Modeling Vigilance Performance as a Complex Adaptive System

Joerg Wellbrink, PhD.
The MOVES Institute
Naval Postgraduate School

Mike Zyda. PhD.
The MOVES Institute
Naval Postgraduate School
Director MOVES Institute

John Hiles
The MOVES Institute
Naval Postgraduate School

Keywords:
Cognitive Architectures, Human Performance Models, Vigilance, Complex Adaptive System, Attention

Abstract: Current cognitive models not only lack flexibility and realism, they struggle to model individual behavior and reduced performance. We propose that reduced human performance can be best modeled as a complex adaptive system. We built a multi-agent model "Reduced Human Performance Model (RHPM)" as a proof of principle. The simulation system realistically simulates the reduction of vigilance that individuals experience during such operations as airport screening, radar-screen operation, and other vital tasks in which attention easily flags. The developed multi-agent system generates individual behavior within a reasonable range. Its use for computer generated forces (i.e. radar screen operator) would improve the realism of simulation systems by adding human like reduced vigilance performance. The model represents a well suited tool to mediate between vigilance theories such as signal detection theory and experimental data. Using the model as a surrogate generates insights that potentially create likely hypotheses to improve the theories.

I. Introduction

The attacks of September 11, 2001, showed not for the first time in Western history (Pearl Harbor, Yum Kippur), a need for more realistic simulation models. Col. Chorev [1] recommended improving the formation of hypotheses – in order to increase the perceived likelihood of alternative interpretations and scenarios that may sensitize analysts and decision makers to discrepant information. Simulation models should be capable of generating or revealing surprises, unintended consequences, and blind spots, thereby forming new hypotheses. One core assumption of this research is that modeling surprises requires simulating realistic reduced human performance allowing non ideal behavior. In a simulation system human-like errors potentially lead to surprising or unexpected outcomes for example by a cascade of errors that analysts did not perceive.
The National Research Council's report in 1998 [2] and follow on research [3,4] on Modeling Human and Organizational Behavior described the State-of–the art in cognitive modeling and the results showed the need for a different approach.

> Even the best of them [cognitive architectures] assume ideal human behavior according to doctrine that will be carried out literally, and rarely take account of the vagaries of human performance capacities [2, p.4].

In order to improve human performance models we must have:

- Detailed data about the behaviors being modeled.
- Greater understanding of fundamentals of human performance.
- Improved architectures for building models.
- Better methods for verification and validation.
- Improved communication about model characteristics.
- More work on individual differences will ensure that models exhibit a reasonable range of responses [4, p.5]

The NRC also encouraged to research alternative paths to modeling human behavior. The predominant approaches in cognitive modeling are symbolism, connectionism or hybrid approaches. One possible alternative solution path was the idea of utilizing complexity theory to build a human performance model. John Holland, a major researcher in complexity theory, commented on effects of complex adaptive system models:

> I just love these things where the situation unfolds and I say,' Gee whiz! Did that really come from these assumptions!?' Because if I do it right, if the underlying rules of evolution of the themes are in control and not me, then I'll be surprised. And if I'm not surprised, then I am not very happy, because I know I've built everything in from the start ([5, p.11]).

Hence we hypothesized that reduced human performance resulting from a vigilance task can be modeled as a complex adaptive system (CAS) and that the resulting computational model can be shown to approximate empirical human performance data under similar conditions.

This multi-disciplinary research covers many different areas like agent based modeling, complex adaptive systems theory, performance psychology (specifically vigilance psychology), cognitive modeling, discrete event simulation, and software engineering. The interested reader can find more details on the areas and how they apply to this research in [6]. It is clearly beyond the scope of the paper to explain the background of all mentioned disciplines. This paper will focus first on defining complex adaptive systems and showing a successful example of using the theory to build a computational model of land combat. Chapter III then briefly describes the vigilance psychology background and how it relates to this research. Chapter IV gives a broad overview on the design of the computational model. Chapter V shows experimental results (human experiments and computational experiments). Chapter VI summarizes the findings and shows the potential benefits for military modeling and simulation.

II. Complex Adaptive Systems (CAS)

There is no standardized definition for a complex adaptive system (CAS). Some researchers say "I know it when I see it". This research uses the following provisional

working definition derived after two prominent researchers from the Santa Fe Institute (John Holland, Murray Gell-Mann):

> A complex adaptive system consists of many agents, acting in parallel without centralized control. The non-linear interactions between these agents lead to adaptive and emergent behavior. The agents organize in dynamically re-arranging non homeostatic structures. The system builds an internal (implicit or explicit) model of the future and acts according to its internal models.

CAS theory has been successfully applied to various sciences like sociology and medicine. Historically, many sciences were founded based on Newton's mechanistic explanation of physics. Newton hypothesized that the universe is mechanistic. He envisioned the universe as a gigantic mechanical clock, where simple rules govern the relationship of the single parts of this clock [7]. Since his rules were very well suited to explain many phenomena (e.g. movement of stars in relation to each other), his approach became the overwhelming approach for almost 250 years. Einstein's relativity theory showed where Newtonian physics fell short. Thus, physics was probably the first science that found complementary theories expanding the mechanistic world view incorporating dynamics of space and time relationships. Dynamic systems constantly change into different equilibria and never maintain a particular equilibrium [8]. Meanwhile many other sciences are beginning to use CAS theory looking at their domain from a different perspective. Economy is a prime example on how CAS theory has changed the perception of a former static theory, called the neoclassical approach. The initial research at the Santa Fe Institute [9, 10, 11] specifically used economics as one application area. Complexity theory has in fact improved the realism of simulation systems, like the artificial immune system (AIS) [12] or Ilachinski's Irreducible Semi-Autonomous Adaptive Combat '(ISAAC) combat model which we will describe next in more detail.

ISAAC Combat Model

For the last century, conventional wisdom regarding the basic processes of war and most current models of land combat has been rooted in the idea of Lanchester Equations (LE). In 1914, F.W. Lanchester used differential equations to express attrition rates on the battlefield. These equations have been modified over the years, but the main assumption is that combat is always driven by a force-on-force attrition rate. This theory ignores spatial relationships and the human factor in combat. It certainly was not adequate to support analysis of the United States Marine Corps' vision of small, highly trained, well-armed autonomous teams working in concert, continually adapting to changing conditions and environments. Thus, Prof. Ilachinski challenged the almost century-old theory by arguing that land combat can (and should) be modeled as a complex adaptive system. He transferred complexity theory into the military domain and showed that land combat properties resemble the properties of CAS [13]. His work has generated a lot of interest in combat modeling especially because tactical behaviors such as flank maneuvers, containment, encirclement and "Guerilla-like" assaults emerged out of his implementation.

