SAND2006-7911

SANDIA REPORT

SAND2006-1111
Unlimited Release
Printed December 2006

Human Performance Modeling for System of Systems Analytics: Combat Performance-Shaping Factors

Dwight P.Miller and Craig R. Lawton

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2006-1111
Unlimited Release
Printed December 2006

Human Performance Modelling for System of Systems Analytics: Combat Performance-Shaping Factors

Dwight P. Miller and Craig R. Lawton

System Sustainment and Readiness Technologies Department
Sandia National Laboratories
POB 5800
Albuquerque, New Mexico 87185-1011

Abstract

The US military has identified Human Performance Modeling (HPM) as a significant requirement and challenge of future systems modeling and analysis initiatives. To support this goal, Sandia National Laboratories (SNL) has undertaken a program of HPM as an integral augmentation to its system-of-system (SoS) analytics capabilities. The previous effort, reported in SAND2005-6569, evaluated the effects of soldier cognitive fatigue on SoS performance. The current effort began with a very broad survey of any performance-shaping factors (PSFs) that also might affect soldiers' performance in combat situations. The work included consideration of three different approaches to cognition modeling and how appropriate they would be for application to SoS analytics. This bulk of this report categorizes 47 PSFs into three groups (internal, external, and task-related) and provides brief descriptions of how each affects combat performance, according to the literature. The PSFs were then assembled into a matrix with 22 representative military tasks and assigned one of four levels of estimated negative impact on task performance, based on the literature. Blank versions of the matrix were then sent to two ex-military subject-matter experts to be filled out based on their personal experiences. Data analysis was performed to identify the consensus most influential PSFs. Results indicate that combat-related injury, cognitive fatigue, inadequate training, physical fatigue, thirst, stress, poor perceptual processing, and presence of chemical agents are among the PSFs with the most negative impact on combat performance.

Contents

Contents (Continued)

Figures

Tables

Acronyms or Abbreviations

ACT-R – Adaptive Control of Thought

AFQT – Armed Forces Qualification Test

BCT – Brigade Combat Team

C2V – Command and Control Vehicle

COGNET – Cognition as a Network of Tasks

EPIC – Executive-Process Interactive Control

FCS – Future-Combat Systems

HACE – high-altitude cerebral edema

HAPE – high-altitude pulmonary edema

HPM – Human Performance Modeling

HOS – Human Operator Simulator

M113 – Family of armored, tracked military vehicles with more than 40 variants

M&S – Modeling and Simulation

MIDAS – Man-Machine Integrated design and Analysis System

MOE – measure of effectiveness

MOS – Military Occupational Specialties

NDS – Nonlinear Dynamical System

NBC – Nuclear/biological/chemical

NVG – night-vision goggles

OMAR – Operation Model Architecture

PERFECT – Performance Effectiveness for Combat Troops

PSF – Performance-Shaping Factor

SAMPLE – Situation Awareness Model for Pilot-in-the-Loop Evaluation

SME – Subject-Matter Expert

SNL – Sandia National Laboratories

SoS – System of Systems

SoSAT – System of Systems Analysis Toolset

SOAR – State, Operator, and Result

UA – Unit of Action

1. Introduction

Current system of systems (SoS) analysis efforts at Sandia National Laboratories (SNL) capture multiple performance attributes of military hardware systems, such as mobility and lethality in the context of user-defined scenarios. This capability is state-of-the-art and unique to SNL. However, the analysis capability is focused primarily on combat technological and sustainment systems; the contributions and performance capabilities of human operators are not included. The problem with this approach is that humans fight wars, not machines. Platforms must be driven into position and operated by humans to wage war. The truth of the matter is that research on human reliability in the military informs us that at least 50 percent of system failures are attributable to human error (Ref.1). In systems of high consequence and safety regulation, such as commercial airlines and nuclear power, the system failures attributable to human error is significantly higher. Not including humans in a SoS analysis will likely lead to overlook the largest performance factor of the SoS simulation. With no human element represented, it would be equivalent to assuming all soldiers performed perfectly all the time. This assumption inevitably leads to the overly optimistic (i.e. non-conservative) results in availability or performance analyses. As the human elements of SoS models become more realistic by reflecting the effects of the environment, weather, adversary actions, and intrinsic soldier variables on soldier performance, and hence system performance, the results of SoS analyses will become increasingly more accurate.

Last year (FY05) the authors investigated what they thought to be the most influential performance-shaping factor[1] (PSF) or behavior modifier, in military scenarios, soldier cognitive fatigue (Ref. 2). A simplistic, stepwise model of soldier fatigue was incorporated into a SoS analysis with dramatic effect. The operational unavailability of C2V platforms rose by 30% when it was assumed that soldiers performed the 72-hour mission without sleep. Having proven the importance of soldier fatigue on system availability, the current report addresses a wide spectrum of combat-related PSFs other than cognitive fatigue that could significantly impact operational availability and performance in large military SoS analyses.

2. Recent Advances in Human Performance Modeling

In the last two and a half decades significant effort has been put into developing models that mimic human cognitive behavior for the purposes of military modeling and simulation (M&S). While it is beyond the scope of this report to summarize all of the work, there are several summaries available in the literature that stand out as authoritative

[1] SNL researchers have traditionally used the term “performance-shaping factor”, however the bulk of the military researchers in this area use the term “behavior moderator.” We will use the former term and its intialism, PSF, throughout this report.

sources. First and foremost is Pew and Mavor's 1998 text (Ref. 3). While it is now somewhat dated, it has remained a solid source reporting the findings of a blue-ribbon panel organized by the National Research Council charged with studying the state-of-the-art in human-performance modeling (HPM) relating to military M&S of cognition, team, and organizational behavior. Pew and Mavor describe and review all of the major integrative architectures extant at the time, including ACT-R, COGNET, EPIC, HOS, MicroSAINT, MIDAS, OMAR, SAMPLE, and SOAR. They also address topics prevalent in the need to model human behavior in military contexts, such as decision-making, situation awareness, planning, learning and memory, attention and multitasking, and unit-level modeling.

The two most popular and most reviewed cognitive architectures are ACT-R and SOAR. The following descriptions of ACT-R and SOAR were provided by Michael L. Bernard of SNL in a recent personal correspondence.

2.1 ACT-R

The original purpose of ACT-R[2] was to model problem solving and learning (for example, predicting the degree and speed of learning when operating a novel operating system). Because of this focus, modeling of the effects (limitations) of working and long-term memory is emphasized. ACT-R is a production-system theory that models the steps of cognition by a sequence of production rules that fire to coordinate retrieval of information from the environment and from "memory." There are two types of knowledge representation in ACT-R—declarative and procedural knowledge. In ACT-R, declarative knowledge is represented in structure called chunks whereas procedural knowledge is represented in productions. Chunks and productions are the basic building blocks of an ACT-R model.

ACT-R generally consists of a goal stack that encodes the hierarchy of intentional guiding behavior, a procedural memory containing production rules, and a declarative memory containing chunks of information. These are all organized through the current goal that represents the focus of attention. The current goal can be temporarily suspended when a new goal is pushed on the stack. Productions are selected to fire through a conflict resolution process that chooses one production from among the productions that match the current goal. Productions consist of a series of condition-action pairs of "if-then" rules. Conditions specify a pattern of chunks that must be present in the buffers for the production rule to apply. The interface between procedural memory and other components are accomplished through buffers. Each buffer can hold one chuck at a time, and the actions of a production affect the contents of the buffers. A buffer consists of a set of patterns to match against the current buffers' contents. If all of the patterns correctly match, then the production is said to match and can be selected.

[2] ACT-R stands for Adaptive Control of Thought

```
(P example-counting                 English Description
  =goal>                            If   the goal is
    isa       count                          to count
    state     counting                       the current state is counting
    number    =num1                          there is a number we will call =num1
  =retrieval>                                and a chunk has been retrieved
    isa       count-order                    of type count-order
    first     =num1                          where the first number is =num1
    second    =num2                          and it is followed by another
                                             number we will call =num2
==>                                 Then
  =goal>                                     change the goal
    number    =num2                          to continue counting from =num2
  +retrieval>                                and request a retrieval
    isa       count-order                    of a count-order fact
    first     =num2                          for the number that follows =num2
```

A chunk can be created explicitly in the initial conditions for the model, as the result of a perceptual request, or as a new goal. Every chunk in ACT-R has a numerical value called activation. The activation reflects the degree to which past experiences and current context indicate that chunk will be useful at any particular moment. When a retrieval request is made, the chunk with the greatest activation among those that match the specifications of the request will be the one placed into the retrieval buffer.

2.1.1 ACT-R's Strengths and Weaknesses

While ACT-R is a powerful cognitive computational architecture, which can be used in training research, there are several limitations to this approach. ACT-R rests on the theory of unified cognition, and like most theories, it is constantly being revised and updated as further empirical data is gathered. While the general architecture of ACT-R was state-of-the-art 20 years ago, newer models of cognition have been developed. ACT-R can be difficult to learn and it can be an arduous process to build a model of a complex task. For example, the modeler has to input the declarative memory and the productions in order to build the model. For complex tasks, this may require a hundred productions or more, which can be a time consuming process. Moreover, the current iteration of the model does not include emotions or higher-level goals.

2.2 SOAR

Like ACT-R, SOAR[3] is an attempt to build a unified theory of cognition. SOAR seeks to describe and realize the fundamental, functional components of intelligence. In the last 10 years, much of SOAR research has focused on building intelligent systems that interact autonomously with complex environments that are inhabited by other entities, including other intelligent agents and human agents. The continuing thread of SOAR research has been to search for a minimal set of mechanisms that are sufficient to realize the complete range of intelligent behavior.

