

Contending with complexity: **Developing and Using a Scaled World in
Applied Cognitive Research**

Brian D. Ehret

Wayne D. Gray

George Mason University

&

Susan S. Kirschenbaum

Naval Undersea Warfare Center Division Newport

VERSION ACCEPTED BY JOURNAL

PUBLISHED VERSION IS SLIGHTLY DIFFERENT

Ehret, B. D., Gray, W. D., & Kirschenbaum, S. S. (2000). Contending with complexity: Developing and using a scaled world in applied cognitive research. *Human Factors*, 42(1), 8-23.

Please send all correspondence to:

Wayne D. Gray

George Mason University

MSN 2E5

Fairfax, VA 22030

1-703-993-1357

gray@gmu.edu

Abstract

Scaled worlds preserve certain functional relationships of a complex task environment while paring away others. The functional relationships preserved are defined by the questions of interest to the researcher. Different scaled worlds of the same task may preserve and pare away different functional relationships. In this paper, we use the example of Ned to discuss the use of scaled worlds in applied cognitive research. Ned is based on a detailed cognitive task analysis of submarine approach officers as they attempt to localize an enemy submarine hiding in deep water. For Ned, we attempted to preserve the functional relationships inherent in the approach officer's information environment while paring away other aspects of his task environment. Scaled worlds attempt to maintain the realism inherent in the preserved functional relationship while being tractable for the researcher and engaging to the participant. [138 words]

Contending with Complexity: Developing and Using a Scaled World in Applied Cognitive Research

Introduction

Commanders of 21st-century submarines will be given computer workstations that will allow them to directly query and receive data regarding hostile, friendly, and neutral targets. (In submariner parlance, all objects in the ocean other than one's ownship, are referred to as targets.) Understanding the information needs of the submarine commander in his role as Approach Officer (AO) is vital to the design of this workstation and is the goal of Project Nemo. This project has required a multi-phase approach that combines the applied tools and of human factors practitioners, the microworlds approach used by those who study dynamic decision making, and a theoretical perspective drawn from cognitive science.

Project Nemo is a long-term research project that, as of this writing, is ongoing. The first phase of the project involved analyzing the nature of the expertise used by AOs in localizing enemy submarines hiding in deep water. Our trials and tribulations in analyzing this novel expertise are documented in Gray and Kirschenbaum (in press). The second phase, documented by this paper, involved using the analyses from the first phase to build and deploy a scaled world. Both papers are in keeping with the themes of the special issue in that their focus is on research methods, not research conclusions. Those interested in the research conclusions of the first phase are directed to Gray, Kirschenbaum, and Ehret (1997) for a short treatment and to Kirschenbaum, Gray, and Ehret (1997) for a detailed account. Those seeking research conclusions from the scaled world phase must await future publications.

This introduction concludes with two overviews. The first contains a brief description of the motivations of Project Nemo and the AO's task domain. The second discusses the role of scaled worlds and the dimensions on which they differ from other simulated task environments.

Overview of Project Nemo and Situation Assessment

The AO performs the role of senior decision-maker during an encounter with a hostile target. His job in identifying and locating enemy submarines is difficult, interesting, and important. It is difficult because it requires locating an enemy who is hidden in a vast and acoustically uncertain ocean environment. It is interesting to cognitive scientists as the expertise of AOs is similar to but different from other, better studied, types of expertise. Finally, whatever the changing nature of warfare, locating hostile targets in an uncertain ocean environment is an important job that is important to do well.

The U. S. Navy is designing a new attack submarine that, among other innovations, will feature a reduced crew and a command workstation for the AO. The results of our project will help inform the design of the command workstation so that the procedures used by AOs in problem solving are supported and facilitated by the workstation. A

prerequisite to building such reasoning-congruent interfaces (Merrill, Reiser, Beekelaar, & Hamid, 1992) is a deep understanding of the cognitive procedures and memory structures used by the AO in his task and how these information processes react and interact with external events.

With these motivations in mind, our task is not to study how well submariners do, but, rather how they do well; that is, cognitive process, not strategic outcome. Specifically, our project goal is to derive a step-by-step picture of the information processing involved in one critical stage of the AO's job, situation assessment.

Situation assessment begins after the hostile target is detected. It entails localizing, or determining a solution for some surface or subsurface target. A solution identifies the target's distance and bearing to one's ownship, as well as its course and speed. It also entails determining the interrelationships among all targets – hostile, friendly, or neutral. For example, is the hostile submarine getting into a position from which to attack a friendly merchant ship? Or, can I move ownship into a position where, if I have to attack the hostile submarine, I will not risk damage to the merchant? When the AO is confident of his solution, he goes on to the firing point procedures stage of the exercise, which may involve either shadowing or destroying the hostile target.

Scaled Worlds and Other Simulated Task Environments

In field research, there is often too much [complexity] to allow for any more definite conclusions, and in laboratory research, there is usually too little complexity to allow for any interesting conclusions (Brehmer & Dörner, 1993, p. 172).

Scaled worlds are one reaction to this complexity. Beginning with a complex, task environment a scaled world seeks to preserve certain functional relationships, while paring away others. There can be multiple scaled worlds of the same task environment that differ on which functional relationships are preserved and which are pared away. The nature of the research question determines what is kept and what is removed.

