4

MILITARY OPERATIONS RESEARCH SOCIETY

MORS
MILITARY OPERATIONS RESEARCH SOCIETY

MINI-SYMPOSIUM PROCEEDINGS

"Human Behavior and Performance as Essential Ingredients in Realistic Modeling of Combat - MORIMOC II"

VOLUME 1

22-24 February 1989
Center for Naval Analyses
Alexandria, Virginia

Proponents: Deputy Under Secretary of the
Army Operations Research
and
Assistant Chief of Staff, Studies and Analyses
HQ US Air Force

90

This Military Operations Research Society minisymposium proceedings faithfully summarizes the findings of a three day meeting of experts, users, and parties interested in the subject area. While it is not generally intended to be a comprehensive treatise on the subject, it does reflect the major concerns, insights, thoughts, and directions of the authors and discussants at the time of the workshop.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	**1b. RESTRICTIVE MARKINGS** None
2a. SECURITY CLASSIFICATION AUTHORITY N/A	**3. DISTRIBUTION/AVAILABILITY OF REPORT** Unlimited, approved for public release
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) MORIMOC II Report	**5. MONITORING ORGANIZATION REPORT NUMBER(S)** None
6a. NAME OF PERFORMING ORGANIZATION Military Operations Research Society, Inc. **6b. OFFICE SYMBOL** *(If applicable)* --	**7a. NAME OF MONITORING ORGANIZATION** Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) 101 S. Whiting Street Alexandria, VA 22304	**7b. ADDRESS (City, State, and ZIP Code)** Arlington, VA 22217
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research **8b. OFFICE SYMBOL** *(If applicable)* --	**9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER** N00014-89-C-0040
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217	**10. SOURCE OF FUNDING NUMBERS**

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO
--	--	--	--

11. TITLE *(Include Security Classification)*
Mini-symposium Proceedings "Human Behavior and Performance as Essential Ingredients in Realistic Modeling of Combat - MORIMOC II"

12. PERSONAL AUTHOR(S) Stephen A. Murtaugh, Editor

13a. TYPE OF REPORT	**13b. TIME COVERED**	**14. DATE OF REPORT** *(Year, Month, Day)*	**15. PAGE COUNT**
Proceedings	FROM 22Feb89 TO 24Feb89	89 May	806

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
12	04	
15	06	

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Operations Research, Human Factors, Human Performance, Combat Data, Combat Modeling, Battle Simulation, Weapons Effectiveness, Micro-SAINT Modeling, (KR)

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This is the Proceedings of the Military Operations Research Society's mini-symposium dealing with the problem of achieving more operational realism in the modeling of combat from the standpoint of accounting for human performance in the battle environment. The meeting was held at the Center for Naval Analyses, Alexandria, Virginia, on 22, 23, 24 February 1989. Included are the 34 contributed papers presented in the five sessions: Human Factors in Decision Issues, Human Performance Models and Applications, Predicting Human Performance/Availability in Combat Environments, Combat as a Data Source, and Representing Human Performance in Combat Models and Simulations. Also provided are the remarks by the session chairs and by the discussants of the papers, plus the several invited presentations. The papers include several by participants from the NATO SHAPE Technical Centre, Great Britain, and West Germany.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	**21. ABSTRACT SECURITY CLASSIFICATION** Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Richard I. Wiles	**22b. TELEPHONE** *(Include Area Code)* (703) 751-2507 **22c. OFFICE SYMBOL** N/A

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

TABLE OF CONTENTS

VOLUME I

SESSION I: HUMAN FACTORS IN DECISION ISSUES

SESSION II: HUMAN PERFORMANCE MODELS AND APPLICATIONS

TABLE OF CONTENTS (continued)

SESSION III A: PREDICTING HUMAN PERFORMANCE/AVAILABILITY IN COMBAT ENVIRONMENTS

---

*Paper not released for publication.

TABLE OF CONTENTS (continued)

VOLUME II

SESSION IIIB: CONTINUATION OF PREDICTING HUMAN PERFORMANCE/AVAILABILITY IN COMBAT ENVIRONMENTS

SESSION IV: COMBAT AS A DATA SOURCE

TABLE OF CONTENTS (continued)

SESSION V: REPRESENTING HUMAN PERFORMANCE IN COMBAT MODELS AND SIMULATORS

1) Methods for quantifying potential performance degradation of individuals and weapons crews in combat environments, based on data from combat, weapons tests, and real-time simulations.

2) Techniques for estimating environmental and workload effects on human performance.

3) Availability and utility of combat data for modeling and analysis inputs.

4) Approaches for including human factors in combat models.

5) Results of recent analyses of combat effectiveness.

A unique feature of this MORS symposium was the participation by representatives of several of our NATO allies (SHAPE Technical Centre, Canada, France, Great Britain, The Netherlands, and West Germany). Significant contributions to the information base cited above, including interesting perspectives on approaches to the MORIMOC II problem, were presented by several of these participants. These include the unique analysis of World War II combat data by David Rowland of Great Britain, the noteworthy study of the effects of soldier fatigue in battle by Htm. Werner Siemon and Helmut Wollschlager of West Germany, and the several penetrating and thoughtful discussions by L. Ron Speight of SHAPE Technical Center, all included in this Proceedings.

From the cumulative information provided by the various papers and discussions, a candidate set of problems have been identified as possible grist for the millstones of the MORIMOC III Workshop - the specific problems will be selected and refined from this set during the MORIMOC II General Session at the 57th MORS Symposium in June 1989.

Finally, an unspecified benefit of the symposium should be noted. This truly became a multi-disciplinary meeting, for unlike the usual MORS activities at which the great majority of the attendees are operations research, social science, and systems analyst professionals, this meeting also attracted attendees and speakers from the human factors, human engineering, psychology, and behavioral science disciplines. The diverse backgrounds but common interest of all these participants were recognized early on, so that time was allocated for mixing and for small group discussions, both during the meeting days and also by a planned social mixer the first evening and a dinner program the second evening. In addition, every effort was made to provide time for questioning the authors of the papers and to select discussants for the papers with an eye to cross-fertilization among the technical disciplines. The extent to which this was accomplished can only be gauged by the enthusiasm shown by the attendees, the sustained attendance throughout the full three days of the program, and the responsiveness of all of the authors and discussants in early submittal of their papers for this Proceedings.

Stephen A. Murtaugh

PREFACE

This Proceedings documents the hard outputs of this landmark multi-disciplinary MORS mini-symposium, which had as its proponents the Deputy Under Secretary of the Army for Operations Research and the Assistant Chief of Staff, Studies and Analyses, Headquarters U.S. Air Force. MORIMOC II provided a forum for description and discussion of efforts by the operations research, human factors, and behavioral sciences communities in identifying and understanding (1) the impact of human performance on combat and (2) techniques for representing such impacts in weapons effectiveness and combat models.

The performance of humans in battle environments, functioning as individuals, crews, or in units, and the ability to model or account for the influence of humans on combat operations are subjects of long term interest to segments of the above-named communities, especially the combat historians/analysts and the human engineers. Over the past decade, many DoD-sponsored activities, including the MORS-sponsored MORIMOC I Workshop in February 1986, examined the shortcomings of combat models and data for such applications as war planning, training, procurement, and logistics decisions. A unanimous finding of all these activities was the lack of accounting for the effects on battle outcome of human actions and performance in the combat environments at all levels of the opposing forces. Yet the human element has overriding importance in all battle operations. The weapons effectiveness and combat analysis/modeling efforts must account for the capabilities and the degradations of the combatants' performance, and must reflect these effects at the individual, crew, and unit levels, as appropriate.

This MORIMOC II mini-symposium was a first step for MORS in creating and hosting a multi-disciplinary program directed to development of the understanding needed to answer important questions about representing human performance (capabilities, limitations) in combat models used to support military decision making: to what extent does human performance affect model outputs, how much of these effects must be accounted for, and how human performance can be included in modeling of weapons effectiveness and combat. In particular, the objectives of this mini-symposium were to:

1) Develop an information base on the present status of modeling of human performance in combat and the effects on the conduct and outcome of battle, and

2) Provide guidance and direction to the structuring of the work areas for the planned sequel, a MORS Workshop (termed MORIMOC III) on this subject, scheduled for March, 1990.

The contents of this Proceedings serves to testify, by the papers presented and the discussants remarks, that both objectives were fully satisfied and some additional benefits (discussed below) also were realized. From the standpoint of the first objective, we uncovered (and provide here) substantial evidence of an extensive information base in the areas of:

PROGRAM STAFF

PROGRAM MANAGEMENT:

STEPHEN A. MURTAUGH, CALSPAN CORPORATION, PROGRAM CHAIRMAN

SALLY J. VAN NOSTRAND, US ARMY CONCEPTS ANALYSIS AGENCY, PROGRAM CO-CHAIR

SESSION CHAIRS:

EUGENE P. VISCO, OFFICE OF DEPUTY UNDERSECRETARY OF THE ARMY FOR OPERATIONS RESEARCH
MICHAEL STRUB, CHIEF, US ARMY RESEARCH INSTITUTE FIELD UNIT, FORT BLISS, TEXAS
SALLY J. VAN NOSTRAND, US ARMY CONCEPTS ANALYSIS AGENCY
CAPT. WAYNE P. HUGHES, US NAVY (RETIRED), US NAVAL POSTGRADUATE SCHOOL
LTC VERNON M. BETTENCOURT, DIRECTOR, TRADOC ANALYSIS COMMAND - MONTEREY

MORS OFFICE:

RICHARD I. WILES, EXECUTIVE DIRECTOR/PUBLISHER
NATALIE S. ADDISON, ADMINISTRATOR/MEETING PLANNER
CYNTHIA L. LAFRENIERE, ADMINISTRATIVE ASSISTANT/SECRETARY
JOHN K. WALKER, PHALANX EDITOR/PHOTOGRAPHER

ACKNOWLEDGMENTS

The draft of this Proceedings was reviewed by the Publications Committee of MORS. In particular, Mr. Clayton Thomas, Chief Scientist, Studies and Analysis, Headquarters U.S. Air Force, read both volumes of this Proceedings, identifying a multiplicity of errors, including various mistakes in analyses and text, several of which were serious. Clayton has been a friend of this editor for many years. The time and dedication he voluntarily devoted to the review and critique of these volumes is a touching and much appreciated demonstration of true friendship. By his efforts, the quality of these proceedings are significantly enhanced.

Within Calspan Corporation, Joan E. Lus skillfully performed the typing of much of the text of the Proceedings plus necessary revisions to both the text and graphics. Alice E. Castren provided secretarial services to support the Symposium. Their efforts through several drafts of text, mailing lists, programs, etc., are appreciated very much. Finally, the Publications Services group provided support needed to prepare the drafts for publication.

MILITARY OPERATIONS RESEARCH SOCIETY WELCOME

Clayton J. Thomas
Chief Scientist, Air Force Studies & Analyses
MORS Air Force Sponsor's Representative

It is a privilege and a pleasure to welcome you today on behalf of MORS--the Military Operations Research Society. Some of you are old friends of MORS, but many here are new to MORS, even a few from other lands. So let me tell you a little about MORS--how it is like a typical professional society, and how it is different.

In many ways MORS is what you would expect. It seeks to enhance the quality and effectiveness of military operations research. It provides media for professional exchange and peer criticism, such as publications and meetings like this mini-symposium. MORS encourages high standards for professional methodology and applications, and high standards of individual excellence and ethics for all practitioners. So far, MORS sounds very typical.

But in other ways, MORS is not at all typical. It does not have several thousand dues-paying members. As <u>actual</u> members, MORS has its board of 28 directors, each of whom is elected to a four-year term. Many friends of MORS who have been to at least one of its symposia think of themselves as members. There are probably several thousand such friends, and they include analysts, operators and managers. Some are in government--uniformed or civilian--and some are in industry or "think-tanks" or academia.

Several hundred of these friends of MORS are the special volunteers involved in the hard work that makes symposia and other MORS meetings actually happen. MORS has a small office of <u>superb</u> <u>and</u> <u>dedicated</u> individuals--the executive director, administrative assistant, and secretary, all of whom you met when you registered--but MORS volunteers are also essential to the success of meetings and publications.

MORS collects no dues, so where does its money come from? The two main sources are registration fees for its meetings--like the fee you paid for MORIMOC II--, and contractual support from its five sponsors. The original sponsor was the Navy. Its Office of Naval Research started MORS just a bit over 30 years ago, and the Navy still manages the MORS contract. For over 20 years, however, the Army and Air Force have joined the Navy in sponsoring MORS, and more recently the Joint Chiefs of Staff and the Office of the Secretary of Defense have also become sponsors.

In its first 25 years, MORS' meetings were mostly two large symposia per year--each symposium having several hundred participants and as many as 30 working groups. To those who have not yet seen a big MORS symposium first hand, I describe it as something like a 30 ring circus. There's more going on than any single individual can watch.

Half a dozen years ago, MORS branched out to new types of meetings. There is still one large symposium each summer, with typically between 500 and 1000 participants. But during the rest of the year there are several smaller meetings. Most of these are either workshops, typically with between 20 and 60 participants, or mini-symposia, which may have as many as 200 participants. So far, MORIMOC II appears to have slightly over 110 participants.

As a mini-symposium, MORIMOC II is, of course, one of the new type meetings, and we hope that it will be one of the best. We have come because we recognize the importance of representing, in some way, the effects of human behavior and performance in the modeling of combat, if that modeling is to be realistic. I suspect that shortly we shall hear more about that importance from Gene Visco and Steve Murtaugh.

I first began to appreciate the generality of the concern with the "human factors problem" three years ago this week, at the MORIMOC I workshop. That workshop looked at how to get more operational realism in the modeling of combat. Each of three working groups looked at the goal of realism from its own perspective--operational, mathematical, and physical-engineering. I had expected the operational group to mention human factors, but I had thought that the mathematical group would probably confine itself to subjects like misuse of expected values, the assumption of statistical independence, etc., and that the physical-engineering group would be largely concerned with measuring quantitities like radar cross-section and silo hardness. The striking result, however, was that each of the three groups had as one of its major concerns the need to do better in representing the effects of human behavior and performance.

Steve Murtaugh and Sally Van Nostrand were at MORIMOC I, and they actually did something about its insight--they planned and spearheaded this meeting. Walt Hollis and Gene Visco and their office supplied to MORIMOC II the impetus of active sponsorship. Those of you who have come to give and discuss papers have provided the final essential ingredient. I salute you all, and look forward to three idea-packed days.

Thanks to all of you for coming.

ARMY SPONSOR'S WELCOME

Eugene P. Visco
Director, Study Program Management Agency
Office of the Deputy Under Secretary of the Army
(Operations Research)

This delightful task normally falls to Mr. Walter W. Hollis, DUSA(OR). I welcome you, not as the corpus Visco but rather, as the spirit Hollis. Other demands were made on his time and I am pleased to represent him as well as myself. We both have strong feelings about the importance of the theme of this mini-symposium. We are equally pleased to see the growing and continuing interest MORS and the military operations analysis community is showing in the issue of human behavior in models of combat, as well as in other forms of analysis.

It is particularly fitting and proper that the Army take a point position about the issue. The other senior service sees itself comprised essentially of ships; we speak of the "600 ship Navy." The junior service sees itself made up of wings and squadrons--and, sometimes, silos or railroad cars. The Army has always seen itself principally as people. Obviously, we form up into units for combat and other activities; regiments, brigades, divisions, corps, armies, and army groups are important elements. But first and foremost the Army is soldiers. Modern events, such as the Vincennes incident, suggests that all of us need to pay much more attention to the human element.

The theme is important. MORS is to be congratulated on leading the way. All the participants, presenters, discussants, and chairs are to be congratulated for recognizing the increasing importance of the theme. Let us proceed.