---

[3] SOAR stands for State, Operator and Result.

SOAR uses an associative mechanism to identify knowledge relevant to current problems and to bring that knowledge to bear on potential solutions. A symbolic pattern matcher compares a representation of the current context to the activation conditions for each element in the system's knowledge base. Any relevant knowledge activates, leading to further elaboration of the context. The agent's reasoning context includes its current beliefs about the state of the world, proposals to perform different tasks, actions to be executed, etc. Thus, the flow of control in SOAR is determined by the associations made in memory. Because many of these associations can be made independently from each other, SOAR allows them to occur in parallel. The functional role of parallelism is that many types of relevant knowledge can be brought to bear on a problem. Thus, for a single problem, a SOAR agent can simultaneously consider problem decomposition, analogy, experimentation, or other solution methods.

SOAR employs a computationally inexpensive truth maintenance algorithm to update beliefs about the world. For example, a pilot agent might believe that an enemy aircraft is pointed at it. This belief would have been derived from environmental inputs such as the enemy heading. When the enemy's heading changes, the agent will automatically update its belief about the enemy.

While it is useful to be able to choose between options, an agent could find itself in a situation where it recognizes no available options or cannot decide between options. SOAR recognizes such a conflict, marks it as a reasoning "impasse", and automatically creates a sub-goal to resolve the impasse. This automatically generates a new problem context on which the agent can bring new knowledge to bear. The agent might use planning knowledge to consider possible futures and determine an appropriate course of action. It might make a choice by comparing this situation to others it knows about and then drawing an analogy. Automatic sub-goaling gives SOAR agents a meta-level "reasoning" capability, the ability to reason about their own reasoning. Further, SOAR's meta-level reasoning is not a separate component of the overall system.

In addition to not knowing what to do, or how to select among alternatives, an agent may be faced with knowing what it wants to do, but not having any immediate operational way to achieve that goal. For example, once a pilot agent decides to intercept a particular target, it still must make a series of decisions about *how* to prosecute the intercept. This situation also results in an impasse in SOAR. In this case, there is a selected commitment, and the impasse (of not having an immediate way to achieve that commitment) creates a goal to find a way to achieve it. The impasse goal corresponds to a task goal, such as "execute a mission" or "intercept a target". Importantly, SOAR makes a distinction between the goal itself and how the goal is achieved. Thus, the same goal can be attacked in different ways at different times, depending on the situation.

Automatic sub-goaling enables task decomposition. At each step in the decomposition, the agent is able to focus its knowledge on the particular options at just that level and ignore considerations at other levels. The process of decomposition narrows a potentially exponential number of considerations into a much smaller set of choices, which is important for agents that have finite computational resources and that must remain reactive to the environment. Recognizing and focusing on only knowledge appropriate to the current problem is the essence of the problem space hypothesis, one of SOAR's theoretical foundations.

2.2.1 SOAR’s Strengths and Weaknesses

SOAR's mechanism of adaptation is both conceptually simple and powerful. All impasses represent a lack of immediately applicable knowledge. However, once the agent reflects and comes to a decision that resolves the impasse, it summarizes and generalizes the reasoning that occurred during the impasse. This process results in new knowledge that will allow the agent to avoid an impasse when encountering a similar situation. SOAR's learning mechanism has been used to: learn new conceptual knowledge, learn new procedures, and correct its knowledge as it gains feedback through experience in its environment. There is a variety of different styles of learning that an intelligent agent might exhibit. Past research demonstrates that SOAR’s simple generalization method can implement many of these styles. The research hypothesis is that SOAR’s method can implement *all* of them. However, the precise way to implement some styles of learning is still a research question, and even some well-understood methods are still difficult to implement in individual SOAR agents. Other adaptive intelligent architectures pragmatically finesse the difficulty of having a single, primitive method for adaptation by directly implementing a variety of high-level learning styles.

The downside of the SOAR approach is that, by dictating very general mechanisms, it under-specifies the capabilities that must be built into intelligent agents. Most of an agent’s competence arises from the encoded knowledge (i.e., the set of rules) that SOAR’s mechanisms operate on. Thus, agent knowledge must be created to realize any high-level intelligent behavior. For instance, SOAR has been used to build planning systems. In comparison to other AI planning systems, SOAR offers little immediately evident power. SOAR only specifies very low-level constraints on how planning can occur, so SOAR agent designers must develop their own plan languages and algorithms, while these are provided in most planning systems. However, what SOAR does provide is a natural, scalable methodology for integrating planning with plan execution -- as well as natural language understanding, reasoning by analogy, etc. Too often, AI researchers develop powerful single-purpose systems that then create new problems when different subsystems must be integrated to realize a complete intelligent agent. By focusing on a uniform substrate that allows knowledge to mediate any decision, SOAR provides a ready tool with which to realize integrated approaches. SOAR therefore trades off powerful, but overly constrained processes for the flexibility to integrate solutions.

2.3 Sandia’s Cognitive Modeling Framework

More recently, J. Christopher Forsythe and colleagues at SNL have developed a cognitive framework that models how an individual sees the world, associates concepts, and responds to patterns of complex stimuli. Unlike the former models that represented a generic human performer (ACT-R, MIDAS, SOAR), the SNL approach grounds each model in conceptual/associative data taken from an individual, customizing each representation to a particular person's knowledge and conceptual associations. It has also been used to represent a cognitive collective, a group of individuals who need to perform a group task, such as evaluating applicants as security threats. The SNL approach has been successful in modeling individuals responding to air-combat threats, simulating security officers' responses to facility intruders, and reflecting individuals' pattern-analysis styles in a collective task of recognizing a security threat from mined behavioral data. The following description of the Sandia Cognitive Model Framework was provided by J. Chris Forsythe in a recent personal correspondence.

The Sandia cognitive framework is a modular software architecture that includes various component classes needed to construct human-like computational cognitive models. The component classes correspond to functional subsystems essential to construct a cognitive model. Figure 1 shows the modules and their associations.

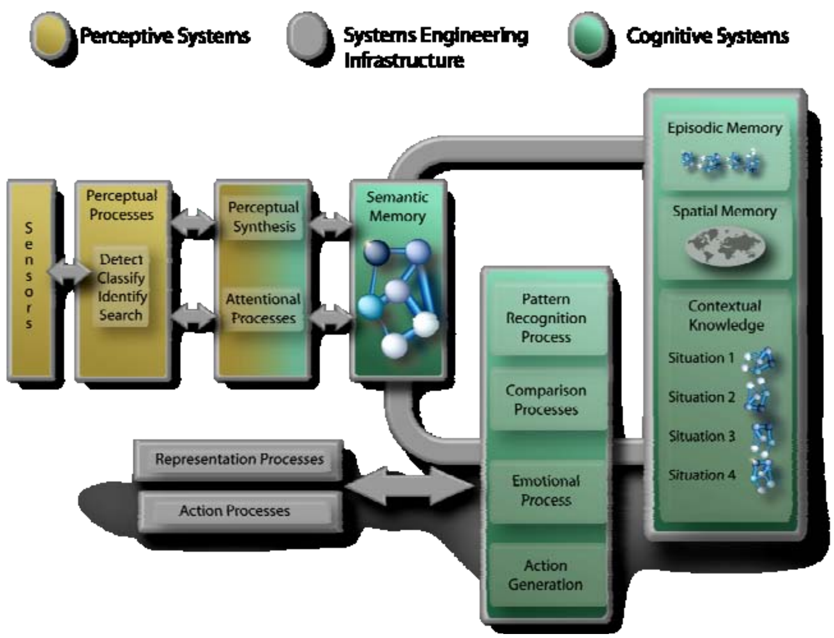

Figure 1. Schematic representation of Sandia's Cognitive Model Framework.

A key component is context recognition. This component utilizes the activation of concepts within semantic memory to recognize patterns learned through past experiences. From an engineering perspective, the context recognition component is a type of Nonlinear Dynamical System (NDS). The inputs to the NDS are processed sensory inputs and the outputs are estimated activation levels of the various contexts as a function of time. The estimated context activations can be considered a trajectory through the state space of the NDS. Other components of the framework include episodic memory for capturing a record of past experiences that provides the basis for context acquisition, spatial memory, which allows the recognition of contexts to be influenced by the location of events (i.e. geospatial coordinates of images) and a comparator that detects and responds to anomalous events.

The basic unit in our cognitive framework is the ‘concept.” Concepts correspond to the most elementary units of cognition enabling an entity to recognize and respond to stimuli. Within this framework, cognition begins with recognition of meaningful regularities within an entity’s sensory experience of their external and/or internal world. These regularities may involve almost any stimuli detectable to the entity’s sense organs, as well as combinations of stimuli that may span sensory modalities and temporal durations.

Many factors may contribute to a regularity being meaningful. These include association with reward, predictive value, causative relationship, distinguishing attribute, cue in guiding motor processes, etc. The foregoing assumes an ability to acquire meaningful regularities and retain the knowledge needed for their subsequent recognition. Once learned, the internal representations of these regularities correspond to the concepts that serve as the basic unit for cognitive processes.

2.3.1 Semantic Memory

Within the cognitive framework, knowledge of concepts (i.e., regularities) is represented within semantic memory. However, an additional function of semantic memory is to store knowledge of the “relatedness” of concepts. Relatedness refers to the awareness that two concepts are somehow associated whether one is a property of the other, they are members of the same category, they operate together in some functional relationship, etc.