Other classes of simulated task environments include high-fidelity simulations, synthetic environments, and microworlds (e.g., Brehmer, 1992; DiFonzo, Hantula, & Bordia, 1998). These environments differ from each other, and from scaled worlds, along (at least) three dimensions: tractability, realism, and engagement (see Figure 1).

----- INSERT FIGURE 1 ABOUT HERE -----

Tractability is the “complexity” issue referred to by Brehmer and Dörner (1993): can the researcher use the simulation to pursue the question of interest? Tractability includes concerns such as collecting the right data, at the right grain size, with the right timestamp. It also concerns whether participants can learn the simulated task environment in a “reasonable” amount of time. The usability of the simulated task environment is a concern as well. If the design of the interface interferes with the conduct of the task, then the data collected have more to do with usability problems than with issues for which the simulated task environment was built.

In our research, tractability extends to a fourth concern: can the computational cognitive models we write interact with the simulated task environment to perform the task. Models and humans must be able to receive and query the same information from the same simulation.

Tractability is a relative dimension defined by the research question. What may be a tractable simulation for one set of research questions may not be tractable for another. For questions that focus on the flow of information to and from the decision maker, we would expect high-fidelity simulations of complex systems to be almost as intractable as the real-world they simulate. In contrast, as there are few constraints on how a microworld or synthetic environment is constructed, these simulated task environments can be built to the researcher's specifications (therefore, if these are not tractable, they have been built wrong). The tractability of a scaled world would be some place between these extremes. Exactly where a scaled world would lie on this dimension depends on which functional relationships need to be preserved and which may be pared away.

The second dimension is realism. The simulated task environment is realistic to the situation to the extent that experiences encountered in the simulated environment occur in the real task environment (see also DiFonzo et al., 1998). Scaled worlds should be more realistic than a microworld or synthetic environment designed to investigate the same functional relationships, but less realistic than a high-fidelity simulation. Scaled worlds are single-mindedly focused on preserving certain functional relationships from the task environment. Maintaining these functional relationships maintains a type of realism. In general, builders of scaled worlds may try to maintain the realism of other aspects of the task environment unless such realism interferes with the tractability of the research questions of interest.

The third dimension is engagement. Engagement describes something about the participant's motivation. Participants may be engaged because we pay them money to do well. They may be engaged because they view the simulated task environment as an interesting game that they like to play. Alternatively, they may be engaged because they have deep knowledge of the real task environment and believe that it is interesting and important. In this case, the scaled world provides a sketch of the real task environment, whereas the participants' deep knowledge provides details that are missing from the sketch and, in this way, may enhance the realism of the scaled world.

The engagement dimension is an interesting one. It can be argued that, for most research questions, if microworlds and synthetic environments are not engaging then they have not been built or used correctly. However, not all real-world tasks are especially engaging. Hence, any high-fidelity simulation of those tasks may not be engaging either. In contrast, designers of scaled worlds have an opportunity to "improve" on the task environment by paring away non-engaging aspects. If such a course is chosen, it must be followed with care and with deep knowledge of how the non-engaging aspects of the task environment affect the functional relationships of interest.

Overview of this Paper

The following section provides an overview of Ned and its displays. This section is followed by a discussion of the purpose of Project Nemo and the goals for which Ned was built. The discussion of goals and purposes is followed by a quick summary of the cognitive task analysis that formed the basis of Ned and of the conclusions from that analysis. The heart of the paper discusses the building of Ned and the issues and tradeoffs for Ned between tractability, realism, and engagement. The data set currently being analyzed comes from 36 AOs. Demographics of this participant population are discussed in the *Data Collection Using Ned* section. The penultimate section focuses on *The Future of Ned*. The paper concludes with a discussion of scaled worlds as task analysis.

Ned: A Scaled World for Project Nemo

The target audience for Ned consists of submarine AOs. These are typically either the executive officer or the commanding officer of the submarine, though we have used Ned with junior officers as well.

Ned users must have a great deal of knowledge just to understand its displays, let alone to use Ned to complete the task scenarios. For example, once we tell our participants that Figure 2 is intended as an unclassified version of a Narrow Band Towed (NB-towed) sonar display they have no problems understanding the meaning of each display element.

----- INSERT FIGURE 2 ABOUT HERE -----

Ned's displays are generated dynamically. Information and displays are updated every second. The menu item OSC (see the bottom-right item of Figure 2) takes the AO to a display for ownship controls. At the OSC display, the AO can check or change ownship course, speed, or depth. Ned responds dynamically to any change the AO makes. Indeed, for the AO, often the information that comes in while ownship is changing course is more important than the course change per se.

The menu item TMA takes the AO to a display showing the output of *target motion analysis* calculations. The TMA module integrates sonar data over time and outputs its best guess at various target attributes, including bearing, range, speed, and course. A different TMA solution is available for data collected by each of the three sonar sensors, NB-towed, BB-towed, and BB-sphere. AOs are expected to disagree with TMA's estimates of each attribute. AOs override TMA's estimates by entering estimates of their own, based on their reading of the raw data (sonar) and understanding of the tactical situation. Forcing one attribute to a certain value (e.g., overriding TMA's estimate of the target's range by changing it from 30,000 yards to 7,000 yards) results in the recomputation of all other values. Other displays include plots of raw sonar data, such as time bearing (TimeBrng) and time frequency (TimeFreq), as well as a geosituational plot (GEOSIT), and a line-of-sight (LOS) display that is designed to assist in maneuver planning. When the AO has what he considers a good solution for the hostile

submarine, he ends the situation assessment phase by clicking on Main. This signals that he is confident enough in his solution to fire on the enemy submarine (a process not simulated in Ned).