INTRODUCTION TO
MORIMOC II: MORE OPERATIONAL REALISM IN MODELING OF COMBAT

Stephen A. Murtaugh
Symposium Chairman

We are gathered here as an interdisciplinary group, to consider what has generally been assessed to be the unquantifiable aspects of war - the contributions of the human being in the so-called "high technology" combat environment. We are interested in behavior and performance of the human in this stressful environment from the standpoints of the individual combatant, the crew of a weapon (e.g., helicopter, tank, artillery piece) and the unit (platoon, company, etc.) - and how that behavior and performance can and does affect the combat capability of the weapons operated by the humans in various types and levels of combat.

The need has been recognized for some years to develop the understanding we are seeking and to apply this new-found knowledge in the analysis of military operations so as to provide substantiated inputs to the planning, preparation, conduct, and evaluation of combat.

Each of us here has not only recognized this need, but has acted upon it, each according to his or her own talents and capabilities. Many of you have offered papers to present at this symposium, papers which provide hard evidence of your resolve, along with that of your employers and your contract sponsors, to develop understanding of the impact of human behavior and performance on combat effectiveness.

As a result, we have been able to put together an agenda for these next three days which is based on good, solid work performed in such areas as experimental research, battle measurements, historical data bases, model and simulation development and applications, and realistic combat training. All of this has the focus of developing better capabilities for understanding and modeling the role of the human in combat so as to meet the needs of the U.S. and NATO countries' military planners.

Of course, what is immediately apparent is that the problem we have defined requires the efforts of a variety of scientific and engineering disciplines. The field of operations research is an inter-disciplinary science, coupling social sciences and operational analysis to provide a powerful tool for quantifying and explaining many of the functional phenomena in which mankind engages. However, the range of professional capabilities which you folks possess as a group, and which are needed to develop the technology we seek, is much broader and is truly impressive. We on the program committee, plus the MORS Board of Directors and the MORS DOD sponsors, are very pleased with your participation in this symposium - and that we have this opportunity to marshall our joint efforts in an interdisciplinary approach to this timely problem.

I do not want to suggest that we expect that all the problems of accounting for and integrating human performance into combat models will be solved in this one meeting. We have been planning and preparing for this symposium for over two years - and during that time received much good guidance and encouragement from various sources, including several people who are on the program: Clayton Thomas and Gene Visco, whom you have already heard from: Sally Van Nostrand and Wayne Hughes, Session Chairs on Thursday and Friday, Ron Speight from STC, who will be speaking to us tomorrow night after dinner, and Ed VanDiver, Director of the Army's Concepts Analysis Agency. This meeting is but one step in a planned sequence to attack this problem.

After this meeting, we will have a general session at the 57th MORS, to be held in early June at Ft. Leavenworth. At this meeting, the session chairmen and I will present a summary of what was accomplished in this program and a listing of the problem areas which have been identified. Also, we shall discuss the formation of a new working group dealing with human factors and quantifying the human. The working group will meet at the annual MORS symposia - to deal with the problems we are concerned with here. Another step is a follow-on MORS-sponsored workshop planned for next year. This workshop will be composed of those of you here who wish to work on and contribute to the approach and solution of the problems we jointly identify as an output of this symposium.

Still another facet of our attack came about this way. Early in our planning for this series of endeavors, Frank Tapparo, of the Office of Assistant Secretary of Defense, Program Analysis and Evaluation, learned of our efforts and contacted me to inform us that the NATO SHAPE Technical Centre (STC) at The Hague, The Netherlands, was planning to start work on a research study project on the representation of human factors in wargames and battle models. He further introduced me to the head of the STC Operations Research Division, Dr. L. Ron Speight, who is participating in our program, and who is directly involved in establishing this NATO workshop. Subsequently, Ron visited me and we devised a plan by which the MORS and NATO symposia and workshops will be complementary and feeding into one another, so that jointly we can achieve more than if they were pursued separately. This is how it has come about that, for the first time, NATO is participating in a MORS program. Dr. Speight has been influential in generating interest among NATO countries to participate in our meeting. As a result of his efforts and those of others, we have several European papers on our agenda and have representatives from STC, England, Canada, France, The Netherlands, and West Germany here to participate in our discussions.

We shall work toward having NATO participation in the follow-on meetings we plan, and we shall provide such support to the NATO workshop, when it is scheduled, as they may find appropriate to their need.

Now, as to the objectives of this symposium. Once it was determined that the subject was of interest and important, it was deemed a next step should be to determine who was doing what work in the field and of what value might it be? Therefore, the first objective of this symposium is to gather an information base to define the present status of modeling human performance and behavior, and the effects of such performance on the conduct and outcome of combat. This includes the status of human factor data base development, the use of such data in combat models for various applications, and the results achieved as this data is applied.

The response to this first objective was somewhat of a surprise to us involved in the planning - I, for one, had expected 15 or 20 abstracts in response to the call for papers - hoping for enough material to have a 1-1/2 or 2 day symposium - truly a "mini." Instead, we received about 50 abstracts -some in considerable detail which indicated the work described to be near completion. We accepted 35 of the abstracts, but found it would be necessary to go to a 3 day meeting, if we did not want simultaneous sessions. So, with the good graces of our host, CNA, we extended the meeting through this Friday afternoon.

Our second objective has two parts to it - both have to do with providing guidance and direction so that the follow-on MORS workshop will make a worthwhile contribution. First, from the perspective I have at this point in time, I feel this purpose will be served in part by the identity and definition of specific problem areas which might be amenable to attack in a workshop environment. Such problems are:

- How do the particular kinds of "decision issues" for which the combat modeling is being performed, impact the degree to which human factors are "essential ingredients" in combat modeling and support of decision makers? That is, how much realism is required for the various decisions required?

- How to develop a conceptual framework for consideration of the MORIMOC II topic? As Ron Speight has so aptly expressed it in a letter to me: "The field of enquiry is potentially vast, stretching as it does into every field of human endeavor in battle, all facets of operation of the many different weapon systems, and all the different judgments and decisions which men are called upon to make when countries are opposed in war. A year-long symposium of random papers might expect to peck holes in all this here and there (especially as basic data are extremely scarce). This being so, there is a very real risk of any symposium being formless and inconclusive, especially as there is no ready-made theoretical structure into which papers could be fit."

Second, the documentation of the papers presented and the discussion of the papers is an important record of what we have accomplished here. Furthermore, such a printed proceedings, issued at an early date, will serve as a useful handbook to those who participate in our follow-on workshop and to our NATO friends in the conduct of their workshop. Therefore, we attach great importance to the early revision and submission by the authors of all the papers presented and the comments by all the discussants.

In keeping with the concern over conceptual framework, so as to bring order and focus to our consideration of this topic, note that the agenda has departed from the sequence of five topics cited in the call for papers and is as follows:

- Human Factors in Decision Issues
- Human Performance Models and Applications
- Predicting Human Performance in Combat Environments
- Combat as a Data Source
- Representing Human Performance in Combat Models and Simulations

Each of these topics is treated in our agenda to the extent that appropriate papers were offered on the subject. As you are listening to the various speakers and discussants, I encourage you to keep in mind not just the paper being presented, but to consider the application of all the good work in the context of the following four areas:

1. Use of weapons so as to take full advantage of potential firepower
2. Cohesion and effectiveness of units, especially as losses are taken and replacements made
3. Individual, crew, and unit performance as constrained by environmental conditions and human performance
4. Innovation/ingenuity-as in decision-making with information missing or in error, or other fog of war circumstances

Two more points I wish to make. To enhance the value to each of you attending this symposium, we have discussants for most of the papers. These discussants will try to apply their experience to the subject of the paper they are addressing so as to provide, on the one hand, a broader view of the utility of the author's work to the overall problem we are considering and, as needed, a critical commentary. In addition, we have allowed time for questions and answers following the discussants comments.

The second point is to recognize again the broad range of professional capabilities and interests you folks bring to this forum. You represent operations research, systems analysis, human factors and human engineering, and behavioral sciences, at least. Each of us sees this meeting as an opportunity to interface with people of different professional capabilities but with a common interest and sense of purpose. Therefore, our agenda provides opportunity for meeting each other and discussing the issues - at the lunch breaks, at the mixer tonight, and the dinner tomorrow night.

The program committee has had a ball putting this symposium and the agenda together. It has been a privilege to work with all of you authors, speakers, discussants, and session chairs. Up to now, we have looked at this as being <u>our</u> meeting. From here on, the symposium is <u>your</u> meeting, to conduct, to participate in, and to learn from. Enjoy yourselves, and I hope the agenda fully lives up to your expectations!

Now I turn the meeting over to the chairman of the first session, Mr. Eugene Visco.

AFTER DINNER ADDRESS

Dr. L. Ronald Speight, Chief, Operations Research Division
NATO SHAPE Technical Centre

Ladies and Gentlemen, we meet on an historic occasion, with the US Military Operations Research Society being joined by elements of NATO to launch an attack on one of the great challenges of military analysis - that of the human factor in battle. I feel that I should open on an historical note.

I work for the Supreme Headquarters, Allied Powers Europe, at their Technical Centre in the Hague. Very many times a year I travel from the Netherlands down the motorway to our military headquarters in Mons, Belgium. As I skirt Brussels I pass the monument of the Butte de Lion. It was there, at the battle of Waterloo, that the Second Earl of Uxbridge and Marquis of Anglesey, Lord Henry Paget, turned to the Duke of Wellington after one of the last French cannonades rolled out and exclaimed: 'By God! I've lost my leg!'

'Have you, by God?' Wellington replied, then supported him for a few moments before galloping on on his horse Copenhagen. Later Colonel Felton Hervey tried to get Wellington to move back to a less exposed position, saying 'We are getting into enclosed ground, and your life is too valuable to be thrown away.' But Wellington would have none of it. 'Never mind,' he said, 'Let them fire away. The battle's won; my life is of no consequence now!'

There are three reasons for sketching in this little historical vignette. First of all, it establishes my British credentials. Secondly, it paints a picture of war as we would like it to be: peopled by heroes, who behave in a cool, calm and collected fashion, if on this occasion a trifle dispassionately. But most of war is not like that. Waterloo itself was full of terror and confusion, and perhaps was lost to Napoleon when the Imperial Guard itself failed to stand. It is the characterisation of behaviour in battle which now preoccupies us.

But, lastly, I am happy to act as an Uxbridge to Steve Murtaugh's Wellington, if Steve can just forgive me for imposing on him this temporary British citizenship. We meet to do historic battle against a great foe. I trust in Steve to lead us to victory, and I shall be proud to have been at his side, even though I suspect that I shall have lost a leg in the process.

The first thing a staunch lieutenant must do is to obey orders, and Steve has commanded me to give you an account of European military OR. And so I shall. I shall give you a view which is entirely personal, opinionated and probably inaccurate, before I rejoin the main battle with the human factor at the end of my address.

You must be aware that we in the rest of NATO look to you in the States, and we see a military operations research scene which appears incredibly complex, and brimming with human and material resources. In comparison our European scene seems austere and simple. Let me first deal with the Northern tier. Both Norway and Denmark have their national Defence Research Establishments, with OR staffs in the region of 20 and a dozen scientists respectively. These sections have to tackle the whole gamut of air, land and sea problems at every level of aggregation.

The Central Region of the Allied Command Europe is by far the most richly provided with OR resources, with Germany being the most plentifully endowed. Their military OR is concentrated almost entirely within IABG - an organisation which is independent, but with the military analysis funded and controlled by their Ministry of Defence. They have some 200 OR scientists near Munich, with perhaps another 50 at Trier. The Dutch organise things in very similar fashion with their TNO, some 50 analysts serving right next door to my own organisation in the Hague.

France holds their military analysts within the Ministry of Defence, with perhaps 100 practitioners operating centrally in the Centre d'Analyse de Defense, and further groups of about 20 in outstations serving the individual services. Belgium has a military wargaming effort, but no additional military analysis to speak of. Lastly, in the Central Region, the UK has one group of about 60 scientists in the Defence Operational Analysis Establishment, concentrating on high level studies for all three services. More detailed studies are tackled in the land, sea and air R & D Establishments near Sevenoaks, Portsdown and Farnborough. These single service groups have something in the order of 20 to 30 scientists each.

And so, turning to the Southern tier, the scene is practically deserted. Spain, Portugal, Italy and Greece have practically no military OR to speak of. Within the Turkish General Staff there is a small analytical cell, peopled mainly by young serving officers. Inevitably they are loaded with a host of small scale, immediately applied, studies; and yet they breed some of our very best European analysts.

And I should not forget ourselves, the SHAPE Technical Centre. Our strength is some 30 scientists, drawn from all the NATO nations, and our raison de etre is that we should provide advice to the Supreme Allied Commander Europe which can be seen to be independent of purely national concerns. Between them our 30 analysts cover all 3 services over the whole of Allied Command Europe. Inevitably, most of our analysis is pitched at the theatre level. Of huge concern now, when NATO is poised on the brink of serious negotiations with the Warsaw Pact, is the question of conventional arms control. How best to characterise the whole military balance between these two alliances, and the factors which affect it? What are the defensive principles which General Galvin must preserve at any cost if he is to have a fair chance of success in executing his mission to defend the Allied Command Europe?

War is becoming more complex by the day. Technology hardly sets the limit any more. Anything you want to do you can do, at a cost and in time. And as measure begets countermeasure and counter counter-measure, and as each reacts with all the others and with the military environment, the battle models become ever more complex. Any analytical agency with pretensions to original thought is embarked on battle model construction to some degree, trying to represent all this complexity, and performing miracles of modular and structured design to keep it all under control somehow. But increasingly each agency has to concentrate its efforts in this regard into fewer and fewer selected paths of endeavour. Outside those favoured areas buying or borrowing someone else's model is the rule. The less well off borrow from those rich in resources. The poor borrow from those in the middle income bracket, who may be expected to have instituted economy measures to make the complex models less profligate in their support needs; because, even if the expensive toy is given you free by a rich Uncle, you have to look anxiously at the bill you are going to have to pay to keep it on the road. Can you afford what it takes to keep it going?

And so we in Europe look across the Atlantic, and in this most advanced of NATO nations we see the trend towards model complexity at its most advanced. In military hardware we advance through the Flying Fortress and the Mitchell bomber to the B2, perhaps to the single ultimate flying machine which will do everything, win the whole war, and swallow the whole US defence budget. And so it goes with battle models. Computer power is no longer the limitation -tomorrow's PC will be like having a million IBM PS-2s on your desk. Why simplify it if you can model it? Why not represent every trigger pull, and the aerodynamics of each individual bullet? Eventually we shall be able to include the expected insect population in each portion of the globe as a function of season, weather and time of day, and, if some poor unfortunate gnat gets in the way of the speeding bullet, we shall be able to map the fractional deviation in its course. I know, you think I'm exaggerating. But there will come a time when we shall have to stop. Most of us in Europe have already got to the stage when we know that we can't support tomorrow's models. Even if we have faith in the algorithms and the model logic; even if the coding is transparent so that we understand it (because what self-respecting analyst will use a model he cannot understand?); how then can we possibly meet and maintain the data requirements? And so, many of us in Europe know that we can't go on like this. There can be no going back to a time of innocence - to those early days of OR, in which real operational problems were represented in simple abstract manner, and then these mathematical constructs were manipulated in an elegant and satisfying fashion. Those days are gone, but the role of conceptualisation is returning.

And so at last I return to the human factor. How does it fit in to this world of increasing model complexity? Can we honestly model the tremor in each trigger finger; the effect on a soldier's concentration of just having seen his companion killed; or even, like Marshall Ney in Waterloo, whether he was in the right place at the right time? And, even if we can break battle down into all its multifarious component parts, where on earth are we going to get the data on each and every effect, let alone their interactions? As we search for ever greater fidelity in the minutiae of battle, do we not run the risk of being wonderfully, gloriously wrong, but with very great precision and in very, very great detail?