2.3.2 Context Recognition

Everyday experience is more than a complex array of differential concept activations. Instead, we recognize patterns within this complex array. Within the cognitive modeling framework, we have referred to these patterns as “contexts.”

At a given moment in time, the overall experience of the world may consist of the somewhat separable experiences that I’m in my home, in my kitchen and eating breakfast. Home, kitchen and eating breakfast may each represent a context. If we consider any one of these contexts, there are corresponding concepts (e.g., morning, orange juice, cereal, wearing pajamas, reading the newspaper, etc.) Each of these concepts is predictive of the context “eating breakfast,” but in isolation, it is unlikely that any of these concepts alone would be sufficient for a meaningful inference. However, to the extent that the concepts present in a given scenario correspond to the pattern associated with the context “eating breakfast,” it may be inferred that the context is “eating breakfast.” As illustrated, a given scenario may consist of multiple contexts that occur simultaneously (e.g., in my home, in my kitchen and eating breakfast). Additionally, contexts may involve differing levels of abstraction (e.g., in my home and in my kitchen).

Context recognition occurs through a process based on evidence accumulation. Here, the relevance (i.e., evidence derived from the activation of concepts within semantic memory) of contexts to the current situation is continually updated. Transitions between contexts occur when evidence for a context exceeds or falls below the threshold for recognition. The experience of a given situation is then based on the combination of contexts that each exceeds some threshold for recognition.

There is an additional feature of the context recognition mechanisms that should be mentioned. There is knowledge accompanying contexts that concerns “expectations.” Knowledge of expectations consists of concepts that are likely to occur in conjunction with a context. For example, coffee would be an expectation for the context “eating breakfast.” Consequently, given other concepts provide sufficient evidence for recognition that the context is “eating breakfast,” any dark liquid within sufficient proximity is likely to be perceived to be coffee. Expectations correspond to a “top-down” influence and lead to a priming of concepts in semantic memory, which may occur despite an absence of corresponding sensory events, following recognition of a given context.

2.3.3 Episodic Memory

As we progress through life, we create a record of our experiences. This record is referred to as episodic or autobiographical memory. Here we make a distinction in that episodic memory refers to the actual record of events stored to memory, whereas autobiographical memory refers to the recall that occurs as a product of constructive processes involved in accessing the episodic record.

Our own instantiation of episodic memory consists of a record of experience based on the activation of other components of the model. The record is not continuous, but instead, certain events may trigger an entry to episodic memory. For instance, entries correspond to the recognition of a context or the transition from one context to another. Likewise, activation of concepts in semantic memory may also trigger an entry, as well as activation of other components that have not yet been described (i.e., comparator and emotional processes). An entry to episodic memory consists of a reference to items activated at the time corresponding to the entry and a record of the relative activation of each item. For instance, with semantic memory, this would consist of the concepts that were activated and their relative levels of activation.

The contents of episodic memory are not stored as distinct episodes with clearly defined beginnings and ends, but instead episodic memory is represented as a continuous record of events. For any given point in time, the episodic record may consist of several simultaneously activated contexts. “Episodes” are a product of episodic memory retrieval. Based on retrieval cues, a slice of the episodic memory record is recalled that may have a distinct beginning or end, with the beginning and end typically corresponding to transitions associated with particularly salient contexts. Thus, in our framework, “episodes” are a product of episodic memory retrieval, and not the basis for structuring episodic memory.

Episodic recall utilizes an evidence accumulation-based mechanism similar to that previously described for context recognition. In particular, the episodic memory record is evaluated with regard to the relative correspondence to one or more memory retrieval cues. For example, if the retrieval cues consist of “baseball game” and “cold night,” the episodic record would be evaluated with regard to the correspondence to these cues. Periods involving both cues, in general, would receive more evidence than other time intervals containing one or neither cue. For periods in which both cues occur, the evidence for a time interval would be a function of the level of activation associated with each cue during the time interval. For example, if it was an extremely cold night with significant discomfort, “cold night” may have been an extremely salient cue.

However, of tremendous significance, the episodic record has a level of “accessibility” associated with it that modifies the evidence derived for a given time interval with respect to such factors as the amount of time that has passed. Recall occurs for the time interval or intervals receiving the most evidence.

At recall, an episode derived from the episodic record may be replayed with activation of the constituent elements of the episode. For example, concepts in semantic memory may be activated. Similarly, while not incorporated in our current models, given a perceptual memory containing rich representations of perceptual entities, these perceptual representations would be activated. It is worth noting that the activation of constituent elements of episodic memory is somewhat consistent with events as they occurred at the time of the episode, but is also a product of current knowledge. As a result, the recalled experience may differ in some regards from the original experience.

2.3.4 The Comparator

The comparator within our modeling framework monitors ongoing context recognition processes and maintains an awareness of current contexts, including expectations associated with each context. The comparator also monitors semantic memory and the concepts that are activated. The comparator is triggered when one or more concepts are activated in semantic memory, which are inconsistent with the current context(s). Additionally, an adjustable threshold is provided that allows the model to be more or less sensitive to out-of-context events to enable situational influences (e.g., lowered sensitivity following positive events or successful goal attainment and heightened sensitivity following negative events or failed goal attainment) or individual differences.

When the comparator is triggered, there is heightened attention placed on the out-of-context event. This occurs through a boost in activation for the out-of-context concept in semantic memory accompanied by a generalized dampening of the activation of all other concepts. As a result, in some cases, the model may abandon the context that had been previously recognized. In other cases, the evidence may be sufficient that a recognized context is not abandoned despite the out-of-context event, although the experience may prompt learning that incorporates the previously out-of-context event, or acquisition of a new context.

2.3.5 Emotional Processes

Our modeling framework has made provisions for separate components that each correspond to distinct emotional processes. For instance, a model may be configured that has distinct components for emotions corresponding to pleasure, dysphoria, frustration-anger, fear, anxiety and disgust. Each emotion may have a behavioral correlate, but in the current framework the greatest impact of emotional processes occurs through its influence on information processing.

Concepts within semantic memory and contexts within contextual knowledge may be associated with an emotional process. For example, the concept “snake” may be associated with the emotional process corresponding to fear. When the concept “snake” is activated, this triggers the emotional process corresponding to fear. As a result, similar to the comparator described above, there is a heightened activation of the concept that triggered the emotional response accompanied by a generalized inhibition of other concepts. This serves to focus attention on the fear-inducing stimulus to the exclusion of other concepts and may prompt a reinterpretation of the ongoing context.

2.3.6 Sandia’s Cognitive Model’s Strengths and Weaknesses

A primary strength of the Sandia cognitive framework is its emphasis on context recognition. This makes the framework particularly well suited for modeling cognitive processes that involve pattern recognition, intuitive judgment and implicit knowledge. The framework has been tailored to support automated knowledge capture providing the unique utility of being able to quickly and easily derive models of specific individuals. Furthermore, the framework has been implemented in a variety of practical applications that involve integration with tools for automated knowledge capture.

However, the Sandia approach is not a generic model, but one that is populated by an individual’s concept structure. Thus, the model cannot simply applied to represent an entity without first collecting personal data and populating the model. To date, there has been limited experimental validation of the Sandia cognitive framework. While the framework emphasizes context recognition, it currently has no mechanism to represent deliberative processes or executive functions. Finally, the framework is an experimental platform with limited availability outside of Sandia that requires significant time investment to begin using. Thus, it has not benefited from there being a community of users and developers, as exists with other modeling frameworks.

2.4 Simplified Approach Needed

The common trait in the models and architectures referred to above is that they attempt to model the performance of an individual performing a task or set of tasks, given environmental conditions and input information. While this level of complexity and detail is appropriate for evaluating new system designs or graphical user interfaces, it can be computationally overwhelming for applications where thousands of humans are needed, such as in a SoS analysis of a US Army brigade, or in Future-Combat Systems (FCS) parlance, a Unit of Action (UA). Even if it were computationally tractable, each soldier's conceptual world model would need to be populated from collected data, and situational inputs would be required for all of the soldiers represented in the modeled scenario. This approach may be the most valid for the few soldiers (high ranking officers) in decision-making roles in the SoS analysis, who make critical decisions regarding the direction of the simulation. In a UA, approximately 9 percent of the forces are officers and the remaining 91 percent are infantry and armor specialists. One approach to be considered is modeling the top-level UA decision-makers with cognitive models.

2.4.1 Human-Reliability Modeling Assumptions

One notion of looking at soldier performance is that if there were no PSFs in effect, the soldier would perform all tasks perfectly, that is, within time constraints and without errors. This is analogous to the previous state of SoS analysis, where there was no modeling of human elements. This approach is also analogous to equipment reliability and availability assumptions—that is, when no faults are extant, performance is as designed, no better, no worse. The notion of a perfect soldier gives us a conceptual baseline to which we can compare more realistic estimates of soldier performance. That is, as more PSFs take effect and get modeled, we would expect more negative impact on [perfect] soldier performance. While we realize that on rare occasions acts of heroism and feats of superior performance can sway the balance of a battle, these anomalies are extremely rare and difficult to model. Perhaps the assumption that, under ideal circumstances, all soldiers perform their tasks on time and without errors, balances the inability to model the occasional act of superior performance.

2.4.2 Our Strategy

What is needed is a simpler set of assumptions or models of human performance that is less detailed and easier to replicate for thousands of soldier entities in a large SoS simulation. The difficult remaining tasks associated with this approach are: 1) identifying which PFSs hold significant impact on soldier performance in combat, 2) deciding which PSFs have the highest impact on a given set of combat tasks, 3) determining if there are interactions among the most significant PSFs that occur simultaneously, 4) finding, developing and using valid dose-performance curves in estimating impact of the PSFs on soldier performance, and 5) assigning appropriate levels of those factors to the affected soldiers during those periods of performance in a combat scenario in an SoS simulation. The remainder of this report concentrates on the first two of these tasks.