Ned is written in Macintosh Common Lisp 4.2. Running on a Macintosh PowerBook 2400, Ned was used to collect data from 36 submarine officers – an estimated 10% of the active-duty AOs.

Goals for Ned and Project Nemo

Project Nemo focuses on the functional relationships between the query and receipt of task-relevant information and the various ways used by AOs to transform this information. To facilitate the project, Ned has been carefully designed to preserve the functional relationships of interest while paring away most other functional relationships.

There are three goals for Project Nemo. First is to describe and simulate the cognitive processes and memory structures used by AOs in locating enemy submarines. Second is to be able to generalize the findings to a new, unknown submarine, with reduced manning, and new instrumentation. In meeting this second goal, we anticipated the warning of DiFonzo, et al. (1998) that, “generalizability . . . may actually be hindered by excessive attention to mundane realism” p. 283. That is, we believed that the only aspects of the AOs’ task that would not change in the next 25 years (i.e., the time it will take for the new submarine to be designed, built, and launched) are the laws of physics regulating underwater sound propagation. The third goal is a contribution to the study of expertise. The AOs perform a complex, event-driven task that is unlike better-studied forms of expertise. Understanding the AOs’ expertise would be a contribution to cognitive theory.

A Cognitive Task Analysis for Ned

Ned is based on a detailed cognitive task analysis of AOs doing situation assessment. In this section, we provide an overview of the methods used to collect data for the task analysis and of the conclusions of the analysis. A standard discussion of the methods used and details of the results from this phase of Project Nemo are given in Gray, et al. (1997) and Kirschenbaum, et al. (1997).

The process of data analysis was quite long and complex and was hampered by all of the usual problems entailed by a detailed protocol analysis as well as our misconceptions regarding the overall cognitive structure of the AOs’ task performance. The iterations and frustrations of this phase of Project Nemo are painstakingly detailed in Gray and Kirschenbaum (in press), so they will be omitted from this account.

Overview of Data Collection Methods

The charge of Project Nemo was to study the information processing needs of the AO in the context of a situation assessment task (i.e., finding a solution on a hostile submarine). For this charge, studying the AO onboard a nuclear powered attack submarine was both too problematic and too complex. It was too problematic in that both security and expertise issues would prevent observers from noting and recording

even a fraction of the AO's information processing. It was too complex in that the social and organizational dynamics in such an environment would obscure the information processing requirements of the task.

Rather than becoming anthropologists of the nuclear Navy, we chose to use a high fidelity simulation of the ocean environment residing at the Naval Undersea Warfare Center Division Newport (NUWC). NUWC's civilian scientists as well as our military experts (namely, our AO participants) assured us that this simulation, the Combat Systems Engineering and Analysis Laboratory (CSEAL), provided the AO with the ability to query and receive all the information relevant to the situation assessment task.

Data were collected from ten submarine Commanders with an average of 20 years in the Navy. All AO interactions with CSEAL were mediated through a computer operator whom we refer to as the ownship-operator. The AOs were instructed to talk aloud. The CSEAL screen and all AO and ownship-operator dialogue were videotaped. Window events, the current state of the simulation, and all ownship-operator actions were timestamped and saved to a log file.

Overview Task Analysis Conclusions

Many of our problems in analyzing the data stemmed from our attempts to force it to fit various cognitive control structures. Tasks involving expertise are often viewed as forming a wide and deep hierarchy. Each level of the hierarchy contains many alternative steps (width). Each step represents a subproblem and each subproblem can be decomposed into another layer of alternative steps (depth). In contrast to expertise, the accepted wisdom is that, for everyday tasks, the search space is limited. The problem spaces for everyday tasks are either shallow and wide (like choosing a flavor from the menu in an ice cream store) or narrow and deep (like following a recipe from a cookbook) (Norman, 1989).

These preconceptions guided our initial analyses. When our data refused to fit our preconceptions, we were unwillingly led to the realization that most AO actions could be characterized as small steps in a shallow goal hierarchy. However, unlike the everyday task of choosing one flavor from a wide but shallow ice cream store menu, AOs make many successive choices. It is the nature of these successive choices that characterize the AOs' procedural expertise.

The awkward phrase "schema-directed problem solving with shallow and adaptive subgoalings" (SDPSSAS) serves to encapsulate our current theory of how AOs solve the localizing problem. The schema is the task-relevant knowledge accumulated over 20 years of experience as a submariner (half of it at sea). It is a set of declarative as well as procedural knowledge structures. An implication of shallow subgoalings is that the knowledge available to AOs is so rich that steps to supplement this knowledge can be shallow.