> In ISAAC, the "final outcome" of a battle -- as defined, say, by measuring the surviving force strengths -- takes second stage to exploring how two forces might "co-evolve" during combat. A few examples of the profoundly non-equilibrium dynamics that characterizes much of real combat include: the sudden "flash of insight" of a clever commander that changes the course of a battle; the swift flanking maneuver that surprises the enemy; and the serendipitous confluence of several far-separated (and unorchestrated) events that lead to victory. These are the kinds of behavior that Lanchesterian-based models are in principle incapable of even addressing. ISAAC represents a first step toward being able to explore such questions [13, p.226].

Ilachinski's work has not died out. Many research projects continue to explore his ideas. Project Albert is an international military research effort with many participating countries (i.e. United States, Australia, New Zealand, and Germany) [14]. The MOVES Institute especially has produced many follow-on projects. ([15] provides a good summary of this work). MOVES also adapted Ilachinski's definition of autonomous software agents that is being utilized in this research:

> The fundamental building block of most models of complex adaptive systems is the so-called *adaptive autonomous* agent. Adaptive autonomous agents try to satisfy a set of goals (which may be either fixed or time-dependent) in an unpredictable and changing environment. These agents are "adaptive" in the sense that they can use their experience to continually improve their ability to deal with shifting goals and motivations. They are "autonomous" in that they operate completely autonomously, and do not need to obey instructions issued by a God-like oracle. [13, p.13]

The paradigm for combat modeling has fundamentally changed and improved insights into the processes. These types of simulation systems will enhance the capabilities exploring policy and concept development as well as force structure development. ISAAC agents simulate human-like behavior without going into the details of cognitive modeling. Our approach goes one step further by modeling individual behavior that emerges from the interaction of multiple autonomous software agents.

Complex Adaptive Systems theory has furthered the understanding of previously ignored (or taken for granted) phenomena. The average behavior assumption (humans show the same reaction to the same stimulus on average) in the social sciences combined with the rationality assumption of human behavior (humans always choose the rational decision) has led to a linear mechanistic world view [16]. By refusing these assumptions, economics, organizational sciences and combat modeling theory have advanced to a better understanding of the relationships between individual elements (firms, groups, and individuals). A natural extension to this view motivated by the success stories of others is to research how individuals are modeled and whether complexity theory could improve the understanding of individual based behavior while overcoming the mechanistic

brittleness of current cognitive models. The next chapter explains the background in human vigilance performance.

III. Human Vigilance Performance

Human beings are born to perform. In a broader sense, we perform every time we engage in a goal-directed activity. Real life performance depends not just on the task, but also on the influence of stress factors such as noise, heat and fatigue. Furthermore, individuals differ in their abilities and motivations when called upon to perform [17].

Vigilance research started in the early 1930s and was established by Mackworth's work on naval recruits. Mackworth was tasked to research the question why so many enemy submarines that were on the radar screen of radar operators still remained undetected. He studied the phenomenon of the vigilance decrement in laboratory settings.

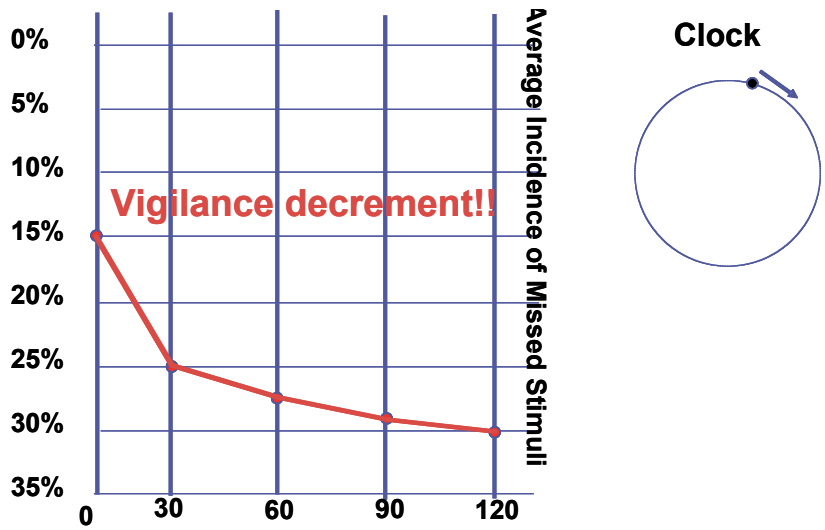

Figure 1: Mackworth's Clock Experiment and Results

Figure 1 shows the results of the Mackworth clock test. It was used to establish the increase in misses and the increase in reaction time. Subjects watched a clock's watch hand for two hours. Whenever the watch hand jumped two instead of one second the subjects had to report it. Within the first 30 minutes the decrement in hit rate was most pronounced. After that the decrement leveled off and stayed at an almost constant level [18].

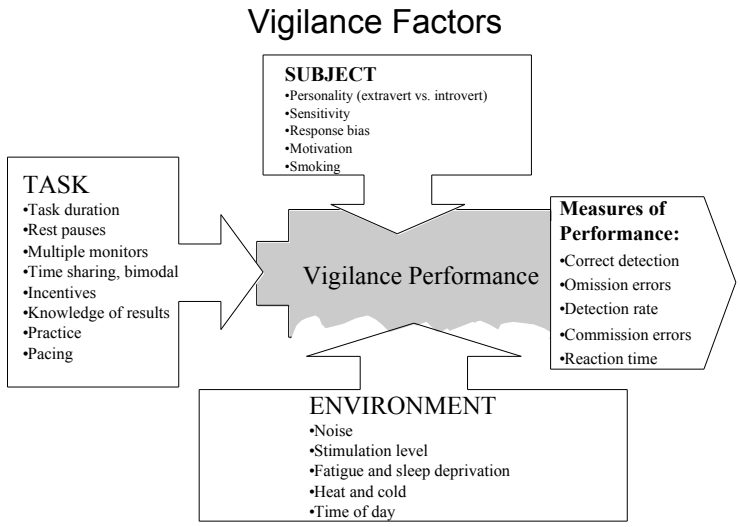

Figure 2 Vigilance Factors

Figure 2 summarizes the findings of several researchers [17, 19, 20, 21]. It shows most of the main factors that influence vigilance performance. It also shows a sample of the different measures of performance (MOP).

There are three main factors that impact vigilance performance: Task factor, environmental factor and subjective factor. These factors are determined by their identified variables (i.e. the environmental factor is determined by the stress level). One of the research questions focused on how much personality influence vigilance performance. Some factors have a stronger impact like feedback of results almost averts the decrement, sometimes factors cancel each other out. For example there is evidence that degraded performance due to sleep deprivation can be counteracted by noise. The measures of performance are typically expressed in correct detections or hits, omission errors or misses, commission errors or false alarms and reaction time.