3. PSFs Addressed in the Literature

Significant efforts have been devoted to attempting to understand the relationships between PSFs and performance in military operations, both in peacetime and during combat. In the push for combat efficiency, the military services have spent untold millions on selecting people that make better soldiers, training them to be better soldiers, and designing war-fighting weapons, equipment, and systems to accommodate human capabilities and simplify military tasks. It is no coincidence that the field of human factors engineering began with military applications in the 1940s and today is institutionalized throughout military system procurement processes. The philosophy is that if you take into account the capabilities and limitations of humans while designing their equipment and weapons, the improved person-system interface will almost always produce better performance. Human factors taken into account for military system design usually include physical strength; body size and proportions; cognition (knowledge, thinking, and decision-making); perceptual abilities; dexterity, manual control and coordination; language skills; circadian factors; and physiological performance in extreme environments (e.g. in deep water, high altitude, very hot and cold). The following sections provide a brief review of the research literature addressing PSFs and their impact on human performance in combat.

3.1 Classification of PSFs

Several authors have attempted to simplify the treatment of PSFs by pointing out that there are several basic categories to consider: external, internal and task-specific PSFs (Ref.s 3 and 4). “External” PSFs comprise those that originate outside the human and typically have negative impacts on performance. These include environmental factors such as heat, cold, altitude, toxins, noise, vibration, etc. “Internal” PSFs are those specific to the individual, such as intelligence, personality, cognitive styles, expertise, cultural values, emotional factors, and preferences. They are difficult to assess and often combine in ways that are hard to predict. “Task-specific” PSFs include task history, location of engagement, boredom, stress, threats, casualty level, task type, availability of on-site task training, interface quality, task-related fatigue, battle conditions, and local casualties (Ref. 4). Beginning with section 3.3, three categories of PSFs are discussed in detail.

3.2 The Relationship Between Peacetime Measures and Combat Performance

To counteract the negative effects of internal PSFs, the military primarily uses selection and training. Selection weeds out undesirable factors that might make a candidate a weak or under-performing soldier, while “boot camp” and training attempt to minimize interpersonal differences by grinding down the ego and building up a physically rugged, team-oriented, command-following, appropriately informed soldier. Regardless, a few months of basic training cannot erase all of the individual differences in recruits. Despite the untold millions spent preparing soldiers for combat, there are surprisingly few peacetime predictors of how well a soldier will perform in combat.

Dover (Ref. 5) studied the relationship of subjective estimates of soldier performance in “routine service” and in combat, the differential efficiency of selection scores in predicting routine vs. combat performance and the construct structure portraying combat soldier performance. Four groups of Israeli Defense Force soldiers were subjects in the study. They were evaluated by ratings obtained by their direct commander, as well as hard-data measures. Ratings of peacetime and combat performance showed significant, but moderately low correlations (r = .40 - .50), with higher correlations for professionalism and promotion, but lower correlations for factors such as work regimen and discipline. Dover reports a study by King et. al (Ref. 6) correlating 43 West Point graduates' Aptitude for Service Ratings of their last year at West Point (r = .52) and final graduation score (r = .43) with combat performance in the Korean War. Grades in applied courses, such as tactics and electricity correlated more highly (r = .20, .24) than those for academic subjects such as mathematics and English (r = .01, -.02).

Dover's factor analysis for combat performance of 1279 experienced combat soldiers' scores on a Soldiers Peacetime Performance Evaluation (questionnaire), taken 6 to 12 months after recruitment produced two factors explaining 65.8% of the variance; "promotion, professionalism and prospects for functioning in combat (59%) and "work regimen and functional performance" (6.8%). A second factor analysis of soldiers' scores on the Soldiers Combat Performance Evaluation shortly after combat revealed a two-factor solution explaining 73.4% of the variance. The dominant factor, including test items such as calm and collected, courage and coping with dangers, and sticking to the goal, was called "combat functioning." The second factor included test items including performance prior to combat, potential beyond squad leader, and technical and tactical abilities, and was called "routine functioning and promotion."

The conclusion that needs to be drawn from the preceding paragraphs is that we cannot assume that because we know something about a soldier's internal PSFs, such as his/her intelligence, marksmanship, aptitude, or strength, that we can assume it will have a predictable effect on combat performance. Therefore, in modeling the human elements of combat-scenario simulations, we need to put more emphasis on the external and task-based factors. With this caveat in mind, let us look at various PSFs considered for modeling in recent studies.

3.3 External PSFs

As mentioned previously, external PSFs comprise those that originate outside the human, including environmental factors such as heat, cold, humidity, altitude, toxins, noise, and vibration. Each of these will be looked at in terms of impact on performance.

3.3.1 Heat

Excessive heat in situations where heavy work is performed has been studied extensively to help develop protective guidelines for the commercial work force and to understand how it affects Army operations (Ref. 7). Dissipating metabolic (body) heat is a complex process that depends on the work being performed, acclimatization, the air temperature, the radiant heat present, the humidity, air speed, and the clothing being worn. A much simpler model is needed. The following guidelines may be applicable to our modeling problem.

Heavy work can extend for 3.5 hours at a temperature of 82 degrees F before heat exhaustion is expected; 2.2 hours at 88 degrees, and 90 minutes at 94 degrees. Military STD 1472D limits work at 102 degrees to 30 minutes, but suggests 12 hours is safe at 85 degrees (Ref. 3). Acclimatized soldiers can withstand the negative effects of heat for longer periods, however acclimatization cannot protect them indefinitely (Ref. 8). While one would not expect effective work to go from 100% to zero at the time limits identified, a set of exposure-performance curves is needed to model the negative effects of heat over time.

3.3.2 Cold

Military history teaches us that extremely cold environments can wreak havoc on the best planned campaigns. If proper insulated garments are not provided to the troops, frostbite and hypothermia can profoundly reduce the effectiveness of the ranks quickly. Cold weather has the effect of reducing tactile sensitivity and impairing performance on tasks requiring fine manual dexterity, such as the manipulation of knobs, switches, screws, nuts and bolts (Ref. 9). Hand-grip strength decreases 21-28% after three hours at -25 degrees C. Modern armies typically have thermally protective garments, and other means of keeping their soldiers out of extremely cold weather. However, as several authors point out, bulky protective clothing (e.g. arctic gloves) often trades the performance deterioration due to cold to that due to the lack of manual dexterity from the heavy gloves (Ref. 9). Military Standard 1472D limits work at 59 degrees F to four hours, and work at freezing to 30 minutes (Ref. 3). Acclimatization can occur within one week, however two or three weeks allows for steady-state performance. There is little support in the literature for the notion that cold can cause deterioration of morale, anxiety, increased irritability, depression, and sleep loss.

3.3.3 Noise

Noise is defined as any unwanted sound. Ambient noise of significant level can be an environmental stressor that can be detrimental to cognitive activity and performance. People vary enormously in their response to noise and can adapt to continuous background noise if below physiologically damaging levels. The primary effect is annoyance. Combat noise can be maximally disruptive, as it is unpredictable, highly variable, and of high intensity, however quantitative models of performance decrement due to extraordinary noise are not yet available (Ref. 3).

3.3.4 Vibration

Whole-body vibration, if of sufficient magnitude, can have serious performance effects on soldiers. For example, the tolerance limit for accelerations of 3 m/s^2 at 5 Hz is one minute, while at 0.3 m/s^2 at the same frequency is 8 hours. For complete curves showing the relationship between acceleration, frequency, and tolerance limits, see Ref. 10.

At very low frequencies, less than 0.5 Hz, motion sickness occurs. At 30 Hz, the resonant frequency of the eyes, vision is disrupted. Multimodal displays, using tactile and auditory redundancy have been used to make up for the performance decrement in using visual displays alone in high vibration environments. Manual control of input devices, such as cursors, buttons, and vehicular controls can also be negatively impacted by low frequency vibrations (Ref. 11).

Low-frequency vibration, combined with being confined in a visually restricting vehicle (e.g. tank or armored personnel carrier), have had disastrous effects on soldiers' abilities to use computer displays and perform combat tasks upon leaving the vehicle, due to motion induced sickness. Recent US Army studies have attempted to quantify the detrimental effects of being "on the move" while working with computer display systems in C2Vs and other vehicles (Ref. 11). The authors report a substantial degradation in individual and task performance while "on the move" and working at the four workstations in the back of the vehicle (See Ref. 12 for more details). The degradation would also carry over to many tasks performed immediately after being released from the confines of the vehicle.

There are some findings that vibration can negatively impact cognitive functions. Sherwood and Griffin (Ref. 13) discovered differences in rates of learning between static and vibrated groups, while Schipani et al. (Ref. 14) saw decrements in time sharing, memorization, inductive reasoning, attention, and spatial orientation after rides in a modified M113.

3.3.5 Terrain and Altitude

While anecdotal evidence would suggest that terrain and altitude play important roles in workload, there are no available literary sources verifying the effects of these two environmental stressors for the general infantry soldier. FM1-100 supplies us with valuable advice for operating Army helicopters in different terrains, including mountains, jungles, deserts, arctic areas, urban terrain, and nuclear/biological/chemical (NBC) environments (Ref. 15).