The second implication is that the AO solves a series of problems, one every 30 to 300 s. The problem is always the same; namely, "what is the state of the world – NOW." The AO is trying to find a quiet target hiding in a noisy environment while remaining undetected himself. The protocol analysis reveals that he takes a series of short steps

that either (a) assess the noise from the environment or signal from the target – NOW, or (b) attempt to reduce the noise or increase the signal from the target by maneuvering ownship. As shown in Figure 3, these short steps result in shallow subgoaling. When a subgoal pops, the schema is reassessed. The result of this reassessment directs the next step (i.e., selects the next subgoal). This step is accomplished, it returns information to the schema, the schema is reassessed, and so on

----- INSERT FIGURE 3 ABOUT HERE -----

The process of subgoaling is adaptive in two senses. First, the subgoal that is chosen next reflects the current reassessment of the schema. Second, this choice is sensitive to both the long-term importance of the subgoal as well as its recent history of success or failure. Regardless of a goal's long-term importance, AOs will not continue to attempt a goal if successive tries fail. Instead, they will choose another goal and return to the more important goal later.

The dynamic aspect of the AO's task plays an important role in this view of schema-directed problem solving with shallow and adaptive subgoaling. First, the state of the AO's world is continually changing – both ownship and target are moving at a given depth, direction, and speed. For ownship, the value of these attributes can be changed, but neither ownship nor the target can be stopped. Consequently, time is an important part of the picture. Second, subgoals are not accomplished once and then discarded. In the AO's world, subgoals bring in certain types of information or accomplish certain changes to ownship. As the world changes, any given subgoal may be revisited (e.g., DET-BEARING in Figure 3).

Building Ned

CSEAL, the simulation used in data collection, was not designed for behavioral research. Throughout the years, displays had been built that could be considered prototypes of every display currently used by the various crew members onboard a submarine. As a result of this uncoordinated development, CSEAL's displays contained features that limited our ability to infer precisely what information the AOs were getting from CSEAL at any given moment. Simply put, CSEAL was not an adequately tractable environment in which to pursue our research questions. Ned was designed to be a tractable environment. We will first describe the problems with CSEAL and then how the development of Ned resolved or reduced these problems.

Problems with CSEAL

Non-uniqueness of information source. The same information could be obtained from multiple displays. Knowing that the AO had requested or was looking at display A rather than display B did not necessarily allow us to infer that he was seeking information that was unique to display A.

Amount of information per display. The copious amount of information contained by each display encouraged information browsing on the part of the AO. Given the slow nature of both the task domain as well as CSEAL, the AOs could easily monitor a slowly

changing display or text field while their eyes darted around the screen surreptitiously gathering other information. Unless this information was verbalized, we could not track it.

Receipt of unrequested information. Getting from one display to another often required navigating through intermediate displays. In some cases, our protocols yield clear evidence that dwelling on these displays caused the AO to change goals or provided the AO with information that he used shortly after. Unfortunately, we believe there are many more cases of this than for which we have confirming evidence.

Ownship-operator as source of extra-display information. The use of an experimenter as ownship-operator had many advantages, but introduced a source of variance whose influence on AO behavior is difficult to determine.

Each of the above four sources contributed to two problems in our attempt to isolate and identify the information and processing required by the AO for situation assessment. First is the most commonly noted problem associated with verbal protocols: not noting things that were attended to (Ericsson & Simon, 1993; vanSomeren, Barnard, & Sandberg, 1994). The ease of information browsing combined with the amount of information per window and the non-uniqueness of information sources meant that we could never be sure exactly what information an AO was getting from his current display. Likewise, when going from one information display to another, even while talking about his reasons for wanting to get to the requested display, there were ample opportunities for gathering information from the intervening displays. Finally, sometimes the AOs immediately responded to ownship-operator prompts and sometimes they did not. Whereas our encodings reflect the fact that they must have heard these prompts, we can not be sure that they actually attended to and incorporated the information offered.

The second problem is less commonly mentioned; namely, commenting on things that were not used in problem solving. Our AOs chatted all the time; even when, essentially, nothing was happening. For example, while watching and waiting for ownship to complete a turn, while waiting for the display they requested to be brought up, or while waiting for the ownship-operator to make some adjustment to a display. All of what they said was encoded, but much of what they said seems irrelevant to the problem they were attempting to solve at the time they said it. We believe that the source of this problem is the overabundance of information, exacerbated by the slow pace of the situation assessment task, as well as by the ease of information browsing.

The Development of Ned

The development of our own scaled world was driven by two factors: We had a clear focus on the information the AO needed for situation assessment, and we understood how aspects of CSEAL and our data collection methods added complexity to our task as analysts without contributing to the information needs of the AO. Ned was built to allow more precise tracking of the query and receipt of information while minimizing the receipt of nonqueried information.

Our primary design goal was to incorporate into Ned the functional relationships critical to supporting the AOs' information processing needs. We began by examining CSEAL's displays for information fields or display actions not referenced in our protocol encodings. Such fields and actions were candidates for exclusion from Ned. As per the foregoing discussion, however; verbal protocols can be incomplete or extraneous, rendering this strategy merely a first-pass effort. To ensure no critical task elements were excluded, Ned underwent an iterative process of design and review by subject-matter experts. As a result of this process, the number of fields in Ned versus CSEAL decreased by an order of magnitude.

Another goal in Ned's design was to make information fields and interface actions unique to a given display. Aside from some fields that represent the same information in different formats, for example graphically and textually, most other information is unique to a particular display. In addition, Ned's displays are traversed through an always-visible palette menu, which eliminates the problem we noted in CSEAL of having to first open one display to get at the control that will call up the desired display. Likewise, in contrast to the delays inherent in opening displays within CSEAL, for Ned a mouse click on the palette menu closes the current display and brings up the requested display almost instantly.