I wish I had taken notes on all that I had written to Steve Murtaugh. Then I could quote myself to you. But let me put it to you, can we hope to succeed by continuing only in a peacemeal fashion? By characterising experimentally how this particular stressor, at this particular level, affects this particular subtask? How subtasks and individuals combine, how stressors interact, and so on, until we have built up a picture of the whole war? I suspect that the onward march of military technology will keep us forever busy evaluating new subtasks; and, in any case, the experimental laboratory of real war is mercifully rare. When it does present itself, the experimental design and the level of scientific instrumentation and recording leaves a lot to be desired. And so can we honestly expect a truly useful picture to emerge from the simple aggregation of separate uncoordinated effort? It seems to me that we need a grand and unifying vision -what I have called a conceptual framework for considering the human factor in battle. Individual results may be fitted into this framework - extending it, or modifying it if needs be.

It seems to me, then, that in this matter we need to think very carefully of our overall strategy. If there is one thing which I hope we can do at this symposium it is perhaps to start to draw up a concept of operations for the future. And so perhaps I was wrong to liken this setting to a battlefield. Who wants to repeat a battle like Waterloo, which led Wellington to exclaim 'Hard pounding this, gentlemen; try who can pound the longest?' And how can we wage war on Napoleon, of all people, who gave rise to the well known phrase 'In war, moral considerations account for three-quarters, the balance of actual forces for the other quarter'.

Instead, I think, we want to forge an alliance, just as we are forging an alliance between your great nation and the other nations of NATO. Together we must hammer out a grand strategy for this alliance. We must think on the strategic and operational scale where, the Soviets would have it, great wars are won. We must plan to produce a victory in the human factors campaign without any need for a Waterloo. Then I can be spared from saying, as Uxbridge did to one of his lovers: 'Well, Marquise, you see I shan't be able to dance with you any more except with a wood leg.' And we can spare Steve Murtaugh the thought that, because the battle is won, his life is of no consequence now.

SESSION I: HUMAN FACTORS IN DECISION ISSUES

SESSION CHAIR: Eugene P. Visco, Director, Study Program
Management Agency, Office of the Deputy
Under Secretary of the Army (Operations
Research)

Four papers comprise this initial session of the symposium; the papers are designed to range over the basic topic of the meeting and thus provide a true introduction to the topic. The first paper, by COL Joseph E. Stull, describes a most important Army "laboratory" for gathering data on human performance under conditions closely approximating combat. COL Stull recently served as Commander, 32nd Guards Motorized Rifle Regiment, the opposing force (or OPFOR) at the Army's National Training Center. The second paper, by LTC William O. Blackwood and Mr. Cooper Wright, focuses on a planned ambitious effort of incorporating human behavior dimensions into national net technical assessments. The work is just getting underway; it will be interesting to revisit the topic in future MORIMOC fora to benefit more directly from the products. Because of scheduling conflicts, the third paper, by Mr. William R. Beuch and Dr. Alan Rehm, was read, with little advance preparation, by Prof. Wayne Hughes. In spite of the difficulties, Prof. Hughes was able to provide basic information on how the Soviet military focus on human behavior in developing operational plans. The final paper in the session, by Messrs. Ed M. Dougherty, Jr. and Joseph R. Fragola, stems in part from their book, <u>Human Reliability Analysis, A Systems Engineering Approach with Nuclear Power Plant Applications</u>. The human performance postulated during nuclear accident management is compared to that in combat and the modeling approach used in the nuclear case is postulated to apply in the combat arena. These presentations provide a good entree to the balance of the symposium.

THE NTC OPPOSING FORCE (OPFOR), A THREAT REPLICATION OF THE FIRST ORDER

By

JOSEPH E. STULL, COL, INFANTRY, USA, CAA

To be an effective analyst, one needs several things: the best data available, a reliable source for that data, and the knowledge of what to do with it once you get it.

In the world of computer simulation we have routinely used data and lessons learned from previous wars to help us represent it accurately. These include: weapon system data; personnel data (casualties, etc.); ammunition expenditures; and equipment data. In the absence of war; or sometimes during war, we frequently have had to use laboratory type data, i.e., the best available.

The questions we must address is: Are we using the best data on war fighting, currently available and tested, that we can get, especially about the threat, the United States forces and allies? Hopefully we are, but, if we are not, I would challenge all of us to exploit every possibility to get and use the very best data available.

The Army may have provided us some great help--either knowingly or unknowingly--with the fielding of its unique and special training centers, one of which is the National Training Center, Fort. Irwin, California. Even though still in the embryonic stages, I believe a great opportunity exists in them--the NTCs; and to stimulate your interest and enthusiasm I am going to give you some insights into one of those centers, the NTC. I am going to take you on a quick tour of the battlefield at Fort. Irwin and give you a birdseye view of "war at the NTC" from the "eye of the enemy."

The purpose of this presentation is to provide you an overview and insights into the NTC (Fort. Irwin) operations with an eye towards applications and utility for the ORSA and simulations community.

The presentation will cover briefly: the role and mission of the NTC; type and concepts of operations; OPFOR operations and techniques; training results and implications; and, challenges for the future.

The Opposing Force The OPFOR has a reputation as a, and many believe it is, truly great fighting force. It has been said (incorrectly, I must admit) that " It has never lost a battle" at the NTC. One recent article in the Federal Computer Week newspaper stated, "it's funny, but the best Soviet Regiment is probably American." You can form your own opinion, but I want you to know up front that the OPFOR is most proud of the fact that it accomplishes its mission and that is to: REPLICATE THREAT FORCE OPERATIONS IN A TACTICAL ENVIRONMENT FOR TRAINING.

To do that and to assist in the critically important mission of training the CONUS based heavy infantry and armored task forces (primarily battalion-sized), it must be: well trained, resourceful, disciplined, dedicated, motivated, tough, and capable.

Missions and Organization. There are, generally, two (2) types of training exercises conducted at the NTC. They are force on force and live fire. Both "BLUFOR" and 'OPFOR" conduct or are assigned similar missions, i.e., attack, defend, delay, movement to contact, and meeting engagements.

During force on force exercises BLUFOR is pitted against a realistic enemy with both sides conducting missions assigned by a higher headquarters. Neither side (BLUFOR nor OPFOR) is provided with the plans or concept used by the other. It is truly free play, but both sides employ laser weapon systems. The live fire exercise is similar, except the BLUFOR is pitted against a simulated and automated threat and live ammunition is used. In all cases the units are training to a standard and using tactics and doctrine found in current publications to the best of their abilities.

The most common sized unit conducting training/exercises for BLUFOR is the heavy battalion task force (mechanized infantry or armor) with an operational brigade headquarters and an appropriate combat support and combat service support slice. The OPFOR is organized as a Soviet-threat styled regiment.

The NTC Environment. Fort Irwin, California, is located approximately 40 miles east of Barstow, California, and is 150 miles north of Los Angles and 150 miles south of Las Vegas, Nevada (midway between the two). It is in the Mojave Desert and it is a harsh environment. Fort Irwin is approximately the size of the state of Rhode Island and presents a varied desert terrain that ranges from flat rolling open ground to deep gullies and waddies to high mountainous terrain. The climate is also varied ranging from extreme heat in the summer to cold and windy during winter. There is very little plant life or vegetation. This is a great place for good (threat) realistic training.

What Makes the 'OPFOR" Tick? The OPFOR is made up of United States soldiers like any other United States unit. Its soldiers or units have no formal special training in threat operations. Prior to joining or becoming OPFOR members, individuals are put through a brief "OPFOR Academy", approximately one (1) week, at Fort Irwin. Training emphasis is primarily on OPFOR formations, techniques, tactics, navigation, and vehicle and weapons (including 'MILES") operations and maintenance. Techniques for survival are also stressed.

The bulk of OPFOR techniques and skills come from on the job training. Fundamentals are stressed and quickly and routinely practiced, i.e., MOPP-4 training for an NBC environment is a way of life.

Making "MILES" work is critical and becomes second nature; MILES is the Multiple Integrated Laser Engagement System and one must not only understand how it works, but must also know how to maintain it and have confidence in it. You live or die by MILES so it has major significance. If the laser truly does what it is/was designed to do, i.e., simulate the characteristics of the real system, then you must make it work because when you transition from a "peacetime training war" to a "wartime war" you will have confidence in the real system.

Hard work, dedication, pride and professionalism are key trademarks of the OPFOR. Simple rules govern operations, i.e., if you get a target in your sight, you must kill it, everytime; always maintain momentum; a dead recon element is useless, stay alive and report on the enemy; every vehicle and weapon must be in the battle, not in the maintenance bay, etc.

The OPFOR was not and is not perfect, but its success is based in part, and to a great degree, on doing the basics well, i.e., planning well, conducting effective reconnaissance, employment of its people and equipment, use of terrain, agressiveness, and maintenance of people and equipment.

Outcomes/Results of NTC Exercises/Operations/Training. The training conducted at the NTC relies on several key entities for the results achieved:

the BLUFOR unit (people and equipment);
the OPFOR unit (people and equipment);
the observer/controllers;
the computer; and,
the Multiple Integrated Laser Engagement System (MILES)

Training outcomes are measured against or achieved within the seven (7) operating systems:

intelligence;
maneuver;
fire support;
air defense artillery;
mobility/counter mobility/survivability;
command and control; and
combat service support

Data collected is also based on these seven operating systems. The specific data elements or sources used at the NTC are: .

instrumented data;
take home packages;
after action reviews;
communications;
operations orders; and,
observer/controllers

Keep in mind that the fundamental purpose of the NTC is realistic training and it takes all of these elements to make it effective. Additionally, a historical flaw in the United States Army training has been the inability to do worthwhile analysis. No practical means was readily available to systematically watch the battle unfold and see where, who, and what went wrong. The NTC system has gone a long way in solving this dilemma through the use of the computer and MILES.

The problem of the "non-shooting" gun has been reasonably solved by "MILES" which consists of eye safe lasers attached to rifles, tank-guns, helicopters and other weapons. All soldiers, tank systems and other weapon systems carry laser sensors that sound an alarm when they are shot by another laser system. The laser pulse is coded so each receiver knows which weapon system fired it. MILES, in effect, insures realism and units/forces win or lose depending on their ability to shoot and the effectiveness of their planning. At the NTC success is based on success. The book does not really matter; commanders can try anything and if it works, it works. The book can be rewritten and often it is based on NTC outcomes. The Army realizes that new weapons will often require changes in tactics and techniques and the best place to find that out is at a center such as the NTC and during peacetime vice actual war.

Potential Applications for the ORSA Community. There is no doubt that there is much potential for the use of data being captured at the NTC. In the analytical world where a war/campaign is routinely simulated as part of the process in determining force requirements or capability, the need for accurate data is obvious. Do we still have to rely solely on laboratory results, i.e., test results, or historical battle outcomes with outdated equipment, organizations, tactics and doctrine? I say "No." We have an opportunity to get current and highly realistic data that can either be used to verify existing data or used, after validation, as replacement data. Data such as:

Weapons effectiveness/engagement outcomes (P_H/P_K)
Ammunition expenditures
FEBA movement
Unit movement rates
Successful tactics and techniques
Casualty results
Human factors under selected conditions
Threat system information
Equipment strengths and/or shortcomings
Other

Even though the concepts for collecting this type of data may need refinement, formats needed for subsequent application by the ORSA Community must be determined, the frequency of collection, methods of validation, etc. may need determination also, the potential is there and it only takes a bright mind to resolve the challenge.

War at the NTC. War at the NTC is realistic and tough. Through the employment of a realistic and tough enemy, the OPFOR, the Army has succeeded in enhancing the "go to war" skills of its combat forces and increased our chances for success when fighting outnumbered. The OPFOR looks at "war at the NTC" as a challenge, and similarly the ORSA community must look at "war at the NTC" as both a challenge and an opportunity to improve its product through improved simulations and analysis by using the very best data available from a reliable source.

The ORSA Challenge. All of you should ask and answer the following question: Do I, or my agency, model the war, within existing capability, the way comanders expect it to go? If the answer is yes, I say go back to sleep; if it is no, I say let us get up and get with the program and improve our products through an improved data base. The OPFOR's motto is: Be the toughest enemy our country's Army will ever face and if the BLUFOR succeeds against the OPFOR, then most assuredly we will handle the real threat. Similarly, if the ORSA community provides our country with the best analysis and recommendations possible, then our leaders will make the best decisions ever.

HUMAN FACTORS ENGINEERING ASSESSMENT IN NET TECHNICAL ASSESSMENTS

by

William O. Blackwood, LTC., US Army
and
Cooper L. Wright, Automation Research Systems, Ltd.

ABSTRACT

This paper reports on a project designed to develop and apply algorithms that address the human dimension of combat effectiveness. Included will be the development of algorithms or subroutines for existing combat models which would support the comparative analysis of personnel in opposing fighting forces. Resting on a common data base, the algorithms will be built for various levels of analysis, and provide insights or answer questions typically raised in the net technical assessment process.

BACKGROUND

Numerous models are used by the Services and Department of Defense (DoD) agencies to develop information crucial to the acquisition decision process. These models are used to predict technical performance, force structure requirements, and cost operational effectiveness analysis. While model inputs are complex and typically include variables related to technology, equipment configuration, cost, environmental conditions, force composition, and threat, they generally do not include the impact of human performance. Although the human contribution to battle outcome has long been recognized, the human's role in the enhancement or degradation of total system performance has not been adequately considered in weapon system design and development. "The truth is that the most expensive weapon that technology can produce is worth not an iota more than the skill and will of the man who uses it."[1]

AirLand Battle Doctrine characterizes the future battlefield in terms of lethality, operational dispersal, leader to led ratios, stress, casualties, sleep deprivation, electronic warfare, and uncertainty about the battle situation. These will have a significant impact on new weapon system characteristics and hence engineering design implications for time, space, and velocity. As figure 1 shows, the human has clearly become a limiting factor.

•DISTANCE

SIZE OF BATTLEFIELD	CORPS AREA	150-180 Km
LOW EARTH ORBIT		112-193 Km

•TIME IN SECONDS

MILLI	10-3	Light Travels	300 Km
MICRO	10-6	" "	300 m
NANO	10-9	" "	30 cm
PICO	10-12	" "	.3 mm

•VELOCITY

SPEED OF LIGHT	9.84 X 10^8 FPS
ESCAPE VELOCITY	26,000-55,000 FPS
SPEED OF SOUND	1,089 FPS
HYPERVELOCITY	6,600 - 132,000 FPS

•SELECTED SYSTEMS

M-1 120mm Projectile	5,400 FPS
SR-71	3000 FPS
Apache	304 FPS
M-1	88 FPS

•HUMAN

REACTION TIME	1/2 Sec
PERCEPTION AS INSTANTANEOUS	1/20 Sec
VELOCITY	30 FPS (9 Sec - 100 yd)
BLACKOUT	9 Gs for 4 Sec or -3Gs for 6 Sec

FIGURE 1

The engineer who works with space, weight, and power problems is forced to consider issues of human isolation/encapsulation, human survivability and habitability, and cognitive stress issues associated with combat operations, if he wants to optimize system performance. Numerous articles have documented problems which have emerged when human issues are ignored. The F-16 program involving "G"induced loss of consciousness" exemplifies the crucial importance of these types of considerations.[2] Technology permitted industry to develop a plane that could withstand higher "G" turns, than the pilot could.

DoD'S INTEREST

DoD's interest in human performance has increased because of pressures to reduce the budget, reform the acquisition process, and improve operational effectiveness. Incidents in the Persian Gulf as well as recent civilian aircraft accidents have focused attention on the high tech debate again.[3] The Services have programs which address to some extent the manpower, personnel, training, safety, human factors, and health issues during the acquisition process.[4] The DoD Directive 5000.53, dated 30 December 1988, discusses these same issues. Congressional interest is also significant; legislation regarding manpower, personnel, training, and safety in the acquisition process has been passed (Title 10 U.S.C., Section 2434). However, much work remains to be done to fully implement these programs and directives.