We know that at high altitudes (above 3,000 m), there is less oxygen and resulting hypoxia can be a problem. We also know that at very high altitudes (above 6,000 m) even highly trained mountain climbers can experience high-altitude pulmonary edema (HAPE) and high-altitude cerebral edema (HACE). However, these conditions are not those experienced by the typical US Army soldier. Difficult terrain, such as deserts, swamps, jungles, and mountains require more effort to traverse, however it is assumed that leadership would account for increased time and effort by providing mechanized transport or allowing more time and rest periods in difficult terrain.

3.3.6 Night Operations

In the past, before the development of night-vision goggles (NVG), operations conducted after dusk until before dawn were much more difficult to perform than daytime operations, i.e. they take longer to complete and more errors are committed. While the NVG technology has closed the gap considerably, the technology has not completely compensated for the effect of no or little ambient light. Assuming they are available for use, the NVGs amplify ambient light to where a user can perform adequately, however, the limited field of view, limited range, tracing effects, lack of detail and color contrast, bulk, and significant eye strain limit performance to that somewhat lower than using natural ambient light.

3.3.7 Toxic Substances

On the battlefield, if toxic substances are used as weapons, we can probably assume that one of three responses will occur: The toxin will be absorbed and the soldier will be severely incapacitated; the toxin will miss the soldier and no/mild negative effects will ensue; the soldier will successfully don the protective gear prior to exposure and his performance will suffer mostly due to wearing the NBC gear. Experimental findings on the effects of the NBC gear range from no decrement to the complete inability to perform any task directly (Ref. 8). The major stressors include the heat build-up inside the suit, which exacerbates fatigue and reduces the soldier's ability to think, maintain vigilance, and make decisions. Due to the heavy gloves, tasks requiring manual dexterity are severely compromised, as the gloves cannot be removed even for short periods. Additionally, the facemasks reduce peripheral vision and visual acuity when in place (Ref. 16).

3.4 Internal PSFs

As mentioned earlier, internal PSFs comprise those that originate inside the human, including intelligence, cognitive styles, expertise, cultural values, personality, emotional factors, and preferences. These and others will be considered in terms of impact on performance and typical modeling approaches.

3.4.1 Intelligence

The Armed Forces Qualification Test (AFQT) is used to place recruits into mental-performance categories, ranging from Category I to Category V (high to low). The AFQT has been used to assign recruits into compatible jobs and evaluate correlations between scores and performance on selected military tasks or job performance. Studies of M-60 and M-1 tankers have demonstrated a general decline in target hits in gunnery exercises as mental category declined. Automation has been attributed to bolstering performance of the tankers with lower categories in the M-1 (Ref. 3).

3.4.2 Perceptual Processes

Few articles exist treating perceptual processes as model parameters for computer modeling of soldiers. However, Ritter and Avaraamides (Ref. 4) suggest that "perceptual processes allow another area for individual differences and for short term effects" (pg. 12). They break out 5 visual- perception factors that could affect military field performance:

1. Attention to objects moving in the periphery – notice moving objects faster
2. Size of visual field – notice stationary objects faster
3. Perceptual accuracy – helps to find searched-for objects faster
4. Scan speed – targets found more quickly
5. Interpretation of ambiguous stimuli – may be linked to level of anxiety

However, no statement as to the quantitative effects of these factors has been offered, as is the case with many of the sources cited in this report.

3.4.3 Expertise and Experience

All branches of the US Armed Forces spend inordinate amounts of time and money ensuring that soldiers, sailors, airmen, and marines get adequate training prior to performing combat activities, operating technical equipment, or maintaining technical equipment. How effective the training is and how long it remains effective is anybody's guess. A few members of the Army Science Board visited Sandia in 2001 and gave a briefing that included a graph indicating how long training remained effective. It indicated how quickly things learned in training were forgotten if not practiced or put into regular use. It looked something like Fig. 2. Following the up-slope of learning is a characteristic down-slope of forgetting, which in many cases is nearly as steep as the learning curve.

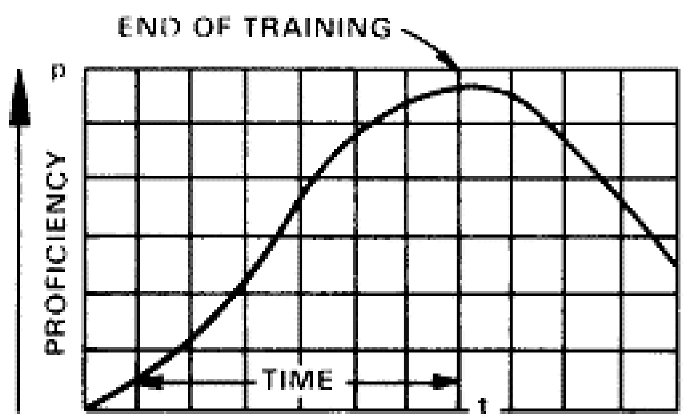

Figure 2. Typical graph showing Army task proficiency during and after training.

Figure 2 demonstrates the worth of having experienced soldiers in the ranks, who have performed their combat tasks many times over a prolonged period. A similarly shaped curve is found in Ref. 8, however, the x-axis represents "Days in Combat," ranging from 0 to 30, with peak performance occurring between 10 and 20 days. The curve is explained as the "new guy factor," where it takes about 10 days for replacement troops to adjust to battle conditions, get oriented, and adapt to the initial high state of fear. After about 3 weeks, the troops begin to get overcome by fatigue and without rest or replacement become totally incapacitated. These data were observed in the Korean Conflict.

3.4.4 Emotions

Eva Hudlicka has written extensively on PSFs associated with individual differences in affective mental states and their impact on cognition (Ref. 17). She concentrates on three emotional states; anxiety, obsessiveness, and depression. She reports that anxiety tends to narrow the focus of attention, can predispose towards the detection of threatening stimuli, and predisposes towards the interpretation of ambiguous stimuli as dangerous. Factors that can increase anxiety in mentally-healthy commanders include isolation of the platoon/company, inadequate time to make decisions, loss of resources or personnel, and vulnerability to the enemy. An anxious leader can overestimate the danger in his situation and make the following command errors: a) commit too many resources or men to a battle, b) choose an overly conservative course of action, or c) interpret signals as approaching enemy and open fire too early.

High obsessiveness, characterized as “checking” behavior, tends to lower confidence in one’s attentional abilities to capture salient features in the environment, narrow conceptual categories, decrease memory for previous actions, slow decision-making speed, and reduce confidence in one’s own capability to distinguish between past events and those that are planned or imagined. Obsessive leaders tend to: a) micromanage, which can lead to inefficiencies and bottlenecks in communication, b) not trust incoming information about the approaching enemy and wait too long before initiating action, c) rehearse more thoroughly, which if time permits, can lead to increased success.

Depression has an impact on memory, in that it can enhance the recall of negative experiences, including negative self-appraisals. It also has the effect of lowering estimates of degree of control in particular inference tasks. Depressed leaders tend to: be less optimistic about outcomes of their decisions, and b) overestimate the likelihood of losing critical equipment, and c) consequently employ overly conservative strategies.

3.4.5 Personality

Personality has been studied since the dawn of the science of psychology, however recent simplifications in modeling personality have taken root in the last 10-15 years. Janis and Mann (Ref. 18) have put forth the view that personality variables, along with situational factors, have important influences on decision-making behavior. According to Pew and Mavor (Ref. 3), more recently, a five- factor theory has become popular (Ref. 19). Janis (Ref. 20) suggests that only three of the five factors are essential in predicting and understanding command behavior; openness, conscientiousness, and neuroticism. Openness has the following characteristics: curious, broad interests, creative, original, imaginative, and untraditional. Conscientiousness is characterized by: being organized, reliable, hard-working, self-disciplined, honest, and clean. Neuroticism embodies worrying, nervousness, emotionality, insecurity, inadequacy, and hypochondria. These and other personality traits can be determined simply by having subjects fill out questionnaires. One study claimed that conscientiousness was the strongest personality predictor of success in the military.

Ritter and Avraamides (Ref. 4) suggest four personality factors that can moderate military performance. Stability moderates other personality factors so that the more stable individuals experience smaller changes in the following personality factors. Individuals with more humor are able to absorb and dissipate their own and others losses and shocks, at least temporarily. Acquiescence, or the willingness to follow orders, varies between individuals and leads groups to perform differently. Eagerness, or the willingness to take on tasks, can lead to problems if someone is too eager, as in a false-courage situation. Unfortunately, the source authors do not offer any quantitative estimates of how these personality traits might impact military-task performance. In our application, it might be valuable to know how these various traits statistically occur in a normal population of military recruits.

3.4.6 Cultural Values

Little research exists on the impact of cultural values and nationality on military performance. However, some research findings from a multinational firm in 67 nation-states suggest that four major dimensions emerge from factor analysis; power distance, uncertainty avoidance, individualism, and masculinity/femininity. Of these, power distance, or the degree to which societies stress status hierarchies, is interpreted as a characteristic of the culture, rather than of the individual-differences dimension. One potential implication of this is that adherence to formal chain of command will be enforced more rigorously in a high power-distance culture (Ref. 3). The authors suggest that one might also infer that the amount of discussion of tactical orders would be minimal. They also cite Brown (Ref. 21), who suggests that risk-taking behavior is much more valued in Western Europe and North American societies than in cultures from other parts of the world.

3.5 Task-Based PSFs

As mentioned briefly above, task-specific PSFs include task history, location of engagement, boredom, stress, threats, casualty level, task type, availability of on-site task training, interface quality, task-related fatigue, battle conditions, and local casualties.