Ned has three simulation modes. In free-play mode Ned can be used to set up and solve various situation assessment scenarios. In ACT-R mode, computational cognitive models written in ACT-R (Anderson & Lebière, 1998) can interact with Ned in the same manner as a human subject; models can query and receive information and can also take action, such as changing ownship's course. Finally, Ned's data collection mode represents our attempt to further control the twin problems of not noting information that was attended and not attending to information that was noted.

In data collection mode, when a display is open, all of its graphics and numerical values are covered with gray boxes. For example, on clicking the palette menu to open the NB-towed display, the window shown in Figure 2 appears with all information covered except for labels. One gray box covers the waterfall display on the left; each of the information fields on the right is covered by its own gray box. To obtain a numerical value or to see the graphic, the AO must move the mouse to and click on the gray box. The information in that information field will then remain visible until the mouse moves out of that field. Although the effort involved in moving and clicking a mouse has not proved burdensome to the AOs, it has reduced the amount of information browsing (Lohse & Johnson, 1996).

Data Collection Using Ned

Data collection using Ned was conducted by a submarine officer as his masters thesis work at the Naval Postgraduate School (Soldow, 1998). The thesis primarily addressed outcome measures as a function of AO expertise and experience. As such, this represents a use of the scaled world beyond that envisioned by its creators. However, the log files and videotapes that were collected are of intense interest to Ned's creators and currently are being encoded and analyzed.

Thirty-six AOs, at two locations (Hawaii and Bangor, WA) completed the study. (In the U. S. Navy, all submarine officers and crew are male.) Of these, 15 were commanding officers or executive officers, 11 were instructors, and 10 were junior officers; 18 served on attack submarines (SSN) and 18 on ballistic submarines (SSBN). Finally, this group had a mean of 23 years in the U. S. Navy of which a mean of 6.4 years was spent at sea.

As this is a methods paper, not a data paper, we will limit ourselves to one aspect of the data that is relevant in evaluating Ned as a scaled world. In a questionnaire presented after Ned was used, the AOs were asked, "Was the situation 'realistic' enough for you?" The results are shown in Figure 4. The 34 AOs who answered this question gave Ned a mean rating of 3.2, the modal rating was 4. Although Ned is a low-fidelity simulation, these ratings suggest that it was accepted by its target population as realistic enough to accomplish the task for which it was designed.

----- INSERT FIGURE 4 ABOUT HERE -----

The Dimensions of Ned

Based on the design of Ned and informed by the results of data collection, we address the three dimensions on which simulated task environments vary; tractability, realism, and engagement.

The Tractability of Ned

As discussed earlier, tractability is a relative dimension that is defined by the needs of the research question. For our purposes, three categories of tractability are important: data collection, data analysis, and computational cognitive modeling.

Data Collection

Ned was designed to be tractable for data collection. Each interface action, including uncovering of the gray boxes, is timestamped and saved to a log file. The record that is written includes the information contained by the field and the duration in milliseconds that the field was visible. The log file provides us with a fine-grained action protocol of the information AOs' query and receive. The timestamp is accurate to the nearest tick (17 milliseconds). In addition, during data collection an s-video output from the computer and a microphone were used to make videotapes of the screen and the AOs' talk aloud protocol as they solved the various scenarios.

Data Analysis

Ned was designed to be tractable for data analysis. First, little information is common to more than one display. Unfortunately, it was impossible to eliminate information overlap without destroying the realism of some of the information sources. For example, the time bearing display (accessed via the TimeBrng item on the menu shown in Figure 2) is primarily used by AOs to estimate a target's bearing rate (the number of degrees per minute in which the bearing to the target from ownship changes). However, it can be

used to obtain the bearing. The same bearing information is available from the corresponding sonar. However, even with such exceptions, the verbal protocols show that an AO's access of an information source largely implies the subgoal for which the source was accessed.

Second, the data collected from AOs can be played back on a geosituational display. (A geosituational display places ownship in the middle of the display. All target movement is relative to ownship's current position.) This playback mode aids us in comparing the true positions of targets during the scenario with what the AO believed the true position to be.

Third, the information record that is written to the log file has proved amenable to an automated encoding of operators and segmenting for goals. The operator level is the finest grain size that our encodings use. It is the basis of goal and subgoal encodings. A sample encoding of one AO's first encounter during a scenario with the NB-towed display is shown in Table 1.

----- INSERT TABLE 1 ABOUT HERE -----

The automated encoding and segmentation of the log file is approximately 95% accurate with the exceptions falling into three categories. First are the exceptions due to how the log file was written. For example, setting a tracker on a trace entails moving the mouse to the trace of the target on a waterfall display (as in the left side of Figure 2) and clicking on it. The timestamp for setting the tracker is written when the mouse is clicked. In contrast, the timestamp for receipt of information from a field is written when the AO's mouse leaves that field (thereby re-covering the field with its gray box). As Table 1 shows, this method of writing out the log file results in some impossibilities, such as a tracker being set on a target (at time 67.02), before the trace of the target is received (at time 68.201). Second are the exceptions due to the later encoding of goals. A comparison of Table 2 and Table 1 indicates that the AO went to NB-towed with the intent of detecting the hostile submarine (DETECT-SUB). As the trace of the submarine was not present, the AO switched goals and decided to put an additional tracker on the merchant. These encodings altered how the log file was segmented. Third are derive operators added after listening to the videotape. The derive operator represents knowledge that AOs compute or know based on combining prior knowledge with information from displays or based on combining information gained from several displays. Most derive operators are added by the automated encoding program (though none of these are shown in Table 1). Reviewing the videotape results in approximately 2-4 additional derive operators per scenario. (In comparison, there are approximately 350-450 automatically encoded operators per scenario.)