Net technical assessments compare U.S. and foreign military weapons or equipment to determine capability changes due to new or different technologies. Although net technical assessments have been performed since the late 1960's, there has not been an attempt to consider the human factors engineering contribution into the assessment at the individual, crew, or unit level. While much information exists about the human contribution to combat and its interrelation with technology, it has not been satisfactorily captured in mathematical algorithms, nor are such algorithms technically adapted for inclusion in existing combat models.

As a result of Congressional and Departmental interest in both human performance and net technical assessment areas, an effort was initiated to develop algorithms that factor human contributions to system and unit effectiveness into the net technical assessment process[5]. This project will develop and apply algorithms that address the human dimension. Included will be the development of subroutines for existing combat models that account for the human contribution to combat effectiveness which would support a comparative analysis of personnel in opposing fighting forces. Resting on a common data base, these algoritms will be created for various levels of analysis, and provide insights or answer questions typically raised in the net technical assessment process.

APPROACH AND SCOPE

The project is a three year effort with built-in evaluation and decision points at the end of each year. Additionally, a study advisory group, composed of approximately twenty sponsor-designated members chosen from industry, academia, and government, will review progress on a regularly scheduled basis.

During the first year, US and threat data will be collected in order to describe the relationships and parameters of a general-purpose human performance algorithm. The new algorithm will be designed as a simplified spreadsheet model, capable of being run on a common personal computer using existing data base management software.

During the second year, the algorithm will be tested using data collected on US and threat weapon systems. Modifications to the algorithm and expansion of the data base will be made as a result of this evaluation effort. The project team has already contacted several contractors of major end-item systems who are willing to provide their in-house data.

During the third year of the project, an attempt will be made to scale up from one-on-one duels to task force engagements. Finally, the algorithm will be incorporated in a selected combat model on a test basis.

The project is designed to gather the data necessary to determine the structure and relationships of the algorithm's human performance factors. An extensive literature review of data bases and information repositories on psychological, sociological, anthropological, medical, and other human factors that affect human performance will be performed. Included in these will be:

Central Intelligence Agency
Defense Intelligence Agency
Foreign Science and Technology Center
Idaho National Engineering Lab
Naval Aerospace Medical Research Laboratory
NASA Ames
National Training Center Data Base
USAF Human Resources Laboratory
USAF School of Aerospace Medicine
US Army Aeromedical Research Lab
US Army Human Engineering Library
US Army Research Institute
Walter Reed Army Institute of Research

The framework of the literature search (Figure 2) facilitates the categorization of human performance measures of effectiveness developed for different organizational levels from system component to echelons above corps. The effort will be structured to capture where and how human performance impacts on force effectiveness.

The baseline for algorithm development will be The Analytical Sciences Corporation's (TASC) model completed in 1983 for the Director of Net Assessment, DoD.[6] The TASC model was part of the umbrella research effort for the Technique for Assessing Comparative Force Modernization (TASCFORM) Project begun in 1978. Although the model provides a means of combining technological quality and quantity in a single measure of force potential, "it is not sufficiently accurate to serve as the basis for selection one type of weapons system over another."[7]

Using the resources of the Science Applications International Corporation's Foreign Systems Research Center, the threat's human factors and engineering design analysis methods will be studied to determine how the threat incorporates human performance into their effectiveness analyses. Rather than assume that threat actions and perceptions are governed by procedures, objectives, and constraints similar to those under which the US operates, the intent of this effort will be to identify the asymmetries between the two. This will provide structure to the relationships and key factors addressed in the threat version of the algorithm.

HUMAN PERFORMANCE MOEs

How do we or the threat take personnel strengths and limitations into account in developing equipment (by category, e.g., Air, Naval, Ground), tactics (by level, e.g., Strategic, Operational, Tactical), and operating procedures?

STATUS UNIT LEVEL	PEACE			MOBILIZATION			WAR			OTHER		
	MOE + Source	Means of Measurement	Data Source	MOE + Source	Means of Measurement	Data Source	MOE + Source	Means of Measurement	Data Source	Models	Change Required	No Change Required
Component												
Sub System												
System (Tank, Plane, Ship)												
Single Mission (Multiple Systems: e.g., Platoon)												
Multiple Missions (Multiple Sys: e.g., Co)												
Tactical (Multiple Systems: e.g., Bn, Squadron)												
(Multiple Systems: e.g., Brigade) (Reg)												
Force Level (Multiple Systems: e.g., Div) (Div)												
Corps (Combined Arms Army)												
Echelons above Corps (Front)												

FIGURE 2

In late 1989, a mini-conference for combat model subject matter experts will be held. The conference will present information on work to-date, including data found during the literature search, and results of the US-threat asymmetry study. Additionally, the new algorithm(s) will be reviewed. During this conference, attendees will have an opportunity to get a hands-on demonstration of the prototype model. The project team will look to attendees for their comments on ease of use, complexity, level of detail, and validity of algorithm assumptions.

PERSONNEL POTENTIAL MODEL

The TASC Personnel Potential Model was developed to quantitatively measure the potential of ground personnel to fully exploit the capability of their equipment in operational situations. The model is a two-sided equation that addresses those factors having the most influence on a soldier's potential to use his assigned weapon system in an operational environment (Figure 3).

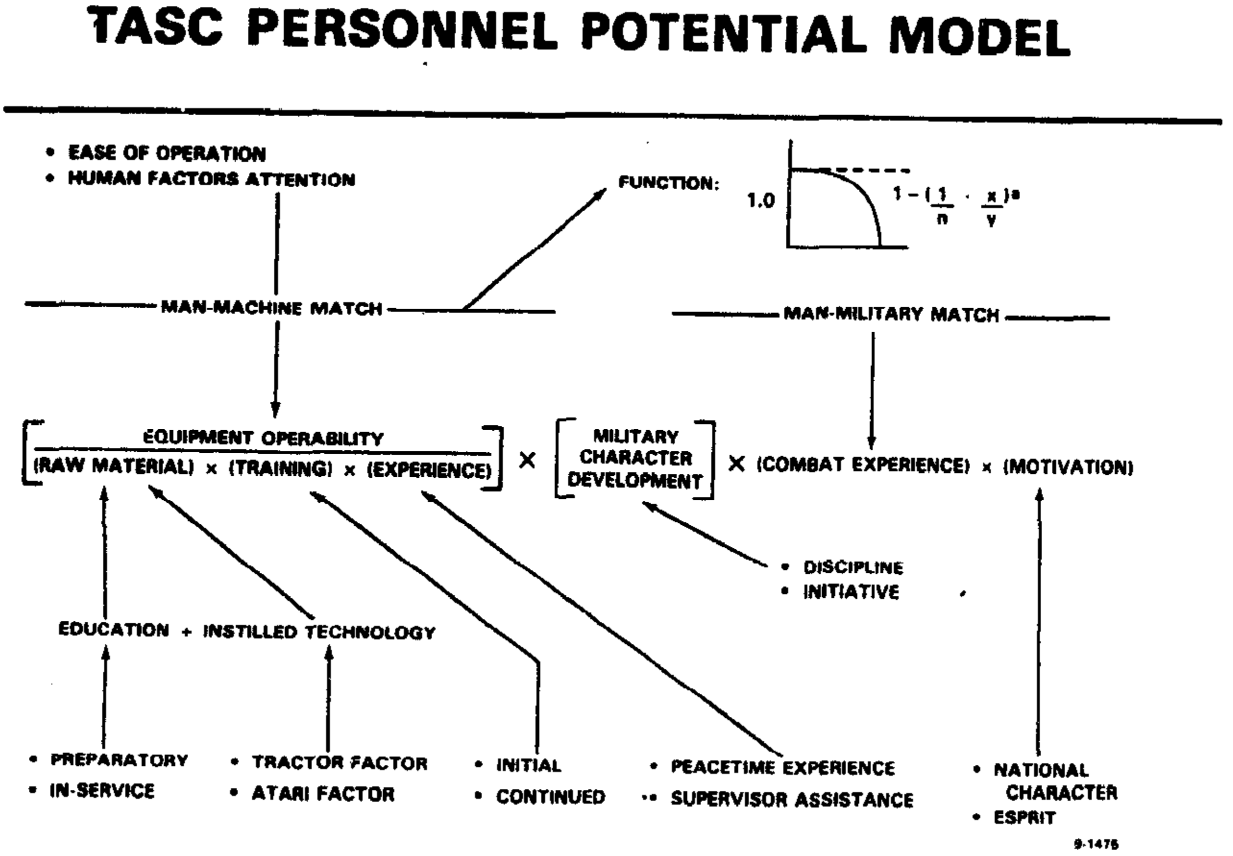


FIGURE 3

The man-machine match term expresses the soldier's ability to use the technology incorporated in his weapon system. This term generally addresses the soldier's skill. The equipment operability factor represents the weapon system's technological qualities, including ease of operations (system complexity) and attention to human factors (how well the system was designed for the soldier). The raw material factor addresses the soldier's education and instilled technology. Education is concerned with the soldier's preparation in reading and mathematics prior to entry as well as his in-service education. Instilled technology indicates the familiarity with technology that a soldier acquires before entering the service. The tractor factor represents his mechanical experience, while the Atari factor indicates his exposure to electronic and microelectronic equipment. Training addresses both initial entry training (preparation to operate and maintain his equipment) as well as on the job training (including large scale field exercises in which the soldier employs his equipment according to specified doctrine). Experience includes peacetime experience (number of years the soldier has operated his assigned equipment), as well as the quality and availability of assistance made available by more skilled supervisors.

The man-military match includes other factors not directly associated with the weapon system, but that affect a soldier's will to perform on the battlefield. The military character development factor includes the process that converts civilians into soldiers. Discipline and initiative are the measures of a soldier's ability and inclination to follow orders in combat while at the same time exercising initiative when required. The combat experience factor recognizes that no training exercise can ever completely replicate the conditions of war. Soldiers who have this experience react automatically to combat situations while the "new guys" in the unit may be slower to react or react improperly.

The last factor in the equation, motivation, consists of national character and esprit. Although thought to be the least measurable of all the factors, motivation remained in the equation because it was widely said by subject matter experts to play a role in combat effectiveness. National character includes human attributes exhibited by citizens of a nation or members of a culture that do not change significantly over time. Esprit, however, is more dynamic and changes quickly as a nation goes to war. Examples of subfactors that impact esprit are the political and economic conditions of the country, the statements of its leaders, and the nature of the conflict. TASC's work will serve as the starting point to produce a family of human performance algorithms that contributes to the DoD acquisition decision process.

LEVELS OF PERSONNEL PERFORMANCE

The models developed under this effort will be built as a series of algorithms, proceeding from individual soldier performance to team (crew, platoon, company) and organization (battalion, brigade, division, army) performance (Figure 4).

At each level of investigation, different factors affect human performance. When considering the individual soldier, his motivation and skill with his assigned weapon are key. The performance of a small group such as a crew, squad, or platoon depends on the contribution made by each one of its members. However, additional factors such as the leadership displayed by the platoon leader and platoon sergeant as well as the cohesion developed among the platoon members have an impact on the performance of the group. At the organizational level, battalion / brigade, performance is dependent on the actions of each subunit sometimes operating independently in combat. However, the additional factors that contribute to organizational success are now more complex and reflect a combination of doctrine, weapon system mix, resource allocations and leadership.

LEVELS OF PERSONNEL PERFORMANCE

$$\text{INDIVIDUAL PERFORMANCE} = f\left[\begin{array}{l}\bullet\ \text{SKILLS AND KNOWLEDGES}\\ \bullet\ \text{MOTIVATION AND ATTITUDES}\\ \bullet\ \text{OPPORTUNITY}\end{array}\right]$$

$$\text{TEAM PERFORMANCE} = f\,[\text{INDIVIDUAL}] \times \left[\begin{array}{l}\bullet\ \text{COLLECTIVE SKILLS}\\ \bullet\ \text{COHESION}\\ \bullet\ \text{EQUIPMENT}\\ \bullet\ \text{LEADERSHIP}\end{array}\right]$$

$$\text{ORGANIZATION PERFORMANCE} = f\,[\text{INDIVIDUAL}] \times [\text{TEAM}] \times \left[\begin{array}{l}\bullet\ \text{DOCTRINE}\\ \bullet\ \text{HARDWARE}\\ \bullet\ \text{SUPPORT}\\ \bullet\ \text{GOALS}\\ \bullet\ \text{RESOURCES}\\ \bullet\ \text{SYSTEM}\\ \quad\text{LEADERSHIP}\end{array}\right]$$

FIGURE 4

The project team will first develop, analyze, evaluate, and refine the high resolution human performance algorithms that are more dependent on the actions and characteristics of the individual soldier. As these algorithms become more refined, work will begin on the lower resolution models that address higher levels of personnel performance. As a result of shared data, links between the algorithms will be established. The challenge for the project team is to identify those soldier performance factors that are critical at each level of interest and trace their importance as the model proceeds from level to level. The algorithms will be further complicated by the requirement to address human performance at three different national levels: peace, mobilization, and war.

HUMAN PERFORMANCE MEASURES OF EFFECTIVENESS (MOE)

The human performance measures of effectiveness matrix (Figure 2) represents the structure for information collection and analysis. The project team will collect human performance information in two dimensions: vertical and horizontal. The vertical component of the matrix establishes ten different organizational levels of investigation, from component to echelons above corps. The models used for analysis at each organizational level will also be listed. Along the horizontal component of the matrix are listed the national readiness postures (peace, mobilization, war). If a measure of effectiveness exists only during a posture period, it will be identified only in that readiness posture column. As an example, the number of casualties a unit suffers affects morale, but that measure of effectiveness can only be assessed during conflict.

To illustrate the function of this matrix, a tank and its crew will be used as an example. The information collection process could begin at the system component level with the eye piece for the tank's thermal sight. Human performance information gathered could include whether the diameter of the eye piece allows soldiers to use the sight while wearing protective masks, how well the rubber eye piece cup fits the full range of soldier face dimensions, and whether the eye piece adjusts to accommodate the vision characteristics of the assigned operator. The implication is that the tank may not be used to its full potential if the soldier either cannot or will not tolerate the discomfort associated with performing the tasks required to fire the main gun unless his survival depends on it.

To continue with this example, information on the tank is gathered at the subsystem level which now includes the entire thermal sight assembly. Information at this level could address the training time and difficulty of tasks to be performed for soldiers operating a thermal sight versus that required for a daylight optical sight. The workload and skills required to support the new tank versus the present tank may have implications for manning the force structure.

At the mission level, the example expands beyond the thermal sight to encompass the entire tank, including its crew and maintainers. Issues that could be addressed include the use of the thermal sight by other crew members, gunner/commander hand off procedures, and the combined effect of noise, heat, blast, and ride quality on the crew's ability to use the thermal sight to detect and hit targets. Even though each one of the tank's individual system hardware components (such as the thermal sight) are precisely designed to perform its assigned function, the addition of the soldier and his integration of the new technologies involved may cause some system degradation. An example of this might be the combination of the damping effect of the tank's suspension system with the viewing angles and motion cues of the thermal sight inducing motion sickness for the gunner.

MEASURES OF EFFECTIVENESS EXAMPLES

Tank ride quality illustrates the difficulty in integrating the various MOEs and associated models over the different organizational levels of investigation anticipated in this project. At the component level, a computer aided design model, such as the Dynamic Analysis Design System, could be used to determine the engineering design requirements for the torsion bar and road arm assembly of the tank's suspension system. At the subsystem level, the interaction of the tank's torsion bars, road wheels, and track over various terrains could be investigated using the vehicle dynamics model to provide an indication of the tank's ride quality. The NATO Reference Mobility Model combines the submodels to provide the amount of absorbed power transmitted to the crew through the tank's suspension system.