3.5.1 Casualty Level

Unfortunately, no models have been found that account for the fitness of the soldier or level of injury incurred during battle. There must be an unwritten assumption in the military that soldiers are either present or ready for battle or are “casualties” and unavailable. However, for our purposes, there may be circumstances where a soldier is injured during battle, but still capable of performing limited tasks. For example, a limb injury or flesh wound might not prevent a soldier from being able to drive a vehicle, fire a weapon from a protected site, or operate a radio at a command post. If we want a casualty model, we’ll have to build our own. Perhaps a three or four-level model would suffice, where levels one and four would be fully fit and injured enough to be removed from battle, and levels two and three would be unable to communicate and immobilized.

3.5.2 Task and Combat-Related Physical Fatigue

While last year's report (Ref. 2) addressed soldier cognitive fatigue due to lack of rest or sleep in a military engagement, it did not specifically address physical fatigue related to the type of task being performed. Workers, foremen, and researchers have known since prehistoric times that heavy work produces more physical fatigue than light work or rest. Depending on the type of work being performed, the amount of time someone can perform the work before exhaustion is reached varies. Combat tasks vary in the intensity of physical work required. Generic models have been developed for combat, based on data collected in World War II and the Arab-Israeli wars. As reported in Ref. 3, Dupuy modeled fatigue as the degradation of ability to produce casualties at different battle intensities. For a unit in contact with the enemy 80 percent of the time, degradation was 7 percent per day. If the unit was in combat between 50 and 80 percent of the time, degradation was 2 percent per day. Recovery, during non-combat periods, was determined to be 6 percent (improvement) per day. In the early 1980s the Army Research Institute developed PERFECT, a model that determines the performance effectiveness for combat troops in a small unit. It combined a task analysis of military tasks with laboratory research on sleep deprivation, noise, visual acuity, and reasoning abilities (Ref. 3). The model and simulation are described in Ref. 22.

3.5.3 Task History

The previous experience the soldier has had with the task (especially complex tasks such as fault diagnosis and problem solving) can influence attitude toward it and ultimate success/failure and time to perform. After many successful same or similar task completions, the soldier is confident of his/her performance aptitude, and has a higher likelihood of performing the current task to satisfaction. Alternatively, after failure, the soldier will likely be more conservative and less likely to attempt risky actions, even if they are routine for other more successful task performers. The success/failure ratio can moderate the mood, motivation, and decision process of the performer (Ref. 3).

3.5.4 Boredom (Low Stress)

Prolonged exposure to a task that contains no frequent changes (such as watching a radar screen in peacetime) will increase reaction time when an event does occur. This is often called the "vigilance effect" and can occur very soon after a boring vigilance task has begun. Soldiers typically can get very drowsy or fall asleep during tasks such guarding a facility or camp at night. In stress models, this is the point below optimal stress, where the soldier is not stimulated enough to perform tasks adequately (Ref. 3). Similarly, if troops are deployed to a forward position with no combat action, waiting for extended periods (weeks or even a few days) can cause recently trained troops to become bored, lose their fighting edge, and often cause a reduction in discipline.

3.5.5 Location of Engagement

Soldiers will typically be more passionate, motivated, and willing to risk his/her life when the task is to protect his/her own country than to defend another's in a remote location (Ref. 3).

3.5.6 Threat Level and Local Casualties

Factors such as knowledge of the size and skill of the opposing force, its proximity, and the direct or immediate threat to self and associates in battle can promote the fear emotion and anxious or panic behavior in soldiers whose level of training does not overcome these natural responses. Local casualties, especially comrades witnessed getting seriously injured or killed, can increase anxiety or depression to levels of virtual incapacitation.

3.5.7 Quality of Technology Interfaces

In combat, where stress levels can be very high, it is important that user interfaces to technological gear be intuitive, i.e. follow the best ergonomic practices. The reason being that recently learned information will be forgotten in high-stress situations, and the users will revert to what that which they expect or are habitually accustomed. Specialized training, if not reinforced by active use over a significant period of time, can be temporarily forgotten and mistakes can be made. The US military has requirements for and spends millions ensuring that their systems are ergonomically designed and fully integrated before testing and purchasing, however occasionally a poorly designed system interface can sneak through with the promise of intensive training, which can overcome (at least temporarily) an unintuitive interface. Consequently, while ergonomic problems are rare, some systems may incur the effect of decreased performance due to poor human-system interfaces.

3.5.8 Task Difficulty

Whether immersed in combat or working in a safe repair depot, task difficulty almost always affects task performance. Difficult cognitive tasks, such as problem solving, fault diagnosis, navigation through unfamiliar terrain, and strategic decision-making are difficult to perform with high reliability. Simple cognitive tasks, such as counting men, looking for someone, listening to orders, firing a rifle at a target, and performing a routine repair procedure on a vehicle can be performed with much higher reliability. Even simple tasks, if prolonged or if consisting of many steps, can be erroneously or slowly performed, if appropriate job aids (e.g. written procedures, checklists) are not used. Technicians and mechanics, if left to their own preferences, often use a written procedure once in performing a new task, and rely on memory for subsequent performances. However, in high-consequence operations, such as nuclear-weapon assembly, written procedures, checklists, and frequent inspections are enforced to be mandatory in every instance.

3.5.9 On-Site Training

If a soldier needs to perform an unfamiliar task on a piece of gear, any training is better than none. The military is developing technologies to deliver on-site training and mission rehearsal when the need arises. The recency of on-site training can make it very effective, especially if the training is followed immediately by successful task completion.

3.5.10 High Cognitive Workload

Cognitive workload is a difficult parameter to define. It is generally understood as a psychological stress due to inability to cope with a difficult task or an overload of information. It is based on attentional and cognitive-capacity limitations, which limit the success and speed of performing difficult or multiple tasks. Some define workload as the portion of the operator's limited capacity actually required to perform a particular task. Others define it as the difference between the capacities of the information processing system that are required for task performance and the capacity available for other tasks. The result of high cognitive workload is usually a detrimental effect of performance on any of the tasks and limiting the potential of adding new tasks or taking in new information. Workload can also be affected by other PSFs, such as environmental factors, knowledge, experience, intelligence, skills, etc (Ref. 23). Lundin (Ref. 24) found differences in how tankers took firing positions, detected targets, target prioritization, depending on high and low cognitive workload. In each case the performance was superior under low workload conditions. For example, in selecting firing positions, tankers selected the best protection, field-of-fire, and distance under low workload, and simply took the nearest position under high workload. In detecting targets, it took 6 seconds from detection until the gunner had it in-sight under low workload, and 10 seconds under high workload. Similar time differences pertained to detecting new targets after hitting a target. Also, under high workload, crews fired upon the closest target, while under low workload, the most dangerous target was selected.

3.5.11 Stress Level

Stress is the physiological response elicited by environmental stressors, perceived threats, social pressures, task difficulty, limited task completion time, lack of personal confidence, and many other factors. When optimal, stress can be a positive motivator for good performance. However, when it reaches levels higher than optimal, performance is almost universally degraded. Although soldiers are trained to perform under high levels of stressors, actual combat conditions can push even very good soldiers into high stress levels, where performance is degraded. At extremely high levels, as can be experienced in intense combat, stress can virtually incapacitate an individual from performing required tasks and incurring casualties among the enemy. Some modelers have such high regard for stress as a PSF they have used it as a multiplier for coincident PSFs, such as training and task difficulty. Swain and Guttmann, formerly at Sandia Labs (Ref. 23), developed a multiplicative model involving stress level (task load), task complexity, and skill level for nuclear power plant operators (see Table 1.).

Table 1. Behavior-moderation factors for human error probabilities based on stress level, task difficulty and worker skill level (Swain and Guttmann, 1983).

Stress Level	Task Difficulty	Skilled Worker	Novice Worker
Very Low		x2	x2
Optimal	step-by-step	x1	x1
	dynamic	x1	x2
Moderately High	step-by-step	x2	x4
	dynamic	x5	x10
Extremely High	step-by-step	x5	x10
	dynamic	use 0.25	use 0.50

The factors in the last two columns indicate the multiplier to be applied to the probability of human error of commission in nuclear–power plant operations. This model is one of the few quantitative models extant that addresses interactions of multiple PSFs on task performance.

3.5.12 Missing PSFs

In addition to the PSFs listed above, there appear to be some that the literature completely ignores. One is hunger. The author spoke with a veteran Special Forces soldier recently and he said things get very different when one is in combat when hungry—which is a lot of the time. The author assumes that the reason he cannot find anything written on hunger in combat is because after the invention of vacuum canning meat for the Napoleonic campaigns, Western armies have generally been very well fed, except for campaigns of attrition.

Another apparently missing PSF is wearing a heavy backpack. We know that US Army and Marines soldiers often carry packs, weighing upwards of 80 pounds. We also know that our soldiers are in good shape and can carry the weight, but how does it affect combat performance? Does it slow movement, restrict hand-to-hand combat proficiency, or impact marksmanship?

3.6 Summary of Effects of PSFs on Military Tasks

In the interest of estimating the impact of PSFs on a representative set of military tasks, a matrix was prepared with PSFs along the y-axis (row names) and tasks along the x axis (column names). Representative military tasks were collected from documentation on the anticipated Unit of Action (UA) complement of personnel and platforms, and grouped into the categories: Concept of Operations Situation Understanding, Mobility, Survivability, Operability, and Lethality. These categories relate closely to measures of effectiveness (MOEs) for the SoSAT software. Within categories, representative tasks were chosen from the 10 most numerously represented military occupational specialties (MOS) classifications for Future Combat System (FCS) Brigade Combat Team (BCT).