No vigilance theory sufficiently explains the entire phenomenon consistently. This poses a challenge and creates an opportunity for a computational model of vigilance. The computational model should allow the representation of different theories and hybrids of theories. The model could then be used to explore strengths and weaknesses of current theories.

IV. Reduced Human Performance Model (RHPM)

This chapter examines the basic idea of using agent based models as an ideal software implementation for complex adaptive systems , the design of RHPM and how these relate to complex adaptive systems theory and vigilance research.

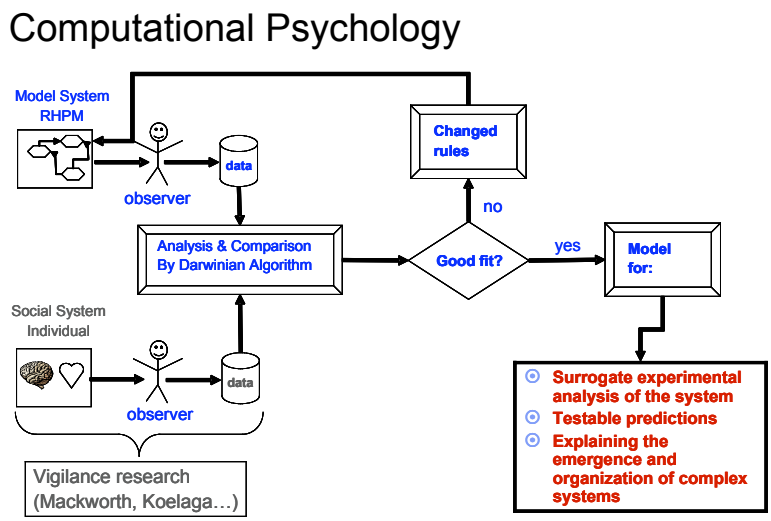

Figure 3 Fitting a Model to a Real System (adapted from [22])

Fig 3 shows how to fit an agent based model to a social system.

This research uses the term Computational Psychology (derived from computational physics) to describe this field as an addition to theoretical and experimental psychology. It fits the reduced-human-performance model (RHPM) to theories and experimental results of human vigilance performance. The analysis and comparison of the model's and the system's output can lead to either a good fit (unlikely in the early stages) or to a change in structures, rules and parameters. This research implements existing theories in vigilance psychology for the computational model and then harnesses complexity by fitting a complex adaptive system to a range of experiments. Once model and system output is sufficiently similar, the model can potentially be used as a surrogate of the system, generate predictions, or explain previously unexplained phenomena [22].
Next we will describe the main psychological models and the main components of RHPM.

IV.1 Design of the Model

The Reduced Human Performance Model (RHPM) uses two psychological models as the blueprint for design and implementation:

- Human Information Stage Processing Model, and
- Multiple Resource Model.

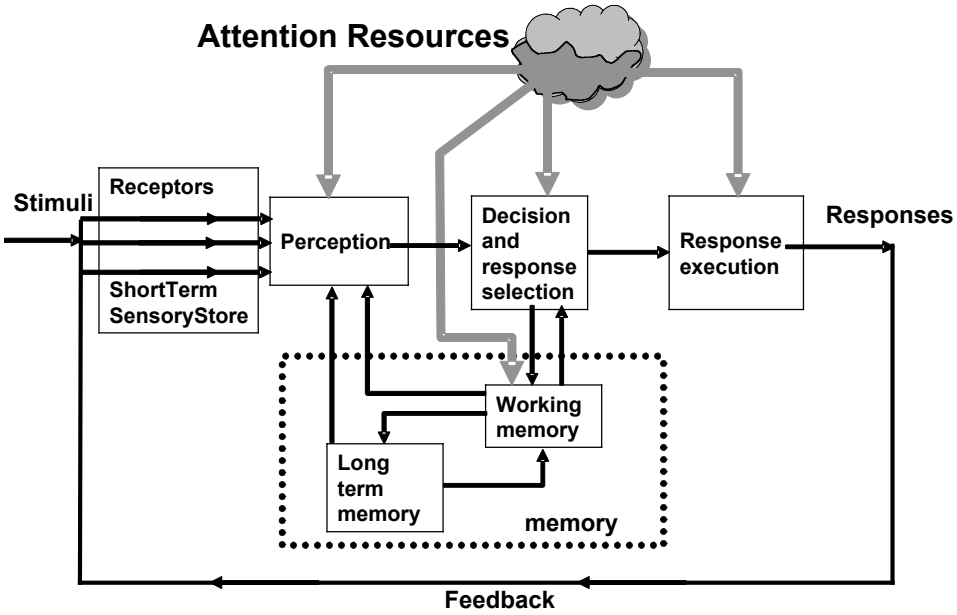

Figure 4: Stage Model of Human Information Processing (after [23])

Figure 4 shows the different stages for the human information processing. A stimulus is stored in the short-term sensory store (STSS) for a few seconds (visual stimulus about 1 second, auditory stimulus about 5 sec; echoic memory). If it is not perceived within this timeframe, it is not a perception. Perceptions are sometimes matched with patterns, likely stored in long-term memory. This is the encoding stage. Next during the central processing stage, the perception is forwarded to the decision- and response-selection system, which uses the working memory to determine whether an action should be initiated. The last stage is the response-execution stage, which leads either to a vocal or manual response to the perceived stimuli [23]

Pew et al. [2] modified this model slightly to show the elements that should be included in an integrative architecture. They left out the STSS and connected the perception to long-term memory via working memory. However, a major alteration to the original stage model is leaving out the attentive resources which seem to be central to modeling reduced performance.
These resources seem to be a key in modeling reduced human performance caused by a lack of attentional resources. There are several theories and models on how humans use there cognitive resources. Wickens' multiple resource model suggests that cognitive resources can be divided into modalities and codes in different stages of the information process.