----- INSERT TABLE 2 ABOUT HERE -----

Computational Cognitive Modeling

Models written in ACT-R (Anderson & Lebière, 1998) can use Ned to solve the same scenarios that our AOs solved. Enabling communication between Ned and models was an important consideration in the design of Ned. As with AO performance, all model

actions are timestamped and saved to a log file. (We discuss the current and future status of computational cognitive modeling using Ned under The Future of Ned.)

The Realism of Ned

The realism of concern is the information provided to the AO concerning the targets and ownship. Properties of the direct-path transmission of sound through water are preserved. Other types of sound transmission, most notably bottom-bounce, are not. The information received changes as a function of the distance and bearing of ownship from the targets, as well as ownship speed (e.g., the faster ownship travels the poorer the signal to noise ratio of sound from the target).

Passive sonar information is transformed and presented to the AO in ways that are familiar. The three types of sonar sensors (BB-sphere, BB-towed, and NB-towed) maintain the functionality that AOs expect. In addition to their displays, these three sensors feed their respective information to other simulated instrumentation including the time bearing, time frequency, and geosituational plots, as well as the target motion analysis module.

An important component of the realism is the scenarios. For each target, the scenario specifies the speed, course, depth, and initial distance and bearing from ownship. Targets may change their speed, direction, and depth at pre-specified times. The AO can change ownship speed, course, and depth at any time. As all simulated sonar information is updated dynamically, in theory the scenario duration is indefinite. In practice, our current scenarios are designed to take approximately 15 min to solve. The time taken by AOs has ranged from 6.7 to 30 min per scenario, with a mean of 17.6 min and standard deviation of 5.8 min.

The Engagement of Ned

As far as we can determine, the AOs found Ned engaging. They willingly used Ned and helped to recruit other AOs to do the same. Although we did not ask them whether they enjoyed Ned, we did ask them to rate the effectiveness of each display. Whereas this measure probably confounds realism with engagement, it is one of the few opinion metrics we collected.

On a five point scale (1 = not effective, 5 = very effective) 9 of the 10 displays were rated above average (> 2.5). The three sonar displays received the highest ratings (> 3.8 each). The one display that received the below average rating was LOS (1.54). LOS – line of sight – does not receive any data. It is intended as an electronic notepad on which AOs can work out solutions on the display rather than in their heads. Very few of our AOs used LOS.

We speculate that much of the engagement of Ned is supplied by the AOs' deep knowledge of the situation assessment task. In our demonstrations of Ned, few people who are not submariners have attempted to solve a scenario. These few have always given up after several minutes.

One source of Ned's engagement is that the scenarios are quickly solved. Events in Ned happen much faster than in real submarines. Whereas a change in ownship course

can take several minutes to execute and several more minutes for ownership instrumentation to become stable, in Ned the same sequence is completed in less than a minute. We see this as a necessary tradeoff of realism for engagement.

Tradeoffs Among Tractability, Realism, and Engagement

The research question dictated the minimum level of realism required. This level of realism dictated some compromises with regard to tractability, primarily with regard to the non-uniqueness of information to one display.

The need to make Ned engaging to AOs required the incorporation of additional realism, primarily with respect to the scenarios we developed. However, the needs of engagement conflicted with realism by changing the time scale of situation assessment from hours to minutes. This tradeoff touches on tractability as well; we doubt whether many senior officers could have been persuaded to invest hours to play a low-fidelity simulation. Obtaining an hour of time from this subject population was an important coup. In that hour, we introduced the purpose of the research, an overview of Ned, and collected data from AOs solving two Ned scenarios.

The Future of Ned

Ned may have several futures. Its next future is that of the host of a budding group of computational cognitive models. The initial motivation for developing computational cognitive models is to determine whether the cognitive control structure that emerged from our task analysis is sufficient to perform the AOs' task. There will be two different tests of these models.

The first test is model tracing (Anderson, Boyle, Corbett, & Lewis, 1990; Gray, 1995; Gray, in press; VanLehn, 1988). Model tracing involves concurrently stepping through the action protocol of an AO to compare the decisions made by the model with those that the AO made. Points of similarity and difference will be identified and used to determine the fit of the model to the data.

The second test is a Turing test. For this test, expert raters (senior AOs) will be given transcripts created from the log files of senior AOs, junior AOs, and the ACT-R models. Comparisons of the ratings for AOs versus junior officers will enable us to determine how well Ned succeeded in providing a vehicle in which the AOs' expertise could be displayed. Comparisons of the ratings of the models with senior and junior officers will provide a stringent test of the sufficiency of our models.

As of this writing, a base level model has been built. The base level model allows the researcher to step through an encoding of an AO's scenario. At each step, the experimenter is presented with a list of production rules that are eligible to fire given the current display and state of the simulation. The model will not run by itself, as the control structure for the model has not been implemented. However, the base level model is an important step for two reasons. First, creating the base level model involved solving a large number of programming issues regarding the communication of ACT-R with Ned. Second, at present, each ACT-R production rule corresponds to an operator that was used to encode the protocols. Indeed, the firing of the production rule results in

an entry in the log file similar to those shown in Table 1. Being able to achieve the same solution achieved by the AO using this set of operators provides a sufficiency proof of our operator encodings. In this way, our strategy for building the models mimics the top-down, bottom-up strategy we used for cognitive task analysis.