These MOS classifications represent approximately 45 percent of the UA force of 3285 men and women. The matrix was first filled out by the author using information found in the preceding sections and four assumed levels of detrimental effects; no effect, mildly detrimental, moderately detrimental, and severely detrimental. The four levels were graphically depicted in the matrix using a background color and three different hue variants of the color red, as shown in Figure 3. Later, two subject-matter experts (SMEs) with military experience filled out a blank matrix. The complete matrix data can be found in Appendix A.

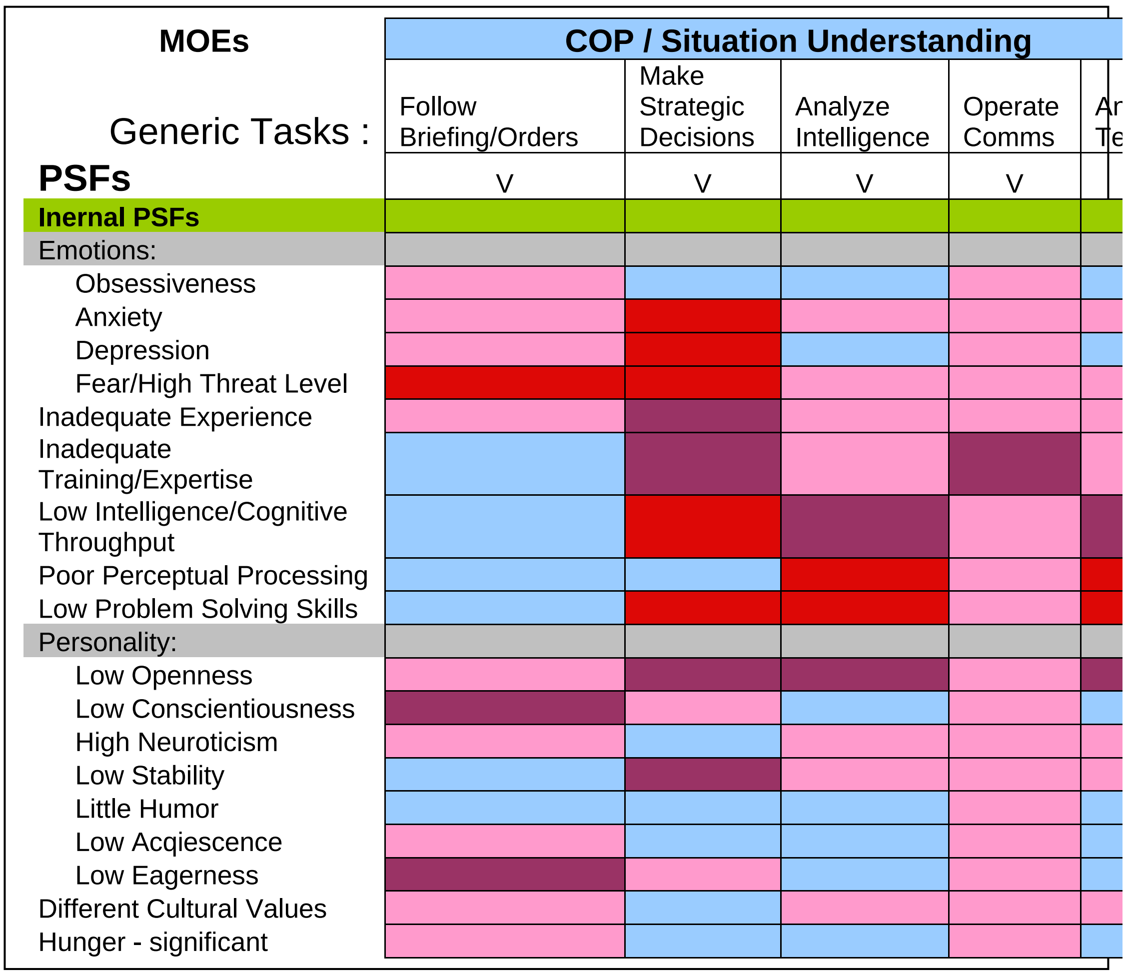

Figure 3. Partial Matrix of PSFs and Typical Military Tasks. The 47 PSFs are listed down the left-hand side of the matrix (along with categorical headers in gray) representing rows. The 22 military tasks are listed across the top of the columns. For every combination of PSF and task, one of four colors is inserted, representing the severity of effect. The default background color (in this case light blue) represents no effect, rose represents small negative effect, magenta represents moderate negative effect, and bright red represents severe negative effect.

3.6.1 SME Matrix Results

Two SMEs with military experience filled out the whole 1034-cell matrix (see Appendix A). Brief biographical sketches of the two SMEs follow:

> Major "X" is an officer in the U.S. Army Reserves. His active duty experience includes command of a Special Forces Detachment (A-Team) and 5 deployments to the CENTCOM AOR. He is a graduate of the United States Military Academy.
>
> Sergeant "Y" is a former United States Marine and veteran of the current conflict in Iraq. He served in Iraq from the beginning of the war until September of 2003. In his role as an Arabic translator, Sgt. Y worked with many different units all over the battlefield, but his primary mission was conducting battlefield interrogations on the front line.

Neither SME discussed the literature findings on PSFs with anyone from the project. Each was e-mailed the instructions shown in Appendix B. On the first attempt, Major X filled out the matrix with mostly bright red. After explaining to him that we needed variation in the responses so that we could judge which ones were the most important, he resubmitted the matrix with a wider range of responses.

Each SME's completed matrix was analyzed for PSFs that had the most cumulative negative impact on the 22 military tasks. A score for each PSF was calculated by adding the number of rose responses with two times the number of magenta responses and three times the number of bright red responses. The highest ten scoring PSFs for each SME were then listed in descending order, along with the results from the literature (see Figure 4.) While almost perfect correspondence was achieved between the literature and Sergeant Y (9 out of 10), less correspondence was achieved with Major X (6 out of 10).

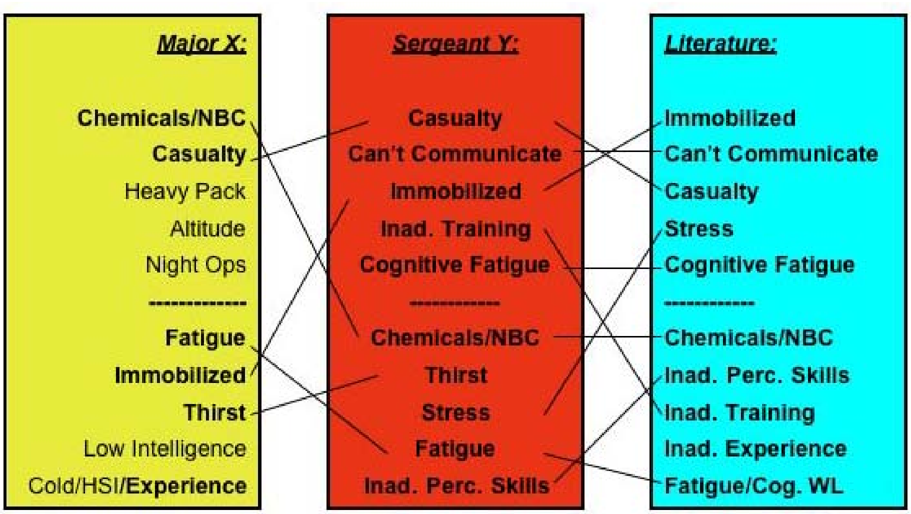

Figure 4. Top ten PSFs from SME's data and the literature.

4. Interactions of Multiple PSFs

Rarely does a soldier going into combat experience only one PSF at any given time, but experiences numerous PSFs, especially as the combat continues for a significant time period and fatigue and other factors accumulate. Given this rational appraisal of combat situations, we would expect previous modeling efforts to have taken the many possible interactions of PSFs into account in their modeling. By interactions, we mean the combinatorial effects of two or more factors, which is more (or less) than a summation or additive relationship. However, we find very few instances where interactions are acknowledged, much less modeled. Exceptions are noted in the following section.

4.1 Interactions Discussed in the Literature

Van Nostrand (Ref. 8) developed a multiplicative model of three factors, noting that an additive model would have been less conservative. The factors modeled were from three different types—internal, external, and task-related: average soldier mental category (IIIA), participation in sustained operations of 24 hours, and experiencing high heat in combat for 7 hours. The model compared combinations of the three factors with a base case of 100 % of expected hits. The mental category reduced performance to 85%; the sustained operations reduced it to approximately 65%; and the 7 hours of high heat reduced performance to about 10%. All three combined reduced performance to about 5%. No data were acquired to validate the model, and no additional combinations of factors were modeled in the report.

Ritter and Avaraamides (Ref. 4) provide a table in which specific combinations of PSFs are combined with brief discussions of effects and examples. For example, the combined effects of noise (external factor) and anxiety (internal or task-related) are the following:

> "Effects: Increase selectivity in attention especially in dual tasks. Anxiety produces no improvement in main-task performance while noise does. Both noise and anxiety reduce secondary-task performance. Anxiety may also lead to overestimating potential danger.
>
> Example: A soldier has two tasks; to track the enemy position and monitor the layout of his teammates. A highly anxious leader…will perform better in the main task (i.e. monitor enemy) but worse on monitoring his fellow soldiers. In the presence of noise in the environment his performance on both primary and secondary task[s] will decrease. An anxious leader may also overestimate potential danger and commit to a task an unnecessary number of troops and resources." Pg. 18.

