous variables on the analysis of the factors of interest. In this case, there was some concern that as teams played the game repeatedly, a learning effect might contaminate the effects of the command and collaboration variables. By using a Latin Square, we hoped to control for such effects.

The resulting Latin Square design called for 6 distinct player teams (each composed of 4 players) to play 6 games each, once with each combination of the 3 command and 2 collaboration factors. The resulting design matrix for the experiment took the following form:

Design Matrix

	GAME G1	G2	G3	G4	G5	G6
Team 1	**B**	**E**	**A**	**C**	**F**	**D**
Team 2	**D**	**A**	**E**	**B**	**C**	**F**
Team 3	**E**	**B**	**C**	**F**	**D**	**A**
Team 4	**A**	**F**	**D**	**E**	**B**	**C**
Team 5	**F**	**C**	**B**	**D**	**A**	**E**
Team 6	**C**	**D**	**F**	**A**	**E**	**B**

A- Cmd by Direction, Shared Viz &Text
B- Cmd by Direction, Push Viz & Text
C- Cmd by Plan, Shared Viz &Text
D- Cmd by Plan, Push Viz & Text
E- Cmd by Influence, Shared Viz &Text
F- Cmd by Influnce, Push Viz & Text

Each of the 6 factorial treatments is represented by a letter, with the meaning of each letter as shown below the above matrix.

The experimental procedure had each team play the sequence of games under the conditions defined by the above matrix. For example, the first game played by team 2 used treatment "D," or command by plan and push visualization (with text chat). The second game played by that team used treatment "A," command by direction and shared visualization.

Experimental measures

As described earlier, we calculated two broad sets of measures for each game, shared awareness scores, and accuracy scores. We calculated these scores for each team for each turn of each game. The scores form the basis for the statistical analysis of the effects of the factorial treatments. There are several possible ways of employing the scores as measures. One approach is to use the final score, that is, the score at the end of the final game turn,[17] as the overall measure of how well the teams were able to build and maintain their SSA or how well they were able to target the SCUDs. Another approach is to look at the average of the scores over the full game. In the case of SSA scores, we used both methods; in the case of the quality (or accuracy) scores, we use only the end-game score.

Sanitized version: removed pages 26-31

17. Most games were played for a total of 5 turns. A few of the earliest games played went on for 6 turns. In those cases, we have used only the data for the first 5 turns of the game, to maintain some consistency with the later games.

Caveats

The results of this experiment need to be interpreted with considerable caution. Some of these caveats stem from technical and logistical difficulties that arose during the execution of the experimental plan.

Other issues arose because of the way that the different command styles were operationalized during play.

Players and observers seemed to agree that the players would have benefitted from additional familiarization training, both with playing the *SCUDHunt* game itself and with using the separate push-visualization application. Under the best of circumstances it is difficult to ensure that all players began the experiment with a fully adequate understanding of how the game worked or of what they were expected to do. Observations indicated evidence of some problems in the way players understood the game program and visualization tools. There were cases of players exhibiting different understandings about how to interpret the command plan, about the capabilities of the various sensors, and about how to use and interpret the various visualization methodologies.

The latter seemed particularly the case with the push-visualization application early in the sequence of game play. On more than one occasion it appeared as if some players restricted their use of the push-visualization tool to record results of searches of their own assets rather than the player's overall view of the battle space.

One observer noted that: the players sometimes seemed to be unsure of how to interpret results of searches in the face of the differing of quality of the sensor information. There was no precise and easy way of discerning true detections from false positives other than comparison of the qualitative assessments of sensor performance and repeated searches. If players lacked a solid grasp of their operational environment in the game, it may prove difficult or impossible to distinguish the effects of the specific command styles from the confusion the players experienced.