Ned is a general-purpose tool for studying AO behavior that may have a life beyond the current project. It is quite simple to write more complex Ned scenarios. Such scenarios might be used to tease out different components of the AOs' expertise.

Discussion & Summary – Scaled Worlds as Task Analysis

The purpose of Project Nemo is to study the functional relationships between the query and receipt of information and the various ways used by AOs to transform this information. There are three goals for Project Nemo:

- to describe and simulate the cognitive processes and memory structures used by AOs in locating enemy submarines;
- to be able to generalize our findings to a new, unknown submarine, with reduced manning, and new instrumentation; and
- to contribute to the study of expertise – understanding the AOs' expertise would be a contribution to cognitive theory.

To accomplish these goals, Project Nemo seeks to combine the applied tools of the human factors community with the theoretical ones of cognitive science researchers. We see the development of a scaled world as critical to this approach.

Scaled worlds extract the essence of more complex environments in such a way as to enable key features of this complexity to be studied. Ned represents our attempt to build a tractable, realistic, and engaging scaled world that serves the purposes of Project Nemo.

Ned could not have been built until we had conducted a detailed cognitive task analysis of AOs performing in a more complex simulation. As our knowledge grew our focus became clearer and we were able to narrow our scope to those aspects of the task environment that were most relevant to the AOs' information processing. Our design goals were to maximize the realism of the AOs' information environment while achieving more than a minimum level of tractability and engagement. Designing Ned to track the AO's information processing necessitated the paring away of functional relationships not central to the information environment. The resulting scaled world is a low-fidelity simulation of the AO's real task environment. We believe, however, that as the design of Ned was guided by our previous cognitive task analyses, Ned's levels of tractability, realism, and engagement are sufficient to enable us to draw inferences regarding our research goals.

Author Note

The work on this project at George Mason University was supported by a grant from the Office of Naval Research (#N00014-95-1-0175) to Wayne D. Gray. Susan S.

Kirschenbaum's work has been jointly sponsored by Office of Naval Research (ONR) (Program element 61153N) and by Naval Undersea Warfare Center's Independent Research Program, as Project A10328.

Endnotes

References

Anderson, J. R., Boyle, C. F., Corbett, A. T., & Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. Artificial Intelligence, 42, 7-49.

Anderson, J. R., & Lebière, C. (Eds.). (1998). Atomic components of thought. Hillsdale, NJ: Erlbaum.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. Acta Psychologica, 81(3), 211-241.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. Computers in Human Behavior, 9(2-3), 171-184.

DiFonzo, N., Hantula, D. A., & Bordia, P. (1998). Microworlds for experimental research: Having your (control and collection) cake, and realism too. Behavior Research Methods, Instruments, & Computers, 30(2), 278-286.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data. (Revised ed.). Cambridge, MA: The MIT Press.

Gray, W. D. (1995). VCR-as-paradigm: A study and taxonomy of errors in an interactive task. In K. Nordby, P. Helmersen, D. J. Gilmore, & S. A. Arnesen (Eds.), Human-Computer Interaction--Interact'95 (pp. 265-270). New York: Chapman & Hall.

Gray, W. D. (in press). The nature and processing of errors in interactive behavior. Cognitive Science.

Gray, W. D., & Kirschenbaum, S. S. (in press). Analyzing a novel expertise: An unmarked road. In J. M. C. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.), Cognitive task analysis. Mahwah, NJ: Erlbaum.

Gray, W. D., Kirschenbaum, S. S., & Ehret, B. D. (1997). The précis of Project Nemo, phase 1: Subgoalting and subschemas for submariners. *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 283-288). Hillsdale, NJ: Erlbaum.

Kirschenbaum, S. S., Gray, W. D., & Ehret, B. D. (1997). Subgoalting and subschemas for submariners: Cognitive models of situation assessment (Technical Report 10,764-1). Newport, RI: NUWC-NPT.

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. Organizational Behavior and Human Decision Processes, 68(1), 28-43.

Merrill, D. C., Reiser, B. J., Beekelaar, R., & Hamid, A. (1992). Making processes visible: Scaffolding learning with reasoning-congruent representations. In C. Frasson, G. Gauthier, & G. I. McCalla (Eds.), Intelligent Tutoring Systems. New York: Springer-Verlag.

Norman, D. A. (1989). The design of everyday things. New York: DoubleDay.

Soldow, D. S. (1998). An assessment of submarine approach officer decision-making and its implications for command workstation design. Unpublished Masters Thesis, Naval Postgraduate School, Monterey, CA.

VanLehn, K. (1988). Student modeling. In M. C. Polson & J. J. Richardson (Eds.), Foundations of intelligent tutoring systems. Hillsdale, NJ: Erlbaum.

vanSomerén, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). The think aloud method: A practical guide to modelling cognitive processes. New York: Academic Press.