Perhaps the reason so little is written on modeling interactions of PSFs is that the factorial experiments have not been conducted, and when it comes to combat behaviors, it would be impossible due to ethical reasons. Silverman, as quoted in Ref. 22 states that:

> "As soon as one tries to integrate across PSFs and synthesize the Integrated Stress…, one rapidly departs from grounded theories and enters into the realm of informed opinion". Pg. 12

Nevertheless, as Swain and Guttmann (Ref. 23) have shown, informed opinion is what researchers can provide prior to the definitive experiment being performed. Swain has told the author that experience in reviewing nuclear-power plant accidents and developing human-reliability analyses of same led to the interactive stress model and that subsequent field operations data have informally validated the model that he and Guttmann developed.

4.2 Other Likely Interactions

As another means of categorizing PSFs, several researchers in the field have made the distinction between PSFs that change during the task (or, in our case, during combat), and those that do not. Looking at the three general categories we've established, most of the internal PSFs don't really change during any given battle, which might take anywhere from a few days to several weeks. The exception is emotion, which can change quickly as the result of a turning point in a battle, getting wounded, or seeing colleagues get maimed or killed.

By and large, however, external and task-based PSFs can change quickly in a given battle, by their very nature and they can combine in non-additive ways. For example, if it were to begin raining in a forest or rocky terrain, military operations would not be as severely affected as if it happened in an area with loose dirt and no vegetation, which could create serious mud underfoot, thus hampering troops movement and vehicle operations. This kind of interaction exacerbates the negative effects of the two factors taken in isolation, making the resultant combination worse than their simple additive effects. Because they affect performance in the same direction, we'll call them *positive* interactions for this discussion. Similar positive interactions can be presumed for high winds and sandy desert terrain, cold and vibration, toxic substances and night operations, noise and fatigue, inexperience and fatigue, etc. Some PSFs can combine in ways that tend to cancel or negate each other's effects. We can call these *negative* interactions for this discussion. Although these tend to be rare, examples would be anxiety and fatigue, rain and toxic gases, and wind and excessive heat. Because we are modeling PSFs that reduce soldier performance from an assumed 100 percent, we will not attempt to model the negative interactions in this report.

Attempts will be made in the next fiscal year to collect data from our subject-matter experts concerning the effects of combinations of PSFs on representative military tasks.

References

1. E.F. Howard and H.G. Lipsett, *Naval Sea Systems Operational Availability Quantification and Enhancement*, NUSC Technical Report, Naval Underwater Systems Command, 1976.

2. C.R. Lawton, D.P. Miller, and J.E. Campbell, *Human Performance Modeling for System of Systems Analytics: Soldier Fatigue (U)*, SAND2005-6569. Sandia National Laboratories, Albuquerque, NM, October 2005.

3. R.W. Pew and A.S. Mavor, *Modeling Human and Organizational Behavior: Application to Military Simulations*. National Academy Press, Washington, 1998.

4. F.E. Ritter and M.N. Avraamides, *Steps Towards Including Behavior Moderators in Human Performance Models in Synthetic Environments*, ACS 2000-1. Pennsylvania State University School of Informational Sciences and Technology, State College, PA, 2000.

5. S.H. Dover, *The Characterization and Prediction of Soldier Performance During Routine Service and in Combat*, AD A399051. US Army Research Institute for the Behavioral and Social Sciences, 2002.

6. S.H. King, J.T. Campbell, C.D. Johnson, W.A. Klieger, and D.W. Yaukey, *Studies of the Performance of Officers in Combat; I Relationship of West Point Measures to Later Combat Effectiveness,* TR969. US Army Personnel Research and Procedures Branch, 1952.

7. K.C. Parsons, Ergonomics Assessment of Thermal Environments, in *Evaluation of Human Work*, J.R. Wilson and E.N. Corlett, ed. Taylor and Francis, Bristol, PA, 1996.

8. S.J. Van Nostrand, *Model Effectiveness as a Function of Personnel,* CAA-SR-86-34. US Army Concepts Analysis Agency, Bethesda, MD, 1986.

9. N.M. Findikyan, M.J. Duke, and S.B. Snells, *Stress Reviews: I. Thermal Stress – Cold*, TR-8. Office of Naval Research, Texas Christian University, 1966.

10. G. Salvendy, *Handbook of Human Factors*. Wiley Interscience, New York, 663-664, 1987.

11. S.G. Hill, and R.A. Tauson, *Soldier Performance Issues in C2 "On the Move."* Human Research and Engineering Directorate, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, 2001.

12. P.S. Cowings, W.B. Toscano, C. DeRoshia, and R.A. Tauson, *Effects of Command and Control Vehicle (C2V) Operational Environment on Soldier Health and Performance,* ARL-MR-468. U.S. Army Research Laboratory Human Research and Engineering Directorate, Aberdeen Proving Ground, MD, November 1999.

13. N. Sherwood, N. and M. Griffin, Evidence of impaired learning during whole-body vibration, in *Journal of Sound and Vibration* vol 152, No. 2, 219-225, 1992.

14. S. Schipani, R. Bruno, M. Lattin, B. King, and D. Patton, *Quantification of Cognitive Process Degradation While Mobile, Attributable to the Environmental Stressors Endurance, Vibration, and Noise,* ARL-TR-1603. US Army Research Laboratory, Aberdeen Proving Ground MD, 1998.

15. Information found at: http://www.cavalrypilot.com/fm1-100/ch3.htm#s2p1

16. J.H. Banks, *Lessons from the NTC: Common Battalion Task Force Training Needs (draft)*. US Army Research Institute for the Behavioral and Social Sciences, Presidio of Monterey Field Unit, 1985.

17. E. Hudlicka, *Modeling Behavior Moderators in Military Performance Models*, Technical Report 9716. Psychometrix Associates, Lincoln, MA, 1997.

18. I.L. Janis, and L. Mann, *Decision making: A psychological analysis of conflict, choice, and commitment*. The Free Press, New York, NY, 1977.

19. R. Hogan and J. Hogan, *Hogan Personality Inventory Manual*. Hogan Assessment Systems, Tulsa, OK, 1992.

20. I.L Janis, *Critical Decisions: Leadership in Policymaking and Crisis Management.* The Free Press, New York, NY, 1989.

21. J. Brown, *Environmental threats: Social Science Approaches to Public Risk Assessment*. Bellhaven, London, England, 1989.

22. A.I Siegel, J.J. Wolf, and A.M. Shorn, *Human Performance in Continuous Operations: Description of a Simulation Model and User's Manual for Evaluation of Performance Degradation,* ARI Technical Report 505, AD-A101950. US Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA, 1981.

23. A.S. Swain, and H.E. Guttmann, *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Operations (U),* NUREG/CR-1278. Sandia National Laboratories, Albuquerque, NM, August 1983.

24. M. Lundin, *Simulating the Effects of Mental Workload on Tactical and Operational Performance in Tankcrew*, TR No. 2004-09-23. Linkoping University, 2004.

Appendix A (approx. 12 pages)

Major X Matrix.xls

Sergeant Y Matrix.xls

Literature Matrix.xls

Appendix B

Matrix form-filling instructions, which were e-mailed to the SMEs:

Here is the matrix we spoke of on the phone. It has behavior moderators, or Performance Shaping Factors (PSFs) down the left-hand column, and typical military tasks across the top. The tasks are categorized into Measures of Effectiveness (MOEs), but don't pay any attention to that. Some of the PSF lines (rows) are gray, indicating they are headers for what follows immediately below (indented), and don't need to be filled out. The three main categories of PSF are shown in bright colors and don't need to be filled out either.

We added an example of how to fill it out in the first column. The repeated columns listing the PSFs are there for visual convenience only—so you don't have to scroll across the screen. The existing background colors differ only to visually separate the MOEs across the top of the columns.

Please leave the cells the original color if the PSF has no effect. We assume a soldier is at 100 percent with no negative effects from the PSFs. Select Rose fill for mild negative effect (10-30 percent decrement), plum for moderate negative effect (35-65 percent decrement), and red for severe negative effect (70-100 percent decrement).

We found when filling out the matrix, that there are some tasks that are similar enough (at least initially in our minds) to cut and paste from an already filled out column. However, these were very few, and when we went down the pasted column, not all of the colors were appropriate for the new task. So if you use this technique, please review all of the cells for the new task and make the changes necessary. We found that no columns ended up matching any other columns exactly.

We purposely did not define the tasks in any detail so that you could use your experience to recall what they consisted of. If you are not familiar with the task, give it your best guess. All of the tasks were extracted from MOS documentation for the following:

Infantrymen 11B10, 11B20, 11B30
Medical specialist 68W10
Petrol and Water Supply 92F10
Cavalry Scout 19D10
Canon Crew 13B10
Armor Crew 19K10, 19K20

We assumed that a commanding officer was making the "strategic decisions."

Results may vary, but when we filled out the matrix using the literature as our guide, it took several 2-hour sessions to get through it all. We recommend you don't try to do it all in one sitting, as cognitive fatigue may throw off your consistency. Feel free to go back and change your previous answers, if it will help you to be more consistent.

Distribution

MS 1002 Stephen Roehrig, 06300
MS 1005 Russell Skocypec, 06340
MS 1011 Robert Cranwell, 06342
MS 1011 Dennis Anderson, 06342
MS 1011 Dwight Miller, 06342
MS 1011 Craig Lawton, 06342
MS 1188 Chris Forsythe, 06341
MS 1188 John Wagner, 06341
MS 0748 Chris Atcitty, 06761
MS 1004 Patrick Xavier, 06344
MS 1188 Eric Parker, 06341
MS 1188 Ann Speed, 06341
MS 1188 Kevin Dixon, 06341
MS 1004 Brian Rigdon, 06344
MS 1188 Michael L. Bernard, 06341
MS 9018 Central Technical Files, 8944 (2)
MS 0899 Technical Library, 4536 (2)