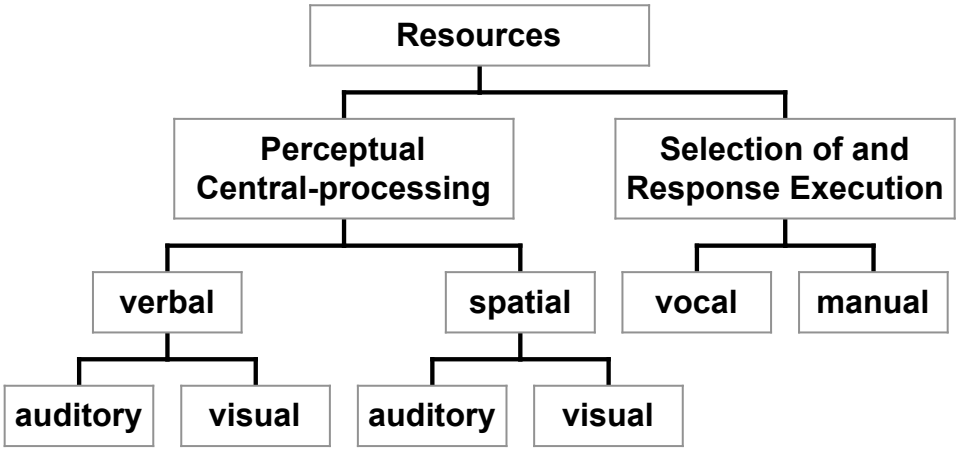

Figure 5 Multiple Resource Model (from [24])

Figure 5 is an adaptation of the better known cube that can be seen in many textbooks [17, 23, 24].
The Multiple Resource Model assumes that humans have two main attentional resource pools: one for the perceptual and central-processing phase, and one for the response-selection and execution phase. These resources can be divided into verbal and spatial, or, respectively, vocal and manual. The structure indicates a hierarchical system. The system is adaptive since humans can focus their attention (selective attention) filtering information to a certain extent in context. Thus, humans adapt cognitive resource consciously or subconsciously (or both) to a changing environment. We envisioned implementing this model by using reactive agents that compete for resources and also supply energy to others. We also expected that the nonlinear interactions between the attentive resources would have different effects on the information processing stages, which eventually result in interesting human-like emergent behavior.

RHPM also implemented parts of several vigilance theories (signal detection theory, expectancy, arousal, resource). [6] describes the prevailing main theories in detail.
Figure 6 shows the main modules of RHPM. The major components, which also show this research's main contributions in terms of modeling, are blue.

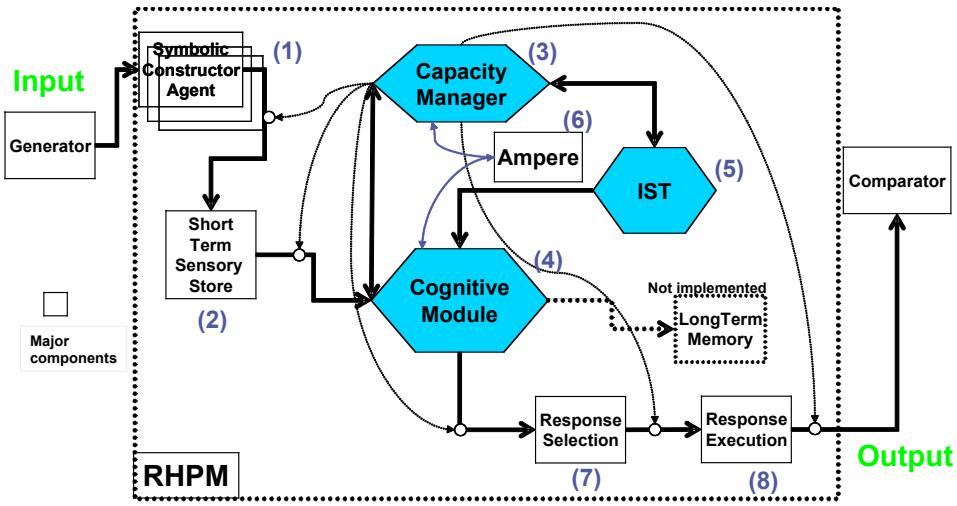

Figure 6 Numbered RHPM Modules

(1) Symbolic Constructor Agents: Symbolic constructor agents (SCAs) encode impressions (input into the system). SCAs represent the perception aspect of this framework. They have been used in a number of projects at the MOVES Institute: see [15] for more details. This model uses two different input modalities, auditory and visual. For every modality, there exists a specialized agent whose performance decreases with time on task to mimic the loss of sensitivity often seen in vigilance tasks. The agent relays the observation to the short-term sensory store.

(2) Short-Term Sensory Store: Chris Wickens describes the functionality of the ShortTermSensoryStore (STSS) in context with the information-stage model:

> Each sensory system, or modality, appears to be equipped with a central mechanism that prolongs a representation of the physical stimulus for a short period of time after the stimulus has physically terminated. When attention is diverted elsewhere the STSS permits environmental information to be preserved temporarily and dealt with later. Three general properties are characteristic of STSS: (1) It is preattentive; that is no conscious attention is required to prolong the image during the natural "time constant" of the store. (2) It is relatively veridical, preserving most of the physical details of the stimulus. (3) It is rapidly decaying [23, p.18].

(3) Capacity Manager: The Capacity Manager is an agent implementation of Wicken's multiple resource model. Agents try to keep their energy flow consistent by means of changing their states. Since there are multiple agents trying to change their flows and interacting with each other the emergent behavior produces a varying degree of performance.

(4) Cognitive Module: The Cognitive Module captures the functionality of the perception and memory parts of the information-stage processing model. This module is a multi-agent system consisting of several heterogeneous, composite reactive agents (see [15] for

detailed explanations and definitions). The Capacity Manager actually provides resources that the agents use to work on a task. The more resource is available the faster they can fulfill their tasks (i.e. Search Agent tries to identify a percept by searching the Working Memory). The energy flow can change during the task which impacts the time to finish the task. However every agent tries to maintain a consistent resource flow changing its states frequently. The interaction generates the desired emergent performance, a decrease in vigilance. The energy flow computation is implemented in the next module—the Ampere Module.

(5) Ampere Module: The agents used within the Capacity Manager and the Cognitive Module only have local knowledge. Ampere is providing this knowledge by computing the different flows for the agents utilizing a mathematical model based on electrical circuit theory. It would be beyond the scope of this paper to explain in detail the background of this part. [6] explains the idea and reasoning behind using electrical circuits in detail.

(6) Individual States and Traits (IST) Module: This module is a pre-planned multi-agent component where emotions and external and internal stressors can effect RHPM's performance. One example is the Distraction Agent trying to increase its energy flow over time to model the effect of decreasing attention during time on task. By increasing its demand it draws energy away from the processes taking place inside the Capacity Manager and Cognitive Module.
Another example is the Expectancy Agent. It sets up expectations by computing statistics such as perceived signal probability and rate influencing and the decision criterion of Response Selection.

(7) Response Selection: This object uses a simple mechanism to determine whether it detected a signal, comparing the nominal value of the percept to a criterion. If the value is below threshold, the percept is classified as noise; if above threshold, as a signal. This is a straightforward implementation of a mechanism know from signal detection theory.

(8) Response Execution: The Response Execution produces RHPM's output. One example is the identification whether or not a perceived stimulus is a signal or noise.
It has uses a stochastic mechanism for producing slips. A slip is an omission error—in our case knowing the right thing to say (yes or no), but saying either the opposite or nothing.