The results of these engineering models must be integrated with human tolerance levels to determine actual total system performance. Absorbed power directly affects crew performance. Studies have shown that drivers will operate 87-93% of the time in a comfort zone that does not exceed six watts of absorbed power even though the tank may be designed to travel cross country at high speeds.[8]

To further expand the MOE identification and collection process, a final example is offered to illustrate the differences in a specific MOE as a result of changes to the national readiness posture. During peace, mobilization, and war, the hardware MOEs for a particular system will most likely remain unchanged. An M-1 main battle tank is still a tank whether firing training rounds at Range 8, Fort Hood or engaged against hostile forces in the Fulda Gap. Hardware MOE values will only change if soldiers make field expedient battlefield modifications such as adding improvised armor plating to the vehicle for increased protection.

Most soldier physical and mental characteristics such as lifting capacity and aptitude will remain the same as the national readiness posture moves from peace to war. However, unlike the tank with its static hardware characteristics, the soldier is a much more dynamic system. A reservist called to active duty will increase his stamina and strength as a result of mobilization training. A soldier in combat will develop mental battle skills that allow him to process key information needed to survive and cope with stress.

As a nation prepares for war the effect of fear and stress on human performance changes. Although the National Training Center has increased the level of stress in training exercises for a peacetime army, the soldiers know that the "battle" has a preassigned completion date. If they can just "gut it out" until then, normalcy will return. This is not the case during mobilization and war. The results of the Iran-Iraq (gas warfare) and Falkland Islands conflicts (British psychological operations) highlight the effect that fear and stress can have on a soldier. Although the hardware in most of these cases remained fully operational, operator and crews no longer effectively performed their combat missions.

The examples discussed point to the difficulties expected with this algorithm development effort. Not only is there an abundance of human performance information of questionable quality, very little of it has been tied together so that it reflects the multiple unit levels or the nation's state of readiness. It will be the project team's responsibility, as advised by the study advisory group, to discipline this process so that the necessary information and relationships are obtained to develop a family of human performance algorithms.

COMMUNICATION & INTEGRATION

Communications amongst groups in a multi-disciplinary study such as this is difficult under the best of circumstances. Attempts to integrate human factors (itself an interdisciplinary field) into combat modeling involves communicating with operational users, intelligence analysts, engineers (design & system), operational research and system analysts, mathematicians, computer specialists, physicists, and others. Compounding this is the perception that everyone is an expert on human behavior. Each of these groups has a body of knowledge with its underlying theories, principles, and taxonomies. Furthermore they formulated their own human performance taxonomy and conditions under which it is applied because a commonly accepted taxonomy does not exist.[9] Hence, lack of a common language has been a primary source of confusion and misunderstanding amongst the communities involved.

The Government has had to bridge the communications problem in order to reduce the technical risk in the contract. Bridging the communications gap involved identifying commonalities and finding a way to relate human performance measures to them. As a result, various paradigms in each of the communities (operational, intelligence, combat modeling, engineering, human factors) were analyzed to determine commonalities. The military organizational hierarchy was found to be common to all communities except human factors. Figure 2 reflects the approach to this issue. This framework relates human performance measures to the military organizational structure over varying national readiness postures.

The vertical integration challenge is difficult, and it remains to be determined if the approach will work. Using an example from the armor/anti-armor area Figure 5 reflects vertical integration from basic science to unit effectiveness. At the mission area level an analysis must be very broad, but as the problem is decomposed each area becomes narrower and more focused.

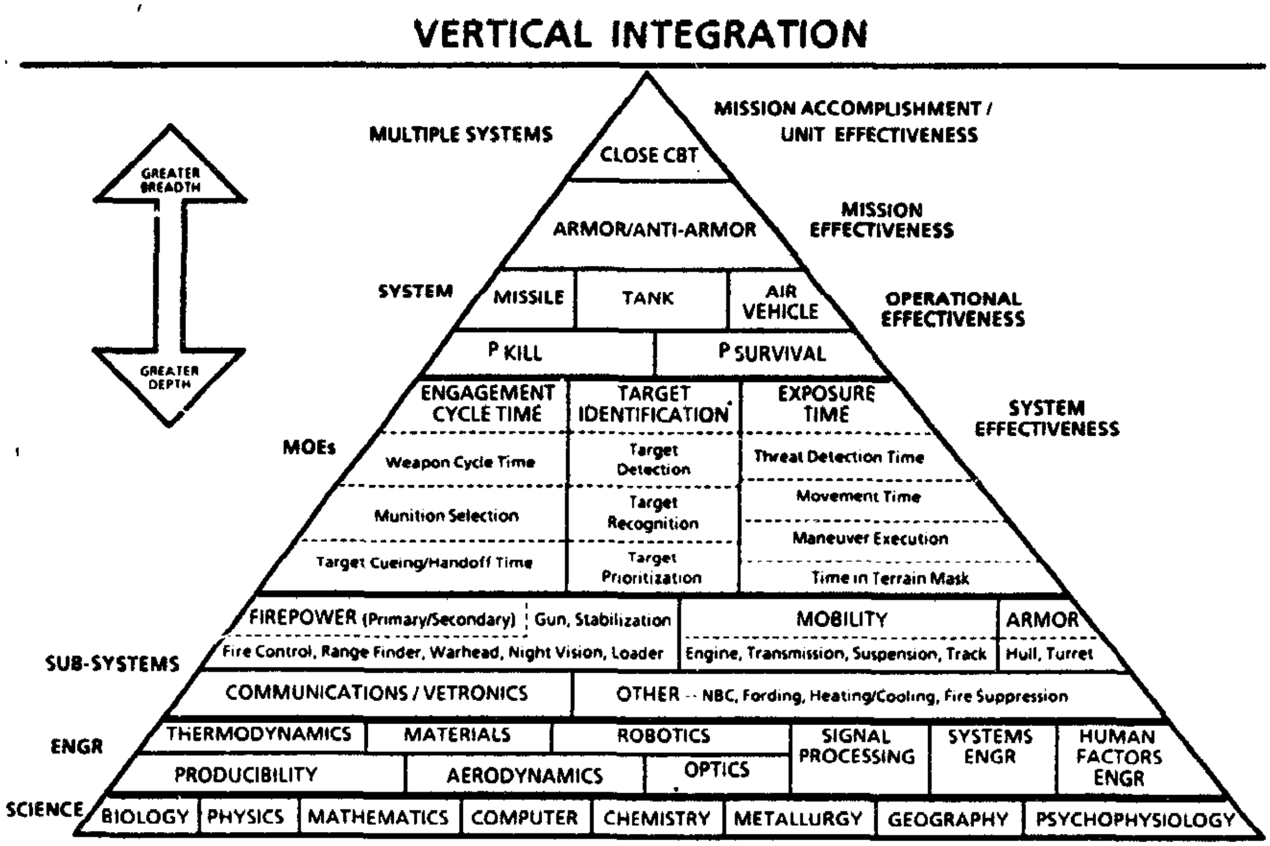


FIGURE 5

The problem is further compounded at the mission area level because systems are designed for multiple roles and missions. Likewise many mission areas overlap so linking systems, capabilities, and mission areas is difficult. Prioritizing systems across mission areas also compounds the analysis and each level of analysis has a group of sponsors and constituents with models that they use. Likewise there are analyses being conducted at the same level that are not coordinated. The vertical and horizontal integration efforts generally do not address common measures of effectiveness. These problems are being addressed by the modeling community, but the question of how to integrate the human role must be addressed in both dimensions at each level.

CONCLUSIONS

Clearly the effort that was outlined is only in its infancy. The work which will be done over the next three years is exciting and has OSD support. The success or failure of the human performance algorithm may be secondary, however, to the greater value generated through improved communications amongst disparate communities on this important topic.

END NOTES

1. Clarke, GEN Bruce C., CG, CONARC (1959), Fifth Annual U.S. Army Human Factors Engineering Conference.

2. Healy, Melissa, "Blackouts Plague Air Force," *Defense Week*, 1 April 1985.

3. Squires, Sally, "The 'Glass Cockpit' Syndrome," *Washington Post*, 11 October 1988, Health Section.

4. Army MANPRINT Program:
    - AR 602-2, Manpower and Personnel Integration (MANPRINT) in Materiel Acquisition Process, dated 12 April 1987.
    - AR 602-1, Human Factors Engineering Programs, dated 15 February 1983.

   Navy HARDMAN Program:
    - OPNAV Instruction 5311.7, Deterring Manpower, Personnel, and Training Requirements for Navy Acquisition, dated 12 August 1985.
    - NAVSEA Instruction 3900.8, Human Factors in the Naval Sea Systems Command, dated 20 March 1985.
    - NAVAIR Instruction 3900.10, Human Factors in Navy Materiel, dated 15 January 1986.
    - NAVAIR Instruction 3900.11, Human Factors in the Naval Air Systems Command, dated 15 January 1986.

   USAF IMPACTS Program:
    - AFR 800-2, Acquisition Program Management, In Staffing.
    - AFR 800-15, Human Factors Engineering, dated 21 December 1984.

5. Title 10 U.S.C., Section 153(a)(2)(C) as amended by Section 201 of Public Law 99-443; and SASC Report 100-117, Calendar Number 247, 17 July 1987/Legislative day 23 June.

6. Leveen, Stephen A., "A Technique for Assessing Personnel Potential of Ground Forces: Methodology and Trial Assessments," The Analytic Sciences Corporation, Technical Report, TR-3798-2-1, 21 November 1983 (The information contained in the discussion of the TASC model was taken from the referenced report).

7. Ibid.

8. Murphy, Newell R. Jr., and Ahmad, Falih H., *Comparison of Measures of Vibration Affecting Occupants of Military Vehicles*, Technical Report GL -86-18, Department of the Army Waterways Experiment Station, Vicksburg, Mississippi, December 1986.

9. Fleishman, Edwin A., & Quaintance, Marilyn K., *Taxonomies of Human Performance* , Academic Press, Inc., Orlando, Fl., 1984.

DISCUSSION OF "THE NTC OPPOSING FORCE, A THREAT REPLICATION"
by J. Stull
and
"HUMAN FACTORS ENGINEERING ASSESSMENT IN NET TECHNICAL ASSESSMENTS"
by W. Blackwood and C. Wright

DISCUSSANT: Eugene P. Visco, ODUSA (OR)

These two papers defy the usual critical commentary. Different reasons justify that remark. The first paper, a fine review with graphic photographs, gave us a good look at the Army's National Training Center. Action at the NTC is said to be the closest to war without live fire. COL Stull knows whereof he speaks; he was once known as "Commander, 32nd Guards Motorized Rifle Regiment," the OPFOR. The main intriguing issue raised by COL Stull is that of the vast amounts of data now emerging from the NTC, data which is potentially of great value to the theme of this meeting. Great care must be exercised in the use of the data, but it is rich and becoming more available.

The paper by Blackwood and Wright represents a plan of work that has just begun. So, we must be patient and reserve our judgment but remain hopeful. Aside from nitpick comments about maintaining consistency in illustrative measurements (e.g., velocity in feet per second, the corps area in square kilometers), the paper is a clear exposition of the plan for the next few years. The authors refer to the Stark, Vincennes and Roberts in highlighting the Department of Defense's interest in the human issue. I think of the human aspects of the three events as: the first representing command failure, the second man-machine interface failure, and the third successful training of sailors to face life- and ship-threatening emergencies. The lessons each case provides are extremely valuable to our considerations of human behavior in combat.

HUMAN FACTORS IN SOVIET RESEARCH

William Beuch
Science Applications International Corporation

and

Dr. Allan Rehm
U.S. Army Concepts Analysis Agency and Center for Naval Analyses

INTRODUCTION

Sally Van Nostrand asked us to discuss Soviet research we had come across in the course of our own work which might provide insights into how the Soviets include human factors in their models. Specifically she asked about Soviet data collection concerning human factors and human factors considered in Soviet models.

Neither of us claim to be human factors specialists. Because what we know about Soviet models is almost exclusively their general approaches, rather than details about particular Soviet combat models, we will mainly discuss the following substitute: human factors considered in Soviet research which appear likely to be those included in one of their combat models.

The paper makes three points: 1) there is a Soviet military human factors community which has been doing research on such matters for a long time; 2) several fundamental questions have been examined again and again over the years; 3) there are several known examples of human factors which are included in Soviet combat models.

Let us start by identifying the group of Soviet officers and the military institutions which seem to be most directly involved in human factors research. Those human factors which do appear in Soviet models are likely to be based upon research by this group.

SOVIET MILITARY PSYCHOLOGICAL RESEARCH

The Soviet military includes perhaps a hundred, or possibly several hundred, of officers with advanced degrees in psychology and pedagogy. Two prominent members of this group are Doctor of Pedagogical Sciences General-Major A.V. Barabanshchikov and Doctor of Psychological Sciences, Professor Colonel N.F. Fedenko who have jointly authored books and articles on a number of subjects. They are on the staff of the Military-Political Academy.

For many years these two wrote on the psychology of small unit cohesion, particularly in a nuclear combat environment. The Russian phrase describing this area of research is the "psychology of the military collective." Neither of us has followed this work, but it apparently deals with how to select members of small units in such a way that they will have the best likelihood of retaining their combat efficiency on the nuclear battlefield. Roughly four or five books have been published in Russian on the subject, but none, to our knowledge, have been translated for wide circulation in English. As will be seen, the question of unit cohesion in a nuclear combat is a recurrent Soviet theme.

The U.S. Air Force "Soviet Military Thought" series of translations includes volumes on both Military Psychology and Military Pedagogy and chapters of the former volume cover summaries of some of this work.

The best known Soviet psychologist is Ivan Pavlov who began conducting his researches on conditioned relfexes using dogs around 1900. Pavlov's results are known through lectures published in English in 1926 and currently available in a Dover reprint. (Conditioned Reflexes, An Investigation of the Physiological Activity of the Cerebral Cortex, reprinted by Dover Publications, Inc. 1960). It is worth noting that those lectures were delivered in 1924 at the Military Medical Academy in Petrograd (now Leningrad). Pavlov's theories strongly influenced Soviet military psychology.

The Military Medical Academy is the second Soviet military institution which seems to have made significant contributions to the study of human factors. The third important institution is a group within the Soviet Air Force which studies the psychology of flight crews.

The extent of Soviet officer research on questions of military psychology from the Revolution up through 1967 was indicated by the Soviet specialists mentioned above in an article in a Soviet psychology journal.

		Number of Publications	
Period	Years	Total	Annual
Inception	1918-1920	10	3 per year
Establishment	1921-1941	150	7-8 per year
Great Patriotic War	1941-1945	30	6 per year
Revolution in Military Science	1945-1967	400	20 per year

Source: A.V. Barabanshchikov, K.K. Pistonov, N.F. Fedenko of the Military Political Academy named for Lenin, "Military Psychology in the USSR, Questions of Psychology, No. 6, November/December, 1967, pages 76-84.

Growth in Soviet military psychology publishing has continued to increase in the past twenty years. The Lenin Military-Political Academy's Department (or Chair) of Military Pedagogy and Psychology (headed in the mid-1960s by Barabanshchikov) appears to be the primary establishment for this research.

The primary subjects of interest to us which we have observed in Soviet military psychology (a biased sample) have been the following:

Troop morale
Psychological stability of personnel
Fighting qualities of personnel
Unit cohesion (particularly on a nuclear battlefield)
Training for leader-subordinate coordination of actions while out-of-communication
Psychological questions connected with camoflage
Man-machine systems
Personnel selection
Scientific organization of labor
Training methods and psychology of learning
Flight psychology

The U.S. Air Force "Soviet Military Thought" series of translations includes volumes on both Military Psychology and Military Pedagogy and chapters of the former volume cover summaries of some of this work.

The best known Soviet psychologist is Ivan Pavlov who began conducting his researches on conditioned relfexes using dogs around 1900. Pavlov's results are known through lectures published in English in 1926 and currently available in a Dover reprint. (Conditioned Reflexes, An Investigation of the Physiological Activity of the Cerebral Cortex, reprinted by Dover Publications, Inc. 1960). It is worth noting that those lectures were delivered in 1924 at the Military Medical Academy in Petrograd (now Leningrad). Pavlov's theories strongly influenced Soviet military psychology.