Problems with the ability of the players to understand and play the game effectively were exacerbated by a series of technical problems that arose during execution. These problems included unreliable or slow internet connections, disconnects, and hang-ups or stalls in the operation of the *SCUDHunt* program or the push-visualization application. All participants maintained a positive and professional attitude, and managed to overcome most of these problems. Nevertheless, the frustrations resulting from the technical problems and a tendency to play as fast as possible when the system seemed to be working reliably may have had unpredictable effects on the results.

During the DARPA experiment,[19] results and observation of play indicated that teams who tended to "bond" together quickly may have exhibited more effective play techniques. During this NWC experiment, team bonding was hindered by time constraints on game play.[20] Players were under pressure to finish the games in the time available, and the initial allocation of time to play each game proved to be insufficient for a measured pace of play. The effects of these time constraints were most severe in games using command by influence, the command style that may be particularly sensitive to interpersonal dynamics.

Another issue affecting the interpretation of the effects of command style is the fact that implementation of the different styles may have been inconsistent across academies, teams, and games. If experimental treatments were not applied in a standardized way, then it is difficult to compare the treatments in the manner required for the analysis. Extraneous sources of variability such as this can create more experimental "noise," possibly obscuring any actual effects of the design variables (i.e., a "signal") that may be present in the pattern of results.

The NWC report on the experiment may discuss these and other issues of experimental practice in more detail. For our purposes, it is enough to note the potential for experimental conditions to make the statistical analysis of results we present here less reliable than they might be had the experiment gone off flawlessly according to plan.

That said, however, this experiment does indicate that *SCUDHunt* can be used to explore the effects of command style on the development of SSA and the making of good decisions. Additional exploration of the data may yield further insights. A single-factor experiment, hold-

---

19. See Perla, *Gaming and SSA.*

20. Initially, players were limited to 45 minutes of play for a particular game. If they exceeded that limit, the game was shut down. This constraint was later relaxed so that not all games were played with the same condition.

ing visualization tools constant (using the inherent shared-visualization tool of the *SCUDHunt* on-line game) and varying command styles as attempted here, seems both an eminently doable experiment and one that may confirm or deny the most interesting speculations that flow from the current analysis.

Some future directions

The NWC's *SCUDHunt*-based C2 experiment demonstrated the benefits of game-based experimentation. Such research offers scientifically based and statistically valid results that can answer practical questions about human performance in C2-related tasks. These insights are crucial if future wargames are to improve their representation of such operational concepts as network centricity, information warfare, and self-synchronization. Modeling those interactions to test hypotheses about key factors is critical to making scientific progress.

If future wargaming systems and technology are to enable researchers to explore the effects of different C2 structures and processes, we need to have enough theoretical understanding of the key elements and dynamics of those processes to guide and supplement other information sources. These sources are of two types: *historical research* into understanding what happened (and why) in actual (or simulated) experience; and *analysis*, to extrapolate from the limited experience available to build general models that support wargaming for all three of our experimental objectives.

The art of the wargame designer is to convert understanding of the decisions commanders make, of the information commanders use to make decisions, and of the variables that determine the outcomes of those decisions, into a working operational model of reality. In the past, wargame designers have focused their talents on building simulations of increasing detail and "realism." But detail is not always the same as realism. Even the most complex modern simulation is not real, and the price we must pay for this flawed mirror of reality—paid in terms of time, and money, and effort to build it—exceeds our ability to repeat its use enough to generate useful statistical data. In addition to such grand and expensive tools, it is time for us to create some simple abstract games and tailored distillations appropriate to the kind of scientific research this NWC project has embodied.