Generator is an object that recreates the experimental scenarios by generating stimuli (simulated visual and auditory stimuli) for the system. The Comparator then compares RHPM's answer to the true signature of the stimulus (noise or signal) and computes the measures of performance such as hit rate, false alarm rate and reaction time.

V. Experiments

This research utilized three different experiments:

1. A five factor personality test.
2. Vigilance experiment with 4 different conditions

3. Computational experiments to calibrate and validate the model.

These experiments were intended to research questions such as:

- Is the military population' sample a normal population?
- Can performance differences be explained by personality traits?

The vigilance experiments were needed to have controlled scenarios for expected experimental data. Only if all conditions of the experiments are known it would make sense to recreate in-silico experiments with a computational model.
The next sections describe these experiments and our findings in more detail

V.1 Personality Test

We used an electronic version of the NEO FFI short version. The five factors are openness, consciousness, extraversion, agreeableness, and neuroticism (OCEAN). The test consists of 60 questions that are answered on a rating scale (from strongly agree to strongly disagree). The test computes the raw scores and standardizes them under the assumption that scores are normally distributed with a mean t-score of 50 [25]. The NEO FFI not only provides the t-scores of individuals in the five dimensions it also correlates their traits describing certain styles of behavior based on the trait assessment.

Participants: Fifty Naval Postgraduate students (mostly military officers) participated in the study (38 US students (5 female) and 12 foreign students from four different countries (Germany, Greece, Singapore, Turkey). Mean age was 34. Participants volunteered and received a personality report printout after conducting all experiments.

Subjective Measures: Participants completed the NEO FFI electronic version before they started the vigilance experiments.

Results:

The results indicate that the tested population is in fact not a normal population. There are biases that might be typical for the military community:

One Sample T-test	P-O	P-C	P-E	P-A	P-N
Mean	53.36	52.92	53.82	46.86	45.60
stddev	9.81	10.68	8.53	11.24	9.51
t	2.42	1.93	3.17	-1.98	-3.27
df	49.00	49.00	49.00	49.00	49.00
alpha 0.025	2.01	2.01	2.01	2.01	2.01
Ho mean=50	Reject	Fail to reject	Reject	Fail to reject	Reject
CI Lower	50.57	49.88	51.40	43.67	42.90
CI Upper	56.15	55.96	56.24	50.05	48.30

Table 1 One Sample T-test for Personality Scores

Table 1 summarizes the result of conducted two-tailed t-tests. Every dimension was tested against the following hypothesis at the alpha level of 0.05:

$$H_0 : \mu_{\text{trait}} = 50;$$
$$H_1 : \mu_{\text{trait}} \neq 50;$$

There are three traits (openness (O), extroversion (E), and neuroticism (N)) where the null hypothesis was rejected, indicating that the means of these traits differ from a normal population. Thus, the sample is more prone to score high in O, high in E, and low in N. The latter score is certainly a desired trait in the military community since a low score in negative emotionality indicates a more relaxed reaction to negative experiences. There was not enough evidence to reject the null hypothesis for conscientiousness (C) and agreeableness (A). The 95% confidence intervals consequently (albeit barely) cover a mean of 50. The result indicates that simulation systems have to take the shown bias instead of an average assumption into account. Statistical analysis that related personality traits and vigilance performance showed that the trait extraversion (E) explained up to 30 % of the variability between subjects.

V.2 Vigilance Experiments

.

The vigilance experiments were conducted on PCs with the SynWinGenerator (TM Activity Research Inc). We produced two data sets; one data set was needed to calibrate the model, the other one was needed to validate the model by exposing it to the previously unseen experimental scenario without changing any model parameters and structure during the validation runs.

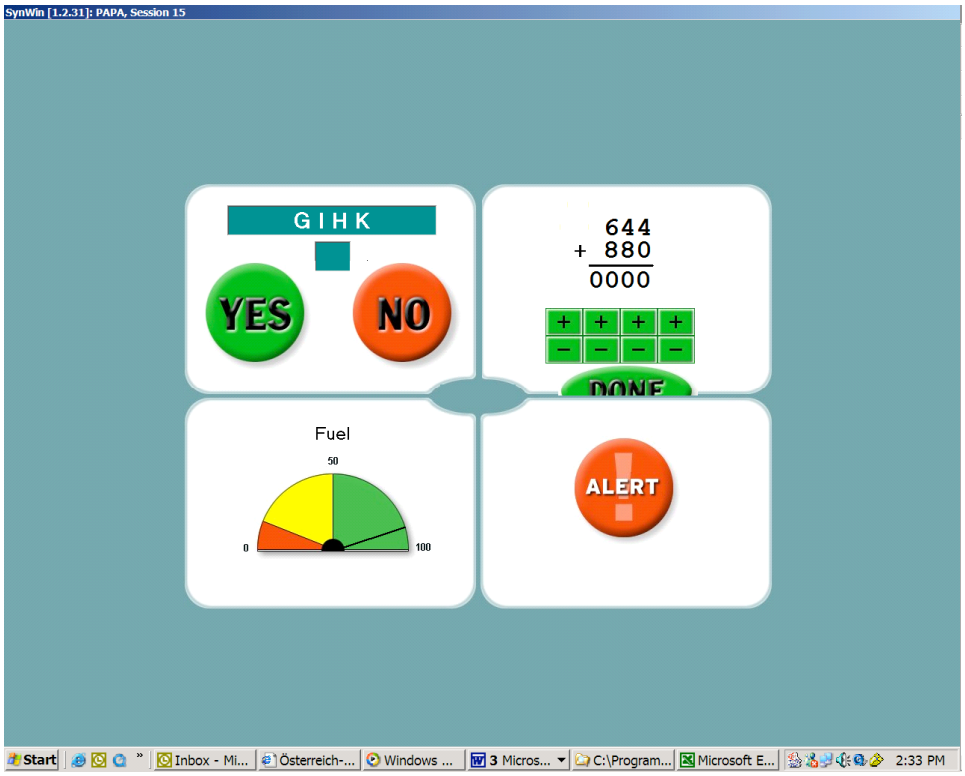

Figure 7 SynWin Generator (TM Activity Research Inc.)

Figure 7 shows the SynWinGenerator displays. The upper left window is a Sternberg Memory task. At the beginning of the experiment four letters were displayed. During the experiment probe letters were randomly displayed (trial duration 8 seconds) and participants had to decide whether or not the letter was in the test sample. Feedback for correct, false or missing answers was given with the help of a point display in the middle of the screen and an auditory signal for mistakes. The upper right corner shows a simple cognitive task, computing digits. Participants could use the + or – buttons to display the sum of the math task. Feedback for correct and mistaken answers was given via the point display and an auditory signal.