Tables

Table 1: Sample automated encoding and segmenting of the log file from near the beginning of one AO's scenario. Time is the time in seconds from the beginning of the scenario. Operators are the finest grained encoding categories defined by the experimenters. Info-source refers to the display or field within a display. Ship identifies whether the information accessed concerns the merchant (MERC), hostile submarine (SUB), or ownship (OS). Attribute and value refer to the type of information queried and the information that the AO extracted. Duration is the length of time (in sec) that the gray box was opened. The information within a segment is information that is revealed when the gray box covering that field is opened.

TIME	OPERATOR	INFO-SOURCE	SHIP	ATTRIBUTE	VALUE	DURATION
62.428	DISPLAY-NAV	SONAR-NB-TOWED				
63.98	QUERY	NBT-WATERFALL				
67.02	SET-TRACKER	SONAR-NB-TOWED	MERC			
68.201	RECEIVE	NBT-WATERFALL	SUB	ON-SONAR	NO	4.221
68.201	RECEIVE	NBT-WATERFALL	MERC	ON-SONAR	YES	4.221
68.201	RECEIVE	NBT-WATERFALL	MERC	BEARING	BEAM	4.221
68.201	RECEIVE	NBT-WATERFALL	MERC	TRACKING	YES	4.221
70.063	QUERY	NBT-CONICAL-ANGLE-FIELD	MERC	CONICAL-ANGLE		
70.724	RECEIVE	NBT-CONICAL-ANGLE-FIELD	MERC	CONICAL-ANGLE	82.15	0.661
71.111	QUERY	NBT-BEARING-FIELD	MERC	BEARING		
72.088	RECEIVE	NBT-BEARING-FIELD	MERC	BEARING	152_OR_314	0.977
72.393	QUERY	NBT-SNR-FIELD	MERC	SNR		
72.987	RECEIVE	NBT-SNR-FIELD	MERC	SNR	6.63	0.594
74.635	QUERY	NBT-ARRAY-STATUS-FIELD	OS	ARRAY		
75.601	RECEIVE	NBT-ARRAY-STATUS-FIELD	OS	ARRAY	STABLE	0.966

Table 2: Segment shown in Table 1 following resegmentation and encoding of goals by the experimenters. The headings are the same as in Table 1 with the addition of three fields for goals and subgoals – levels 1 (L1), 2 (L2), and 3 (L3). No L3 goals are encoded in this segment.

TIME	L1	L2	L3	OPERATOR	INFO-SOURCE	SHIP	ATTRIBUTE	VALUE	DURATION
	DETECT-SUB								
62.428				DISPLAY-NAV	SONAR-NB-TOWED				
63.98				QUERY	NBT-WATERFALL				
66.82				RECEIVE	NBT-WATERFALL	SUB	ON-SONAR	NO	4.221
	POP								
	LOCALIZE-MERC								
		SET-TRACKER							
67.02				SET-TRACKER	SONAR-NB-TOWED	MERC			
68.201				RECEIVE	NBT-WATERFALL	MERC	ON-SONAR	YES	4.221
68.201				RECEIVE	NBT-WATERFALL	MERC	BEARING	BEAM	4.221
68.201				RECEIVE	NBT-WATERFALL	MERC	TRACKING	YES	4.221
		POP							
		DETERMINE-CONICAL-ANGLE							
70.063				QUERY	NBT-CONICAL-ANGLE-FIELD	MERC	CONICAL-ANGLE		
70.724				RECEIVE	NBT-CONICAL-ANGLE-FIELD	MERC	CONICAL-ANGLE	82.15	0.661
		POP							
		DETERMINE-BY							
71.111				QUERY	NBT-BEARING-FIELD	MERC	BEARING		
72.088				RECEIVE	NBT-BEARING-FIELD	MERC	BEARING	152_OR_314	0.977
		POP							
		DETERMINE-SNR							
72.393				QUERY	NBT-SNR-FIELD	MERC	SNR		
72.987				RECEIVE	NBT-SNR-FIELD	MERC	SNR	6.63	0.594
		POP							
	POP								
	EVALUATE-ARRAY-STATUS								
74.635				QUERY	NBT-ARRAY-STATUS-FIELD	OS	ARRAY		
75.601				RECEIVE	NBT-ARRAY-STATUS-FIELD	OS	ARRAY	STABLE	0.966
	POP								

Figures

Figure 1. Relationship between three dimensions of simulated task environments.

Figure 2: The sonar, narrow-banded towed display used in Ned. NB-towed is one of Ned's 10 displays and one of three that include sonar sensors. The display shows two targets (indicated by the white streaks, or waterfall, in the left) display being tracked (indicated by the circles attached to each streak). The NB-towed waterfall display indicates targets by the frequency of the sound they emit (from 0hz to 400hz at the top of the display). Target bearing from ownship is indicated by the band in which the target occurs beginning with 0-22.5° and ending with 157.5-180°. Thus, NB-towed yields an ambiguous bearing as it does not indicate which side of the ship the target is on (i.e., 0-180° or 180-360°). The table on the right indicates, for each target, its angle from ownship, its ambiguous bearing (for target S001 the bearing is either 296° or 344°), and the signal-to-noise ratio. On the bottom-right, the display indicates that the array of sonar sensors used for NB-towed is stable. To the right of NB-towed is the menu of the other displays available to the AO. Ned is generated dynamically and all information is updated every second.

Figure 3: Schema-directed problem solving with shallow and adaptive subgoalting (SDPSSAS)

Figure 4. Ratings of scenario realism (1 - Very Unrealistic, 5-Very Realistic).