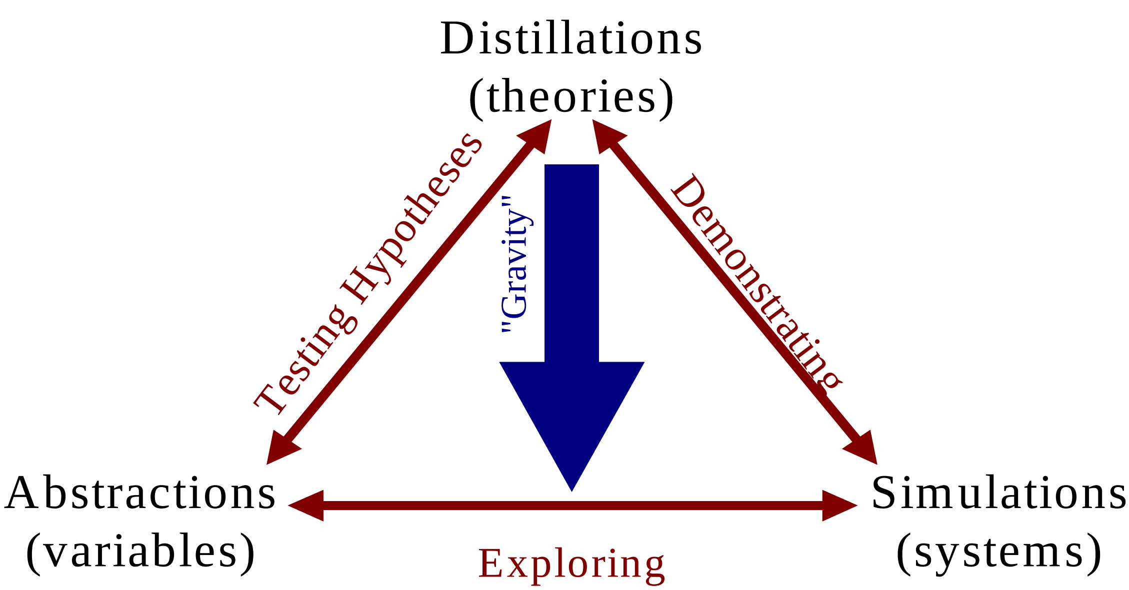
This project highlighted and confirmed the potential value of game-based experimentation as a scientifically rigorous approach to exploring fundamental command and control issues. It also demonstrated the power of using a virtual organization to do so. The WARD was the core and engine of the project, but the involvement of an FFRDC (Federally Funded Research and Development Center) (CNA), a contractor (ThoughtLink, Inc., or TLI), and two service academies (the U.S. Naval Academy and the U.S. Air Force Academy) created a unique virtual team for pursuing this research. WARD provided the impetus and initiative for the project and a link to broader issues of importance through the overall NWC program and Global War Game. CNA and TLI contributed subject-matter expertise in the science and analysis, and in the design, hosting, and execution of the game. The service academies provided players to run the experimental regimes and also integrated the games and experiment into their educational program. The shared research interests of the organizations helped create a situation in which each member of the team not only benefited from participation, but also helped advance the research agendas of all team members.

We need to maintain the momentum. *SCUDHunt* is a simple game, designed to meet limited objectives. What we need now is the capability to develop and support a range of games designed at the levels of abstraction and distillation to supplement the overemphasis on simulations. We need laboratories for game-based experiments. Researchers can use such laboratories to conduct rapid, focused, and affordable experimentation to study variables and issues of high interest.

Such variables and issues may arise from a wide range of sources. Some fertile sources include current military operations, exercises, or simulation games of many different types. The Naval War College's Global War Game series is the quintessential large-scale simulation of the type that can and does generate such issues. Indeed, much of the impetus for the current experiment arose from the experiences of Global 2000 and Global 2001, as described earlier.

We can argue that the steps of an ideal scientific approach to using game-based experimentation begin with the use of abstract games to explore reality and identify important variables. Distillations can help us elaborate and scientifically test operational hypotheses about relationships among these variables. Simulations can help us demonstrate how those relationships work in actual situations. The reverse flow—working from simulation back through distillation to abstraction—is also fundamental to a scientific approach to dealing with these issues.