The visual monitoring task is on the lower left side. Participants watched the fuel gauge and mouse-clicked on it when the needle went into the red zone. Lapses were defined as either clicking too early or letting the needle touch the bottom.

The alert button belonged to the auditory vigilance task. A sound was played periodically every 3 seconds. The noise sound was 1000 hz and 0.15 sec in duration. The signal sound was 1025 hz and 0.15 sec in duration. The participant's task was to click the ALERT button following the signal sound, before the next sound occurred. The probability of the signal sound was 0.1. Measured results (a snapshot was taken every 10 minutes) contain number of hits and misses, number of false alarms and correct rejections, reaction times for hits and false alarms.

There were four different treatments.

- Low workload treatment (visual and auditory monitoring tasks)
- High workload treatment (all tasks)
- Going from high to low back to high workload
- Going from low to high back to low workload.

The first two treatments produced the data set for calibrating the model. The latter two treatments produced the validation data set.
Table 2 shows the summary of results for the first two workload conditions

	Reaction Time		Misses		False Alarms	
	Low	High	Low	High	Low	High
10 min	0.81	1.05	1.69	5.77	3.81	3.63
20 min	0.78	0.96	1.45	1.53	3.21	3.67
30 min	0.83	0.98	3.62	3.05	2.43	3.40

Table 2 Measure of Performance for Low and High Workload

The results clearly show the impact of a higher workload on the overall performance. The higher workload especially impacted the reaction time and initially the miss rate. The mixed treatments' results showed the impact of differing working conditions on vigilance performance. For example there are also distinct differences in reaction times. The error rates were generally lower which indicates that subject learned to better distinguish signal from noises. Even subjects that could not clearly hear differences in the first two experiments improved their performance over time. Since they were not given any feedback on their performance there could be a perceptual learning effect that deserves more research.

One surprising result was the outcome of the false alarm rate in the high low high condition. Despite an increase in workload subjects further decreased their false alarm rate. This result was counterintuitive and further research is needed to explore reasons. The final step was calibrating the model with the first two treatments and validating its output versus the mixed treatments' results.

V.3 Computational Experiments

RHPM underwent extensive computational experiments for calibrating and validating the model. The validation process did not only comprise a data validation but also a stretching of the model to see the bounds of its performance. The initial validation strategy was derived after M. Carley [26]. She suggested a strategy on how to harmonize and validate computational models.

Calibrating RHPM

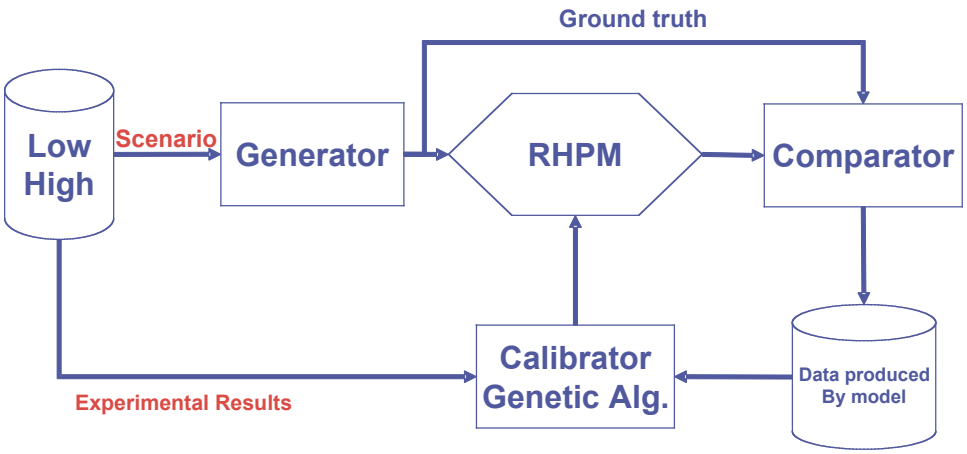

Figure 8 Calibrating RHPM

The scenarios for the low and high workload treatment were artificially duplicated and used to generate input to the model RHPM. The Calibrator uses genetic algorithms to produce better fitting parameter set ups. This clearly was a risky procedure because there was no guarantee that the model would converge to the desired outcome. In that case the model's structure and rules had to be changed and a new calibration process would be started.

A population of 200 different model parameter set ups was generated. Each set up was tested 30 times with the high and low workload condition to compute the statistics. The Calibrator computed the resulting score with the squared differences between model and human output (better known as sum of squared error) as the fitness function. After identifying the best score, this score was used as a benchmark for the creation of a new population. While the population size did not equal 200, set up's were drawn from the original population.

The drawing followed Goldberg's idea of a wheel [27]. We used the minimum score as nominator and the current score as denominator. If a uniform random number was below that ratio the parameter set up progressed into the mating pool. Thus the higher its score the less likely it was going to be in the mating pool. Then we applied the crossover and mutation to the new population and started the process all over. The goal was to calibrate the model such that it would have a resulting score lower than a defined threshold. This threshold is critical to generate a robust parameter set-up.
We hypothesized that a good enough fit would probably be more adaptive to unknown scenarios. The computed fitness value is the sum of squares error computed from 18 different measure points (every MOE was collected at 10, 20 and 30 minutes in two varied conditions.) Thus the goal was not to have a perfect fit of individual performance curves but to have a sufficiently close result for all 18 measurement points. The calibration process worked surprisingly well. After 78 generations the fitness value was below the selected threshold. The next critical step was to see whether or not the computational model could match human results with the computed optimal parameter set up.

Validating RHPM

The model was tested with 24 repetitions (comparable to 24 subjects during the experiment). Every single test run result was treated like a subject's result. The MOEs were computed and statistical analysis was conducted. The individual results were used for paired T-tests to see whether there are significant differences in means and variances.

The next figure shows that RHPM very closely resembles human vigilance performance in terms of reaction time.