The Military Medical Academy is the second Soviet military institution which seems to have made significant contributions to the study of human factors. The third important institution is a group within the Soviet Air Force which studies the psychology of flight crews.

The extent of Soviet officer research on questions of military psychology from the Revolution up through 1967 was indicated by the Soviet specialists mentioned above in an article in a Soviet psychology journal.

		Number of Publications	
Period	Years	Total	Annual
Inception	1918-1920	10	3 per year
Establishment	1921-1941	150	7-8 per year
Great Patriotic War	1941-1945	30	6 per year
Revolution in Military Science	1945-1967	400	20 per year

Source: A.V. Barabanshchikov, K.K. Pistonov, N.F. Fedenko of the Military Political Academy named for Lenin, "Military Psychology in the USSR, Questions of Psychology, No. 6, November/December, 1967, pages 76-84.

Growth in Soviet military psychology publishing has continued to increase in the past twenty years. The Lenin Military-Political Academy's Department (or Chair) of Military Pedagogy and Psychology (headed in the mid-1960s by Barabanshchikov) appears to be the primary establishment for this research.

The primary subjects of interest to us which we have observed in Soviet military psychology (a biased sample) have been the following:

Troop morale
Psychological stability of personnel
Fighting qualities of personnel
Unit cohesion (particularly on a nuclear battlefield)
Training for leader-subordinate coordination of actions while out-of-communication
Psychological questions connected with camoflage
Man-machine systems
Personnel selection
Scientific organization of labor
Training methods and psychology of learning
Flight psychology

Most of these questions are of long-standing interest in Soviet research. This paper will give only a few examples. The references at the end of the paper can be used for tracing down more of the Soviet literature in depth.

The Soviet military thus has long had both officers and institutions specifically devoted to military psychology and human factors. Next we examine a few specific examples of the subjects of human factors research.

HUMAN FACTORS IN WARTIME PLANNING AND COMBAT MODELS

An article in the Soviet Military Encyclopedia which discusses correlations of forces and means implies that in addition to armament and other combat equipment the Soviets would like to take the following into account when analyzing force balances:

Peculiarities of the organization of the sides
Level of combat preparation
National composition of troops and their morale
Capabilities of reconnaissance forces and means
Troop combat experience
Stability of troop control
Materiel-technical support
Character of the terrain and its engineer preparation

The list terminates with the phrase, "and other factors which lend themselves to mathematical expressions which can be compared with the help of this or that coefficient, and the remainder expressed by the determination of 'better' or 'worse'." It is not specified which factors can be represented by coefficients, and which simply through the qualitative assessments of who has the advantage. A few examples are known.

National factors for troops have been cited by Christopher Donnelly, and others are given in General-Colonel Valentin V. Druzhinin and Colonel David S. Kontorov's very interesting 1976 book, Problems of Military Systems Technology. Donnelly reported factors for Allied troops ranging from 1 for Germans, and from 0.25 through 0.8 for other NATO troops. The numbers are apparantly used against troops who have not previously experienced combat in such applications as modifying norms for the number of rounds in an artillery bombardment required to produce a standard degree of suppression.

Druzhinin and Kontorov discuss a combat model that includes coefficients of troop efficiency which are to be determined either "from experience" or varied as parameters. As an example they examine a battle in the Caucasus during World War II against the German 17th Army. They cite the following coefficients of troop efficiency which they state have been assumed "arbitrarily," implying they are not the true values.

Soviet troops	1.0
German troops	0.7
German allies	0.5

Most German allies of World War II are now Soviet allies.

As further evidence of the importance of troop quality in Soviet models, General-Major Konstantin V. Tarakanov, in his 1974 book, Mathematics and Armed Combat, analyzed a World War II attack against the Romanians in which the Soviets attacked, even though outnumbered. Tarakanov notes that the Rumanians were "not noted" for determination and will for victory when compared to Soviet troops. Western analysts examining Soviet use of force ratios in the correlation of forces normally expect a tactical advantage of 3:1 or more. But quality of troops can modify these requirements considerably, if Tarakanov's example is representative. Tarakanov's table is given below.

Forces and Means	Soviet 157 Rif Div	Rumanian 13 Inf Div	Soviet 421 R Div	Rumanian 15 Inf Div	Soviet 3 Marine Regt	Rumania 15 Inf Div
Riflemen	950	1330	540	2300	250	200
% TO&E strength	22.5	30.0	13.0	35.0	13.0	35.0
Light MG, %	24.0	9.0	14.0	5.0	14.0	5.0
Tripod Mtd MG, %	24.0	7.0	14.0	5.0	14.0	5.0
Artillery subunits personnel	54	110	0	250	0	0
% TO&E strength	10.0	10.0	0	20.0	0	20.0
Artillery, %	20.0	20.0	0	22.0	0	22.0
Prob of Success	29.0	35.0	57.0	9.3	57.0	9.0

Tarakanov's criteria for defeating the enemy is destruction of 30% of the artillery and 40% of the infantry. In the process, the Soviet side on the offensive is expected to hold their losses to no more than 20% of the artillery and no more than 50% of the infantry.

If these are representative examples, Soviet models do include national factors and they probably include some of the other factors cited in the Soviet Military Encyclopedia article. The coefficients are based on historical combat experience or are varied parametrically. These values may be human factors, or human factors plus equipment and organization. Whatever they are, they indicate a Soviet willingness to use quantitative estimates of human performance in models and probably also in operational decision-making.

NORMS

The Soviets provide performance standards for almost every task one can think of in the military. Soldiers are rated in their ability to meet these norms and performance records are kept in their personnel file. Unit readiness is assessed on the basis of individual and group performance. For example, there are norms for how long it would take to deploy a camouflage net over a gun, how long to dig in a tank, how long to break down a machine gun from firing position to backpack, or how long to perform artillery calculations for firing.

Soliders have reported that there are more stringent norms for second year recruits than for first year men. There are standard multipliers for performance under a variety of conditions which vary from the base case or else there are specific values stated for each special condition in the book of norms. For example, norms would increase under the following conditions: during limited visibility, in extreme environmental conditions, in rough terrain, and so forth. Thus, the norms at night might be 50% more than during the daylight hours for deploying a gun in fighting position from a march.

There are three levels of acceptable performance: excellent, good, and passing. A military organizational unit receives an overall rating depending upon how its members perform on a variety of tasks.

Soviet tactical combat models might well include variables to account for human factors and group task performance based on this data. The readiness of Soviet units is regularly assessed. A simple assumption would be that combat performance is correlated with readiness measured against norms. The peacetime measurements could therefore serve as surrogates for the qualitative side of wartime combat performance. So far as we are aware, there is no direct evidence that Soviet combat models use such data, but it is an obvious first choice for that purpose.

RECOVERY OF EFFICIENCY FOLLOWING NUCLEAR ATTACK

Considerable Soviet research seems to have been devoted to the question of recovery of personnel efficiency following nuclear strikes. One example was given by General Tarakanov who discussed combat efficiency (boyesposobnost) and individual performance during recovery following nuclear exposure explosions. The graph is taken from Mathematics and Armed Combat (Moscow: Voyenizdat, 1974, pages 166-168 by Tarakanov. The book states that the researh was done by V.K. Korovkin.

The graph shows the upper and lower bounds on the percentage of combat personnel who are able to perform at a specific level of prestrike efficiency following a nuclear detonation. No numbers are given for the actual lengths of time to, t1, ... , tn.

The question of small unit cohesion under similar conditions is the subject of considerable Soviet literature. Barabanshchikov and Fedenko's, Psychology of the Military Collective, is an example. Neither of us have investigated this literature, but the subject may be of interest to others. See the references at the end.

BATTLEWORTHINESS

Unit breakpoints are a major consideration in many US models. The Soviets are obviously interested in similar questions. They appear under the term "boyesposobnost" which gets translated variously in different contexts as "combat readiness" or "battleworthiness." In the article about the term in the Soviet Military Encyclopedia (Volume 1, page 544, 1976), there is a statement that casualties and equipment losses lead to unit states of "partial" or "complete" loss of battleworthiness.

A regiment or division partially retains its battleworthiness when it takes personnel and equipment losses of 50-60% and control is not maintained. This is as close to a statement of what the Soviets consider to be a breakpoint as we have found in the Soviet literature.

Further, Robert G. Poirier and Albert Z. Conner in their book, The Red Army Order of Battle, have noted articles in Military History Journal which refer to "partially" battleworthy divisions in World War II operations which had between 300 and 1200 men. This would indicate that units with between 90% and 97% of TO&E strength were considered battleworthy when troop control was effective. There are numerous references of cases in World War II where units had taken large percentages of attrition over lengthy periods of time, but had remained battleworthy despite strengths well below what is normally considered the range of breakpoints. But part of this is a function of the time duration over which the losses are taken.

RECOVERY OF EFFICIENCY FOLLOWING NUCLEAR STRIKES

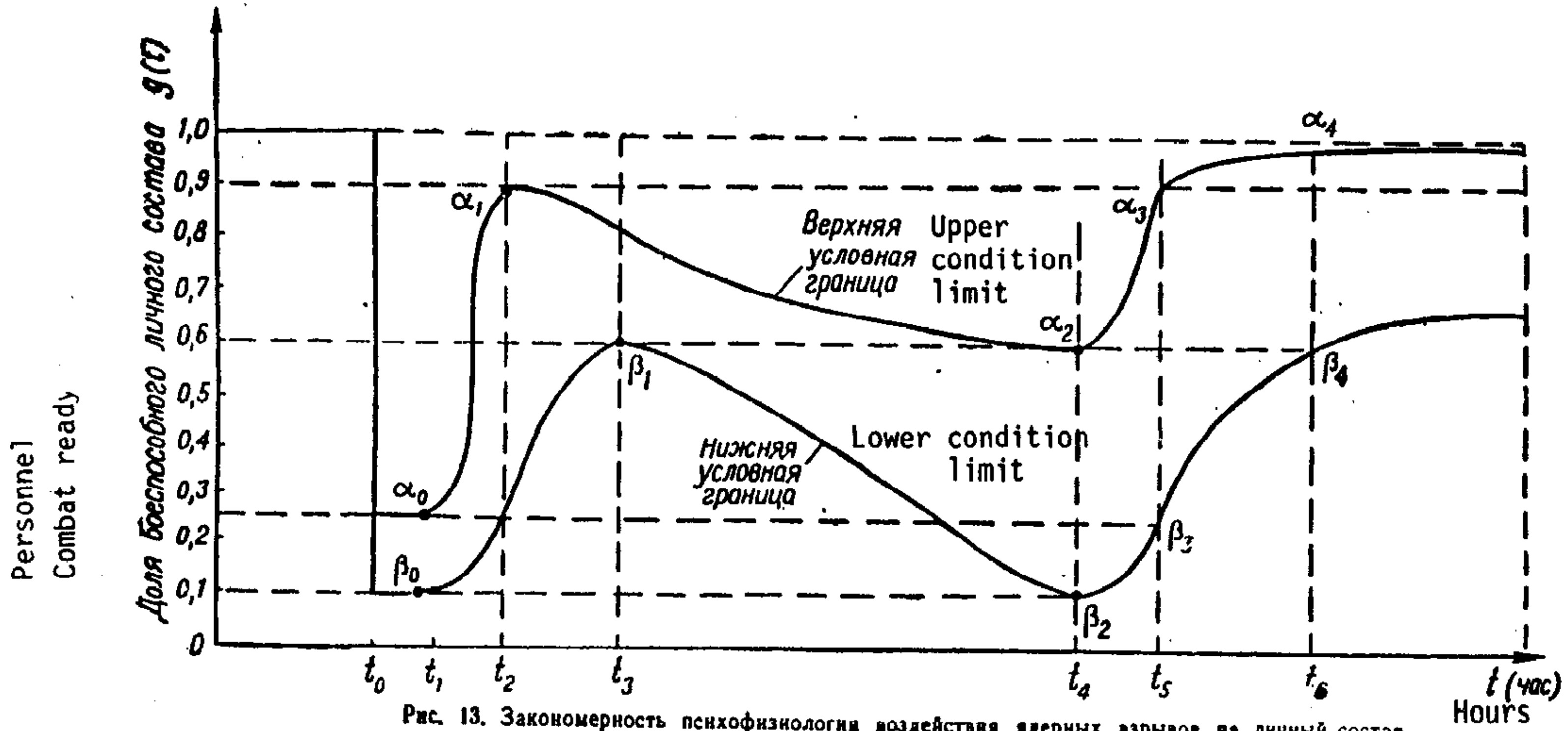


Рис. 13. Закономерность психофизиологии воздействия ядерных взрывов на личный состав (графики составлены на основе исследований В. К. Коровкина)

<u>MATHEMATICS AND ARMED COMBAT</u> (MOSCOW: VAYENIZDAT, 1974, PAGES 166-168)

GEN-MAJOR K. V. TARAKANOV: PERSONNEL EFFICIENCY IN NUCLEAR RECOVERY CONDITIONS.

Lieutenant Colonel Zavatsky, in an article in Military History Journal, examines losses of tanks and the personnel of tank corps and brigades, and the means of restoring their battleworthiness.

"The experience of the last war shows that tank (mechanized) corps, as a result of the losses suffered by the 8th to 10th day of fighting in the operational depth of the enemy defenses, and on the 4th to 5th day in breaching enemy defenses, had exhausted their initial capabilities and were forced to conduct measures to recover battleworthiness; that is, restore disrupted command, establish battleworthy formations, make up losses in combat equipment, weapons, personnel and replensih expended supplies of material."

A variety of methods were used to reestablish losses. Losses in corps staffs, for example, were sometimes made up by taking personnel from the brigade staffs. In another operation, following loss of tank corps headquarters personnel, subordinate tank units were transferred to a rifle corps for use as close support tanks instead of for independent tank operations as part of the corps. At times, composite units were formed from the remnants of partially destroyed units. One case cited involves forming a composite unit of tanks, infantry, anti-tank crews, mortars, and cannons. In the middle years of the war, tank corps maintained replacement pools of tanks, rather than simply relying upon repair of damaged tanks.

This article is a good example of the Soviet penchant for drawing upon historical data for an understanding of operational factors in combat. The books by Tarakanov and by Druzhinin and Kontorov both contain several examples where combat models have been used to examine historical battles, and they employ data from articles similar to Colonel Zavatsky's.

It is possible that Soviet modelers use historical data to assist in the verification/validation problem for combat models. Often, however, the combat model is assumed to be valid, and the analysis is intended to help explore the historical situation by throwing light on how alternative courses of action might have turned out. Zavatsky's article concludes with a note that implies the purpose of his paper was to provide a study of experience for effective training of troops under present day conditions. That is, the purpose was trianing and not specifically to provide data for modelers.

TROOP CONTROL

Another topic which merits a paper in itself is Soviet concern with stability and continuity of troop control, which is closely related to command and control in our terminology. There are several discussions related to human factors in Althukhov (1982). SAIC's FSRC has written extensively summarizing Soviet research on this subject and we will not attempt to repeat it here.

Druzhinin and Kontorov mention research and experimentation intended to enhance compatible decisions between superior and subordinate levels of command who are without benefit of communiction. This amounts to training subordinates to make decisions in keeping with the superior's mission and style.

MISCELLANEOUS

Druzhinin and Kontorov mention an interesting example of a human factor connected with pilot willingness to rely on their own skills versus those of their maintenance crews.

"Sociological studies have shown that during World War II pilots did not shrink from air battles, knowing that the probability of success was only 0.3 (this figure was known to them from analyses of previous battles); they would go to almost certain death, if in addition to this, appreciable losses were inflicted on the enemy; nevertheless they shunned flying aircraft the rated operational reliability of which reached 0.9 (a high figure, but it does not guarantee safety)."