Large simulations raise issues that need answers beyond the capability of the simulations to provide. As was the case with Global, the costs and manpower requirements of playing such large simulations preclude iteration to generate statistically significant amounts of data in reasonable spans of time. Distillations like *SCUDHunt*, on the other hand, provide a way to pose questions as testable hypotheses. We can explore issues and test ideas and concepts far more efficiently, faster, and cheaper using well-designed distillations than we can by using elaborate and expensive simulations. The results of those tests may lead to a redefinition, or improved understanding, of fundamental variables. This new perspective, in turn, can call for even more basic research using abstract games. Although all three categories of games can be useful over a range of topics, each is best tailored to specific component parts of the overall task.



The arrow labeled "gravity" represents a danger we must protect against. Game designers are all too familiar with the tendency—almost a force of nature—for games begun with a focus on this elusive middle ground to slip (at times inexorably) into becoming an over-complicated simulation. The reasons for this tendency include:

- The drive by designers to make the game more "realistic" and more "sophisticated"
- The drive by players to make the game more similar to their everyday routines and so (they think) more relevant to their everyday needs
- The drive by bureaucrats to make the game a bigger more important program to enhance their own prestige and perceived power in the organization
- A healthy dose of "gee whiz wouldn't it be cool if" on all levels.

At the other end of the spectrum, some designers have a penchant for overdoing the necessary design process of making "powerful simplifying assumptions." Such designers focus on the essence of their own perception of the fundamental decision problems players should address in the game, stripping away secondary issues to a greater extent than may be wise. When carried to extremes, such tendencies can transform distillations into abstractions—potentially valuable in their own right, but removing some of the operational richness that may allow for a broader look at issues the designer may undervalue. In some ways, the evolution of chess exhibits this tendency.[21]

The dangers are that game designs, and research programs depending on them, can drift away from their most useful applications and cross into the dangerous waters of promising more than they can deliver, or delivering less than the sponsors of the research expect.

From that perspective, what's missing most in the application of gaming to DoD's C2 research is a coherent program to apply distillations and abstractions to explore issues arising from simulations, and then to feed the results back into the simulations—to demonstrate

21. See Perla, *Art of Wargaming*, for a brief description of this evolution.

and explore the operational implications of the scientific insights. The current effort by the WARD is the beginning of such a process. By connecting a research program at the WARD with the Global game, the NWC is ideally positioned to bring all elements of the process together effectively.

The current project has had some additional benefits. By integrating the experimental efforts with the service academies, this project was able to draw on a ready pool of game players to allow the experiment to achieve the critical mass of participants necessary for the analysis desired. In the process, the academies received educational materials to enrich the experiences of their students. They also benefited from the practical experiences associated with playing the game and experiencing firsthand some of the elements of C2 the game was designed to explore. This same collaboration between the game-designers and non-military academic institutions—reflected in research at the University of Arizona and George Mason University in Virginia—indicates another potential beneficial fallout from the use of game-based experimentation. Such collaborations provide the schools with research opportunities of their own, as well with materials for both instruction and practicum.

How can we continue to build on the foundations laid in this project? There seem to be a couple of opportunities worth pursuing immediately.

First, we should document the principles and practical experiences we have gained from both this experiment and CNA's and ThoughtLink's earlier *SCUDHunt*-based research. One suggestion, made only partially in jest, is to combine such material into a publication titled *Game-Based Research for Dummies*. Such a publication can help other researchers begin to apply this technique from a baseline benefiting from our hard-won experience and guidelines.

Second, it's time to begin building the sort of game-based research laboratories this paper has described. Such laboratories are less a specific facility than they are assemblages of critical components and expertise—the games to serve as experimental test beds; the game designers to create the games; and the analysts and scientists to formulate the problems, design the experiments, and analyze the data.

Such laboratories could be “virtual” organizations, bringing together subject-matter experts from across the United States and other nations as well. A “virtual, distributed laboratory for game-based experimentation” can help advance our understanding of command and control, information operations, network-centric warfare, and other critical concepts that, in many ways, remain buzzwords rather than realities.