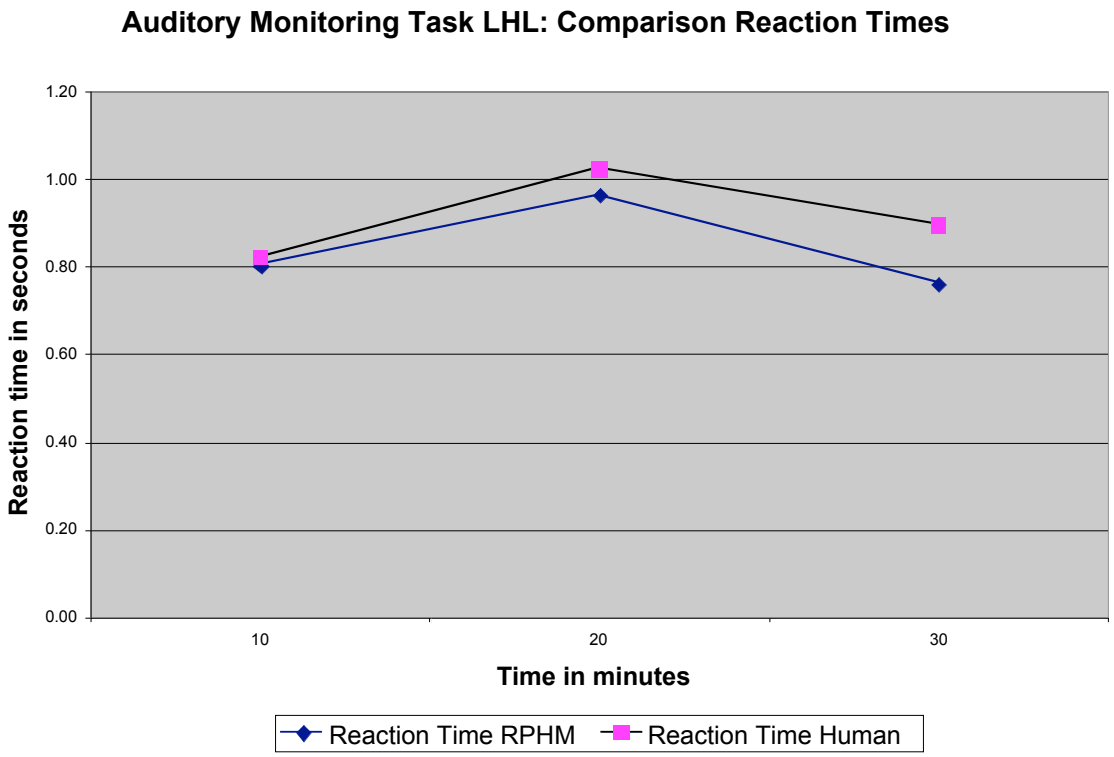

Figure 9 Comparison Reaction Times during the LHL treatment

RHPM showed its contribution by closely matching human vigilance performance degradation within four different experiments.

MOE vs. treatment	Reaction Time			Misses			False Alarms		
	10	20	30	10	20	30	10	20	30
Low	X	X	X	X	X	X	X	X	X
High	X	X	X	/////	X	X	X	X	X
Low-High-Low	X	X	X	X	X	//////	X	X	X
High-Low-High	X	X	X	//////	X	X	X	X	//////

Table 3: Comparison of MOE Fitness

Table 3 shows that during the validation runs between human and computational results, there was not enough evidence to reject the hypothesis that the compared data had an

equal mean indicated by (X). Statistically significant differences occurring are indicated by (/////) (p value < 0.05). Two of these differences occur at the initial time phase for a high workload. This indicates that there is a transient phase that neither the theory nor the model captures. This research captured the difference more by accident by not allowing an initial warm up for subjects. However, this appears to be a significant finding, since normal operators will presumably not have a warm up period. The difference could be modeled by giving fewer resources initially or by introducing a task difficulty factor that would require more resources to process the task at hand.

The differences could be eliminated by including a mechanism of sensitivity increment into RHPM based on the number of signals over different experiments. The differences between RHPM and experimental data are minor considering that there were 36 measurements (4 experiments * 9 MOEs) and only four differed from each other. There seems to be a perceptual learning effect for human subjects, which enables them to distinguish noise and signals more easily after a certain number of experiments or exposure to number of signals. This could be modeled by changing the values of signal and noise parameters (mean and variance) over time. Thus the sensitivity (or the ability to distinguish signals) would increase with gaining perceptual experience over time. The sensitivity decrement would still occur however it would start at a different point. This is certainly a topic for further research and ongoing discussion with vigilance researchers.

Another interesting finding is the start up effect in the high workload condition. It took subjects a while to re-adjust to four different tasks. Normally subjects get a warm up period before the experiment. However, in this case there was no warm up phase at all. This very closely resonates with operational monitoring tasks, a radar screen operator starts immediately working and might be prone to more errors initially before adapting to the task again. RHPM can be used to show that by adjusting parameters the differences in performance are minimized and thus these differences can enable us to gain insights into the explanation of the phenomenon. However, it would be questionable to just change some parameters without researching the theoretical implications. Hence a change to the structure of this model would only make sense if further human experiments validate the hypothesis.

Another important research question considered whether or not RHPM generated a reasonable range of behavior. Range of behavior can be measured as the standard deviation for the MOPs. RHPM is a stochastic model and there are further opportunities to introduce randomness. The original set up already shows very reasonable ranges approximating human performance especially for misses.

Standard Deviation LHL

	FalseAlarm RHPM	FalseAlarm Human
10	1.21	2.04
20	1.29	2.76
30	1.32	2.72
	Misses RHPM	Misses Human
10	1.38	2.30
20	1.33	1.75
30	1.88	1.92

Table 4: Comparison of the Standard Deviation in the LHL condition

Table 4 shows a comparison of the MOEs' standard deviations of RHPM and human subjects in the LHL condition. Human data is more dispersed, however the differences especially in misses are small. Variability of RHPM can easily be increased by introducing more stochastic elements into it. Coupled with a close approximation to the mean MOEs, RHPM is certainly neither mechanistic nor brittle. The random number generation allows for a repeatability of runs while RHPM's pseudo-randomness is making it very difficult to precisely predict the next outcome.

VI. Conclusions

This research suggested a new cognitive model that simulates individual reduced human performance. The human experiment shows evidence that personality traits (especially extroversion and agreeableness) do in fact influence vigilance performance. However, personality traits' multiple regression models only accounted for approximately a third of the variance in the data. We found evidence that a military population is not a normal population. Military population studies exude biases in certain personality traits. There were also hints at cross-cultural differences in further examination of these populations. (see [6] for details).

Furthermore, this research generated three notable hypotheses in terms of vigilance theory improvement:

- Humans need initial time to adjust to a vigilance task. This influence seems to correlate with the difficulty of a task or the overall workload, since this effect was very pronounced with high workload.

- There are two forces influencing the sensitivity: One is the known decrementing force over time. However, there could be an incrementing force correlating with the number of perceived signals. The influence of the latter one indicates a perceptual learning effect that gains more importance (compared to the decrement factor) over time.

- The sensitivity decrement, as well as the shift of response bias, has limits. It appears likely that the rate of change towards these limits decreases which would be a possible explanation for the leveling off effect.

By looking at the performance variation it is very obvious that the average assumption for behavior of performance degradation is neither true for a single population nor for cross-cultural populations. The pitfalls of mirror imaging (thinking and even modeling that others should think and act like ourselves) loom behind simulation systems that do not take these differences into consideration.