MILITARY SOCIOLOGY

Starting in the mid-1960s a group of officers led by General-Major Konoplev began to create a place for military sociology with the Soviet military sciences. Their work was discussed in Adelphi Paper 76 written by David Holloway. The questions that they addressed and their research apparently threatened to overlap and possibly replace work done by the party political officer in the Soviet military. For this reason the introduction of military sociology was not always smooth. The research examined issues such as why young men choose the military for a career, how a military career compared with a variety of civilian careers in terms of attractiveness to school children, and the question of the psychology of the military collective.

Holloway reports on a Soviet survey taken which showed that the military ranked 26th out of 80 in desirability as ranked by Lenigrad school children. Most popular were careers like "research physicist" because the factors determined by the research to affect career desirability were, in order of importance: opportunity for being creative, growth potential, social status, and earnings. The Soviet Union was pushing scientific education as much as possible at that time and continues to stress it. Out of 40 professions the military had the following rankings:

Creative	40
Growth	15
Social Status	25
Earnings	8

Holloway concluded the Soviet research indicated their concern about motivation for a military career, and anxiety about young people's attitude towards the profession. It would be interesting to know what similar research would show today following the war in Afghanistan.

Holloway provides more detail about Soviet military sociology, and its growth over time within established organizations. Around 1969 General Konoplev was appointed as the Director of the Department of Military-Sociological Research within the Main Political Administration, which is the Communist Party organization within the military. We believe that Konoplev was later given a newly created position for sociology within the Ministry of Defense in the 1970s or early 1980s.

MORIMOC AND MILITARY HISTORY

Finally, the Soviets have made a concerted effort to exploit their military historical archives to help put more operational realism in combat models. Statements can be, found again and again that norms are based upon historical experience of the Great Patriotic War, often supplemented by field exercise data which is required where technology and equipment have changed. Examples include:

1. Suppression effects of artillery -- criteria for neutralization and annihilation
2. Massing required to breakthrough
3. Reconstitution of staffs
4. Effects of loss of command and control
5. Supply and ammunition consumption factors
6. Lanchester attrition
7. Basis for norms, correlation of forces and means criteria; tactical densities, etc.

There are several papers available on Soviet uses of military history so we will simply conclude with the comment that Soviet concern with human factors appears to be increasing.

REFERENCES AND BIBLIOGRAPHAY

We have not been able to find translations of several of the following. From skimming tables of contents we believe they may be of interest to participants at this conference who may wish to request translations.

[1] Altukhov, Gen.-Col. P.K. (Editor): *Osnovy Teorii Upravleniye Voysk [Fundamentals of Troop Control].* Moscow: Voyenizdat, 1984. Translation: JPRS-UMA-84-022-L. There is also a second English translation. Chapter 4, section 3 covers measuring the effectivenes of control systems.

[2] Barabanshchikov, A.V., Pistonov, K.K., Fedenko, N.F. of the Military Political Academy named for Lenin, "Military Psychology in the USSR," *Questions of Psychology*, No. 6, November/December, 1967, pages 76-84.

[3] Barabanshchikov, A.V., Glotochkin, A.D., Fedenko, N.F., Shellyag V.V.: *Problemy Psikhologii Voinskogo Kollektiva [Problems of the Psychology of the Military Collective].* Moscow: Voyenizdat, 1973.

[4] Barabanshchikov, A.V. Fedenko, N.F. (editors): *Osnovy Voyennoy Psikhologii i Pedagogiki [Fundamentals of Military Psychology and Pedagogy]*, second edition Moscow: Voyenizdat, 1981.

[5] Danchenko, Col. A.M., Vydrin, Col. I.F. (editors): *Voyennoy Pedagogika [Military Pedagogy].* Moscow: Voyenizdat, 1972. Translation: USAF Soviet Military Thought No. 7, 1976.

[6] Druzhinin, Gen.-Col. V.V., Kontorov, Col. D.S.: *Problems of Military Systems Technology.* Moscow: Voyenizdat, 1976. Translation: source not indicated.

[7] D'yachenko, M.I., Osipenkov, Ye.F., Merzlyak, L.Ye.: *Psikhologo-Pedagogicheskiye Osnovy Deyatel'nosti Komandira [Psychologico-Pedagogical Bases of Activity of a Commander].* Moscow: Voyenizdat,

[8] Holloway, David: *Technology, Management and the Soviet Military Establishment*, Adelphi Paper Number 76, London: The Institute for Strategic Studies, 1971. Contains a good discussion of Soviet military sociology shortly after it first emerged in the mid-1960s a a serious separate discipline within the Soviet military establishment. See the postscript section 2 in addition to the section in the main paper.

[9] Korobeynikov, Col. M.P.: *Contemporary Combat and Problems of Psychology.* Moscow: Voyenizdat, 1972. Translation: JPRS 57,726.

[10] Kovalev, V.N.: *Sotsialisticheskiy Voinskiy Kollektiv [Socialist Military Collectives].* Moscow: Voyenizdat, 1980.

[11] Lutskov, Col. V.N. *Metody Obucheniya Sovetskikh Voinov [Methods of Training Soviet Warriors].* Moscow: Voyenizdat, 1970.

[12] Malopurin, I.I.: *Psikhologicheskiye Osnovy Obucheniya Voinov Sukhoputnykh Voysk [Psychological Bases of Training of Troops of Ground Forces.* Moscow: Voyenizdat, 1987.

[13] Pavlov, I.P.: *Conditioned Reflexes, An Investigation of the Physiological Activity of the Cerebral Cortex.* New York: Dover Publications, Inc., 1960.

[14] Poirier, R.G., Conner, R.Z. *The Red Army Order of Battle,* Presidio Press, 1985.

[15] Podolyak, Col. Ya.V.: *Prakticheskiye Voprosy Voyennoy Psikhologii [Practical Questions of Military Psychology].* Moscow: Voyenizdat, 1987.

[16] Shelyag, V.V., Glotochkin, A.D., Platonov, K.K. (editors): *Voyennaya Psikhologiya [Military Psychology].* Moscow: Voyenizdat, 1972. Translation: USAF Soviet Military Thought No. 8, 1976.

[17] Tarakanov, Gen.-Maj. K.V.: *Matematika i Vooruzheniya Bor'ba [Mathematics and Armed Combat].* Moscow: Voyenizdat, 1974. Translation: FTD-ID(RS)T-0577-79.

[18] Zavatskiy, Lt. Col. Yu.I.: "Restoring Battleworthiness of Tank (Mechanized) Units and Formations in the Course of Offensives from the Experience of the Great Patriotic War," *Military History Journal*, October, No. 10, 1986, pages 73-77. There is a JPRS translation.

ANALOGIES IN HUMAN PERFORMANCE BETWEEN NUCLEAR PLANT ACCIDENT MANAGEMENT AND MILITARY COMBAT

by

Ed M. Dougherty, Jr.
Science Applications International Corp.
708 S. Illinois Ave., E101
Oak Ridge TN 37830

Joseph R. Fragola
Science Applications International Corp.
342 Madison Ave., Suite 1100
New York NY 10173

Abstract. The management of a nuclear plant accident can be assessed using stochastic–time situational modeling that arose in the development of probabilistic analysis and human reliability analysis of nuclear plants. The human performance postulated to occur during accident management has several analogies to combat and this modeling approach may carry over into this arena. These analogies are described and the techniques of time reliability correlations are demonstrated as a stride toward technology interchange between the two modeling areas.

INTRODUCTION

The perceived risks from commercial and government nuclear power or production plants can soar into the hundreds of thousands of immediate or latent fatalities. Of course, no such massive numbers of deaths have ever occurred in an industry that is forty years old. But there are risks—both economic as Three Mile Island can attest to, and lifethreatening, as the victims of Chernobyl would attest to. The plant operators, their advisors and management play critical roles in preventing theoretical risk from becoming real accidents.

While human factors analysis was developed elsewhere, the nuclear industry has developed—*almost from scratch*—the discipline called human reliability analysis, designated as HRA. One of its founders, Dr. Alan D. Swain[1] made the analogy that combating a loss of coolant accident, the major design basis of a commercial plant, was like combating the stress of war. Although this analogy stretches the truth somewhat, there are similarities.

Human performance in nuclear power plant accident management is characterized by on–the–spot decision making under stressful conditions, and decisions once made are carried out by a distributed group of individuals operating in potentially hazardous environments. In accident situations, the perceived need for quick action to prevent financial loss and possible loss of lives and the perception of personal risk exists. Of course, that risk perception may not be as strong as that in a combat situation [except maybe in the most extreme, conceivable circumstances]. The trauma associated with decision making is not new research. It has been studied extensively, and studies such as those of Janis & Mann[2] suggest that all human deciding is traumatic even without extreme stress. However, decision making has yet to be extensively studied in the nuclear plant setting, and the little available research has focused on actions taken within the main control room. The nuclear industry is just now beginning to investigate the hazards associated with actions outside the control room, where, at least during normal operations, as much as 65% of the human errors occur.[3]

This conference has an unique opportunity to ask whether human performance in combat and human performance in accident management is analogous enough so that a technical interchange may reward both areas. To begin this dialogue, this paper will discuss general areas in nuclear HRA and accident management that seem ripe for analogy to combat modeling. Proposing answers to your concerns would seem premature; the nuclear industry has enough trouble coming up with its own answers. However, the exchange of information may be of assistance in redirecting or revitalizing research on common problems and might provide each industry with a fresh perspective.

ANALOGIES IN HUMAN PERFORMANCE

Figure 1 depicts the situational analogy between accident management and combat. There are also other areas that characterize human performance in accident management that may carry over to military settings, including:

the probabilistic nature of human performance analysis and the role of time
the "simulator game" and other uncertainties in data application
the stochastic element of decisional conflict and diagnostic confusion
burden vs. workload as a concept
the stochastic human factors of remote activities
the need to accommodate a myriad of human factors
the uncertainty of a model—the subjectivist stance

In particular, Table 1 shows three areas where the analogies are particularly suggestive. These and the other analogous areas are discussed below.

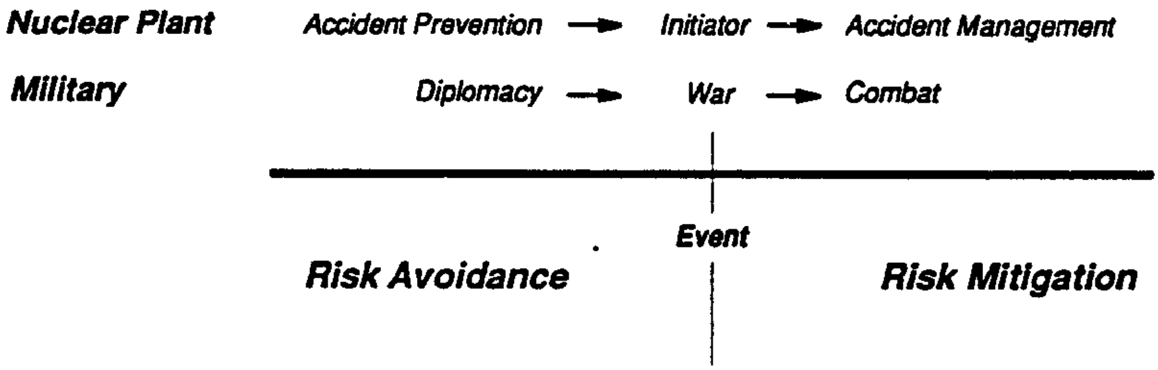


Figure 1. The situational analogy between accident management and combat.

Table 1
Areas of Analogy

	Accident Management	*Combat*
Probabilistic Nature	uncertainty of phenomena uncertainty of people & equipment	uncertainty of tactics & logistics uncertainty of people & equipment
Simulators	sparseness of data the simulator game	desire for limited engagements fidelity to combat conditions
Distributed Decision Making & C^3	manager vs. operator in–CR vs. ex–CR	general vs. field leader HQ vs. field

PROBABILISTIC NATURE OF ANALYSIS

It is the nature of the world to be risky. Since the concept of risk assumes the concept of uncertainty, risk has been mathematically modeled using probability theory. In the mid-1970s, the nuclear industry developed an analytical tool called probabilistic risk assessment [PRA—which is strictly speaking redundant, since risk *is* probabilistic] and is now about to require the PRA of all of its commercial plants over the next few years. Probabilistic analysis also makes explicit what other analyses may hide, namely, that all models contain various kinds of uncertainties and have degrees of fuzziness in their results. Our knowledge is only as good as our models are correct.

The risk from a nuclear plant arises from sequences of events, all of which must occur to result in some undesired consequence. The assessing of risk therefore is performed by the modeling of accident sequences. HRA is tasked with assuring that these sequences make sense operationally and from a human factors engineering [HFE] perspective, in particular, by including and modeling human failure events.

The mission of accident management is to win a "race" between time-consuming recovery activities which are required to mitigate a plant upset, or initiator, and the time-evolving severity of undesirable plant conditions. This race necessarily leads to a *stochastic-time situational modeling* approach. The mathematical characteristics of general probabilisitic modeling are well-known[4] and will not be included here. For example, the performance index of interest in accident management can be formalized as:

$$\Pr[\text{undesired consequences}] = \int_{t}^{\infty} f(t)\,dt \qquad (1)$$

where Pr[...] represents th probability of ... and the density, f, represents the probability over time of winning the race described above. However, even though the mathematical development is mature, the application of probabilistic modeling to human performance is only beginning to be generally available.[5]

Traditional human factors uses a timeline concept as one of its various task analysis tools. Its purpose is to assure that the assessed time required to respond in a certain way meets the specified available time. The task analysis subdivides the tasks, among other things, in order to allow subject-matter experts to reliably estimate times with which they are familiar. The sum of these times is the task response time needed for the task. Then, the claim is typically made that if the estimated required time is less than the available time, the task is "doable", i.e., can reasonably be allocated as a function to people, whereas if the required time exceeds the available time, automation of the task or redesign of the situation must be opted.

It is undoubtably true that if an estimate of the required response time is less than the specified available time, then the task should be re-examined at the least. However, the subtask time estimates are by no means certain—they are estimates, often made outside the estimator's experience base. It may turn out, therefore, that although the estimated response time meets the specification, it may not do so with any confidence. To demonstrate this, take the simplest probability distribution, the exponential density. Suppose the estimated response time is 5 min for an exponentially distributed time and the specified time is also 5 min. Then,

$$\begin{aligned}\Pr[\text{the task takes no more than 5 min}] &= \int_0^5 \frac{1}{5} e^{-t/5}\,dt \\ &= 1 - e^{-1} \\ &= 0.63\,.\end{aligned}$$

This means that there is only a sixty-three percent confidence that the task will meet the specified time, which is not much assurance. If, instead, you wanted 95% confidence, then either (1) the response time would have to be reduced to *less than 2 min*, or (2) the specified time would have to be relaxed to about *15 min*. Thus, when the probabilistic aspects of human factors analysis is factored in, the results may vary from those of the traditional approach.

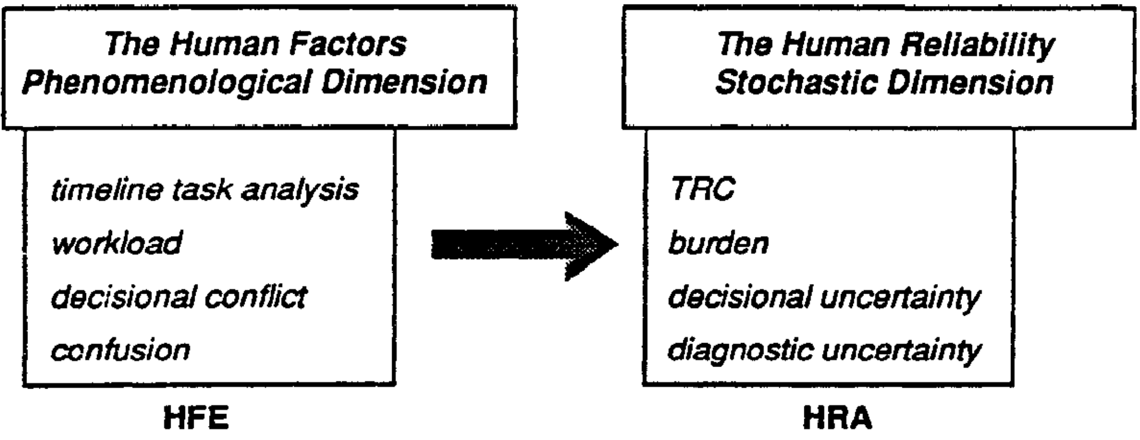


Figure 2. The transformation of "deterministic" phenomena into stochastic models.