Further contributions contain evidence that a paradigm shift in human behavior modeling taking vagary into account is suggestive. The proposed framework for the next-generation cognitive architecture has shown advantages in terms of robustness and being adaptive. The open and flexible architecture shows a possible path of cooperation between modelers. The implemented parts of the cognitive framework show their contribution by modeling the challenging problem of vigilance decrement.

RHPM has been validated with quantitative and qualitative analysis. The model has limitations and potential improvements were mentioned (again [6] shows weaknesses and strengths in more detail). These improvements should occur in cooperation with vigilance researchers.

This research started with the hypothesis that human performance can be modeled with a complex adaptive system. Cognitive modeling with complex adaptive system is a new approach that has yet to show its value. This research contributes to its valid claim of being a new promising avenue by successfully modeling the phenomenon of vigilance decrement. It is possible to harness a complex adaptive system in such a way that it can produce desired emergent behavior; in our case the realistic occurrence of a human-like vigilance decrement. The inherent capability of CAS to learn and to adapt to an ever changing environment seems to be an ideal fit to human performance modeling. However, the implementation of these ideas is not easy since some of the mechanism can only be modeled rudimentarily. This successful proof–of–concept implementation needs to be explored further as it is only evidence for a promising avenue.

RHPM can help to gain more insights into the phenomenon of vigilance decrement and more generally into human performance degradation. It appears to be a step in the right direction. There are many potential applications (civil and military) for a model that reliably simulates reduced human performance. The military applicability includes modeling individuals that conduct monitoring tasks such as radar screen operators.

Especially in a network centric warfare environment, it will be critical to understand the imperfect human information process and to be able to model human behavior. This begins by being able to learn more about the target acquisition and engagement process of infantry soldiers (and how to model them realistically) up to developing procedures of new command and control processes, utilizing modern information technology. The process needs to take human limitations into account in order to develop the right procedures with the right tools. There are known phenomena that can occur if humans are

overwhelmed by information. One example of the negative consequences of information overflow is that decisions are made prematurely indicating a bias that is not going to change with additional information. Another consequence is the delay of decision making because information is still flowing. Such a delay can lead to devastating consequences in military operations. A model capturing human deficiencies can help to improve theory and also help researchers to prioritize human experiments based on computational results. Thereby the model can help generate insights, guide theory development (i.e. via concept development and experimentation processes), and utilize resources more efficiently by applying simulation method to real world problems.

References:

[1] Chorev, M., Colonel Israel Armed Forces (1996). *Surprise Attack The Case of the Yom-Kippur War*. Washington, Industrial College of the Armed Forces.

[2] Pew, R. W. and A. S. Mavor (1998). *Modeling Human and Organizational Behavior: Application to Military Simulation*. Washington D.C, National Academy of Science.

[3] Ritter, F. E., N. R. Shadbolt, et al. (1999). *Techniques for Modeling Human Performance in Synthetic Environments: A Supplementary Review*. Nottingham, UK, ESRC Centre for Research in Development, Instruction and Training: 1-92.

[4] Tenney, Yvette et al., (2003) , *The AMBR Project: A Case-Study in Human Performance Model Comparison*, 2003 Conference on Behavior Representation in Modeling and Simulation (BRIMS) Scottsdale, AZ.

[5] Waldrop, M. M. (1992). *Complexity: the Emerging Science at the Edge of Order and Chaos*. New York, Touchstone.

[6] Wellbrink, Joerg (2003), *Reduced Human Performance Modeled as a Complex Adaptive System*, Dissertation MOVES Institute, Monterey, Naval Postgraduate School.

[7] Newton, I. (1729). *Mathematical Principles of Natural Philosophy*. http://www.marxists.org/reference/subject/philosophy/works/en/newton.htm accessed Jan 2004.

[8] Gell-Mann, M. (1994). *The Quark and the Jaguar Adventures in the Simple and the Complex*. New York, W.J.Freemann and Company.

[9] Arthur, W. B. (1994). *Inductive Reasoning and Bounded Rationality (The El Farol Problem)*. Amer. Econ. Review (Papers and Proceedings).

[10] Cowan, G. A., D. Pines, et al. (1994). *Complexity: Metaphors, Models, and Reality*. Boulder, CO, Westview Press.

[11] Arthur, W. B. (1999). "Complexity and the Economy." Science 284: 107-109.

[12] Grilo, A., A. Caetano, et al. (2000). *Immune System Simulation through a Complex Adaptive System Model*, CiteSeer Scientific Literature Digital Library.

[13] Ilachinski, A. (1997). *Irreducible Semi-Autonomous Adaptive Combat (ISAAC): An Artificial-Life Approach to Land Warfare (U).* Alexandria, VA, Center for Naval Analyses.

[14] Horne, G. E. and M. K. Lauren (2000). *Operational Synthesis Applied to Mutual NZ/US Questions. Part I: Quantico*, VA, Marine Corp Combat Development Command: 1-10.

[15] Hiles, J., VanPutte, M. et al. (2001). *Innovations in Computer Generated Autonomy at the MOVES Institute.* Monterey, Modeling, Virtual Reality and Simulations (MOVES) Institute, Naval Postgraduate School.

[16] Tosey, P. (2002). *Teaching at the Edge of Chaos*. Teaching & Learning Forum for Colleagues in the School of Educational Studies, Surrey, UK.

[17] Matthews, G., D. Davies, et al. (2000). *Human Performance Cognition, Stress and Individual Differences*. East Sussex, UK, Psychology Press.

[18] Mackworth, N. H. (1950). *Researches on the Measurement of Human Performance*. London: HMSO, Medical Research Council.
[19] Davies, D. and G. S. Tune (1970). *Human Vigilance Performance*. London, Trinity Press.
[20] Davies, D. and R. Parasuraman (1982). *The Psychology of Vigilance*. New York, Academic Press.
[21] Warm, J. (1984). *Sustained Attention in Human Performance*. New York, John Wiley & Sons.
[22] McKelvey, B. (2000). *Complexity Theory in Organization Science: Seizing the Promise or Becoming a Fad.* Emergence 1(1): 5-32.
[23] Wickens, C. (1992). *Engineering Psychology and Human Performance*. New York, Harper Collins.
[24] Wickens, C. (2002). *Multiple Resources and Performance Prediction*. Stanton Handbook, CA.
[25] Costa, P. T. and R. McCrae (2000). *NEO Software System for Windows Manual*. Odessa, FL, PAR Psychological Assessment Resources Inc.
[26] Carley, K. M. (1996). *Validating Computational Models*, Department of Social and Decision Sciences Carnegie Mellon University.
[27] Goldberg, D. E. (1989*). Genetic Algorithms in Search, Optimization & Machine Learning*. New York, Addison Wesley.