Human reliability analysis already accommodates probabilistic analysis and therefore extends traditional human factors analysis. Figure 2 shows several human factors concepts and their extension using probabilistic concepts into HRA. Each of the other three factors will be discussed later. The point is that probabilistic analysis extends human factors analysis in an intuitively plausible way and takes advantage of the mathematical formalism of stochastic–time situational modeling.

TRCs as Probabilistic Human Performance Indicators .

Thus, to extend human factors analysis to HRA, a human reliability distribution, analogous to f in (1) is needed. Such a concept has been developed extensively in the area of accident management human reliability. The concept used is a *time reliability correlation* [TRC].

As in the previously described human factors approach, a response time is assumed to exist. This time is the stochastic random variable that represents the time required for an accident management *crew* of people to successfully respond to a specified situation. Because this time is stochastic, it has a characteristic density, f, so that:

$$\text{Pr[successful response takes at least t min]} = \int_t^{\infty} f(t)\,dt \qquad (2)$$

Note that the event described—"successful response takes at least t min"—is a *failure* when the specified time available is t or less. So (2) is an indicator of unacceptable performance and merely translates (1) into a human perfomance setting. The mathematical form of (2) is a cumulative complementary distribution function [CCDF], which in HRA is called a TRC. It presumes that a *success* density, f, can be determined and that the density has reliability parameters, e.g, a median value, m, and an error factor, ef, that can be determined from data or from theory of human perfomance [see Figure 3]. This is where simulators can be used.

SIMULATORS AS A SOURCE OF DATA

The Nuclear Regulatory Commission [NRC] began the use of simulators to obtain human performance data in order to make regulatory specifications according to probabilistic criteria.[6,7] Time–dependent data was plotted on *log–probability paper* and the resulting point distributions were noted to be sufficiently close to straight lines. This meant that simulator responses could be assumed to be *lognormally* distributed. The result, although not defined as such until later,[8] was a lognormal family of TRCs. Exponential and Weibull densities have also been popular in basing TRCs. Figure 4 shows that the choice of underlying density makes a difference in what is called the solution rate, analogous to the hazard rate for hardware failure densities. A peaked solution rate has been conjectured to have human performance basis,[9] showing a transition from more reliable rule–based behavior to the slower knowledge–based behavior.

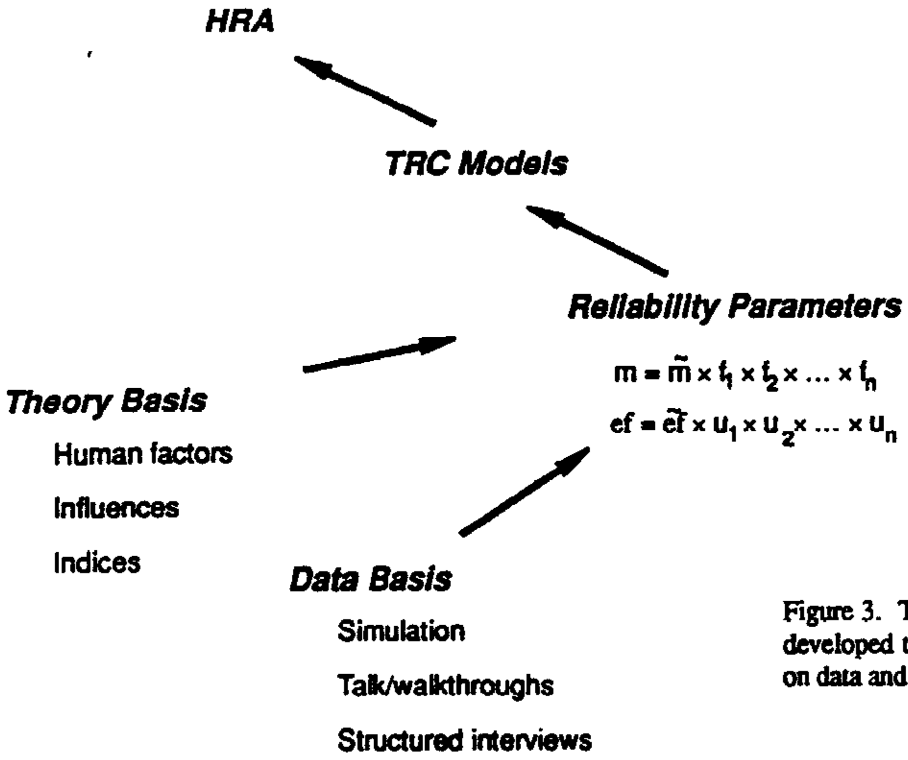


Figure 3. Time reliability correlations were developed that accommodated insight based on data and theory.

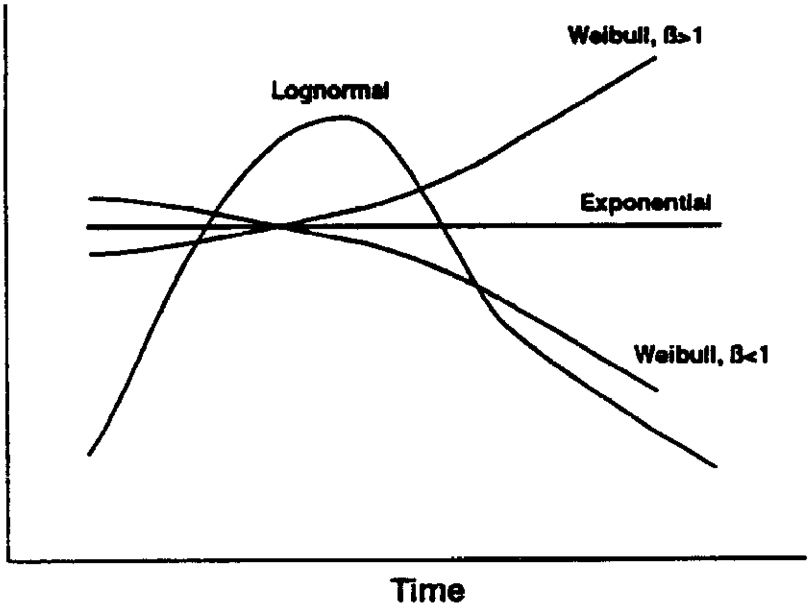


Figure 4. There were even behavioral reasons to choose the lognormal distribution family.

Potentially of more importance was the appearance of clustering in the simulator data taken[8] [see Figure 5]. Although simulator data remains frustratingly sparse, there seemed to be some evidence of TRC clustering which might correlate to the type of behavior dominating the response. The Hollnagel categories[10] were used to describe these reliability regimes—skilled behavior, rule–based behavior, and cognitive [or knowledge–based] behavior. These categories are rather controversial and later data collected do not necessarily support them.[11] But the clustering of TRCs according to behavior is too suggestive to be dropped as yet.

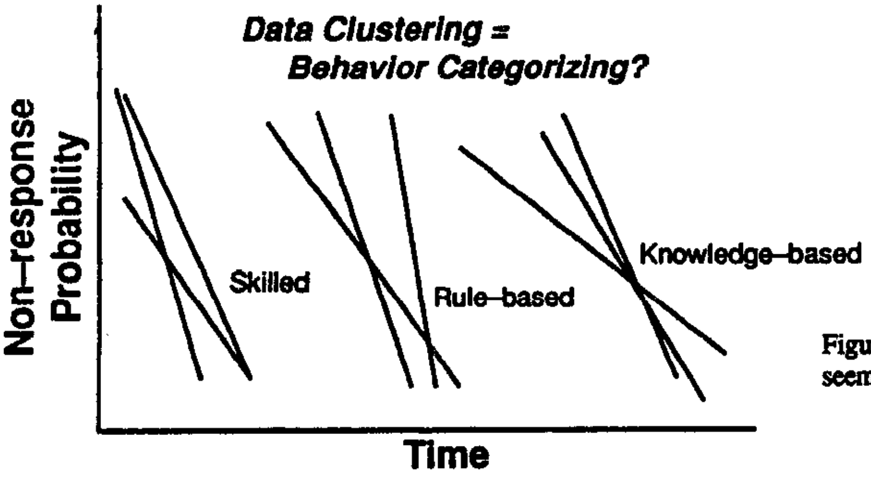


Figure 5. Simulator data, although sparse, seemed clustered.

Another pattern in simulator data was that some TRCs seem to be highly vertical and others quite horizontal. The lognormal density is characterized by two reliability parameters—its median response time, m, and a measure of the dispersion around the median, the so–called error factor, ef. [A random variable has a lognormal density if its logarithm has a normal density.] The greater m becomes, the farther to the right in Figure 6 the TRC is shifted. There was some evidence that the slower, i.e., right–shifted, TRCs represented greater difficulty or uncertainty in diagnosis that accompanied the response. Similarly, a greater ef indicated a more horizontally sloped TRC and this seemed to correlate to a difficulty or conflict in decision making.

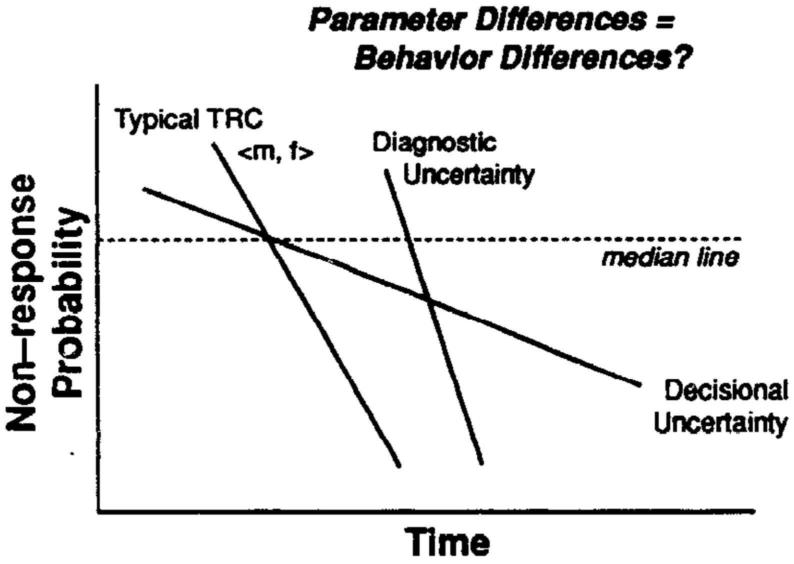


Figure 6. Differences in the basic reliability parameters derived from simulator data seemed a possible source to reflect behavioral differences.

Based on this admittedly speculative analysis, a TRC system was developed that accounts for the probabilistic considerations in HRA of:

1. available time,
2. decisional uncertainty,
3. diagnostic uncertainty,
4. any human factor influence that can be accounted for in a success likelihood index[12], and
5. model uncertainty.

The details are fully documented elsewhere,[5] but the most distinctive characteristics are described next.

Before going to other technical topics, however, a phenomenon that may be peculiar only to nuclear plant simulators must be described. This is what may be called the "simulator game". The NRC has recently required each commercial nuclear power plant to install a full–fidelity control room simulator that is capable of running postulated, design basis accidents. These simulators are primarily used to qualify and re–qualify operators on an annual basis. However, the nuclear utilities are beginning to see other training uses.

One crucial problem with nuclear plant operation is that it has two major, occasionally conflicting, goals. One is the production of electric power, which is the economic goal of a reactor. The other is safety to the public, which is a goal imposed by the realities of radiation and its dispersive, dangerous impact on human bodies. Safety takes priority only when challenged. As a result, postulated simulations can occasionally produce situations in which the operator would have to intelligently balance the power goal with the safety goal. Since the NRC regulates safety only, the operators can choose to "err" in favor of the safest actions. They won't ever fail an NRC test opting this strategy. However, if questioned about an action taken in a simulator that clearly violates an optimal tradeoff of safety and productivity, the operator—off the record—may say something like—"Well, of course, I wouldn't do that in the plant". This is the "simulator game".

The issue is that no matter how high the cognitive fidelity of a simulator in the area of diagnosis, a power plant simulator cannot simulate having a $4 billion plant under the operators' feet. Decision making uncertainty remains high. This may be a technical area in which military experience may be helpful in modeling plant operation.

DECISION MAKING AND DIAGNOSIS

Decision making during a nuclear plant accident is assisted by procedures and other tools that formalize the anticipations of engineers and analysts of plant performance during upset conditions. The "game" of accident management is predominately against the world, e.g., the physical phenomena that evolve because of lost core cooling. In principle, the behavior of the opponent can be completely known—the random factor of the impact of the human interactions that comprise the management strategy notwithstanding. In combat, however, the opponent is just as wilely and motivated as the game player. In comparison, the laws of physics may seem tame.

In an accident situation, decisions must be made within the following environment:

1. A plant consists of some 60,000 or more components, any of which may be failed and needed at the time of the accident.
2. The status of most of the critical components is instrumented in the main control room.
3. Instrumentation of performance parameters, however, is often indirect.
4. Operators are taught explicit plant responses to numerous and various offnormal conditions.
5. There are alternatives to almost all recovery actions.
6. These alternatives have a priority established by design engineers.
7. These alternatives have a priority as inferred by operators from their experiences.

Item 1 shows why the diagnostic role of accident management is so complex. Item 2 points to the problem associated with trading off omniscience and simplicity. Item 3 indicates one reason why operators, even under the guidance of the best procedures, must still use interpretative skills in accident management. Item 4 indicates the NRC influence on operator training and suggests how misdiagnoses may arise because of faulty anticipations. Item 5 again both indicates the positive influence of having options and the negative influence of complexity. Item 6 indicates that operational bases of a plant do not initially come from those who will operate the plant. Item 7 indicates that operators are indeed people and establish their own criteria in decision making. This may have positive effects or negative, but introduces a fundamental uncertainty into the analysis of human performance during accident management.

It is known from research in cognitive psychology and from analysis of actual incidents that people do what their belief structures direct them to do. Beliefs arise from training but also from idiosyncratic cognitive styles and experiences that are unique to individuals. Thus on the one hand, it is in principle possible to describe with high reliability what a person will do when his or her knowledge–base is known. On the other hand, specifying a person's knowledge base completely is in practice impossible. So, although a sophisticated cognitive modeling environment has been developed that can be coupled to a plant simulator in order to model the total person–machine system,[13] building but the most simplistic cognitive models is still not acheivable. Thus, diagnosis and decision making will always have a stochastic aspect if only because of our ignorance in the human sciences.

This is not to say, however, that the stochastic–time situational modeling approach has no recourse in accounting for the uncertainties associated with C^3–like behavior. Expert judgments can be elicited in structured ways so as to provide a cognitive index [as yet only a developmental idea] that can be introduced under desired constraint as a factor that adjusts the reliability parameters of the TRC. For example, in the SAIC TRC system, when a situation is covered extensively by procedures so as to make the diagnosis of the response rule–based, rather than knowledge–based, the median response time is halved. When a situation is considered to involve a decisional conflict, so as to increase the decisional uncertainty, the error factor is doubled, increasing the uncertainty factor.[5] These are to–date only arbitrary adjustments that could be supported by simulator and human factors analysis data. Figure 7 shows the effects of halving the median and doubling the error factor on a lognormally distributed TRC.

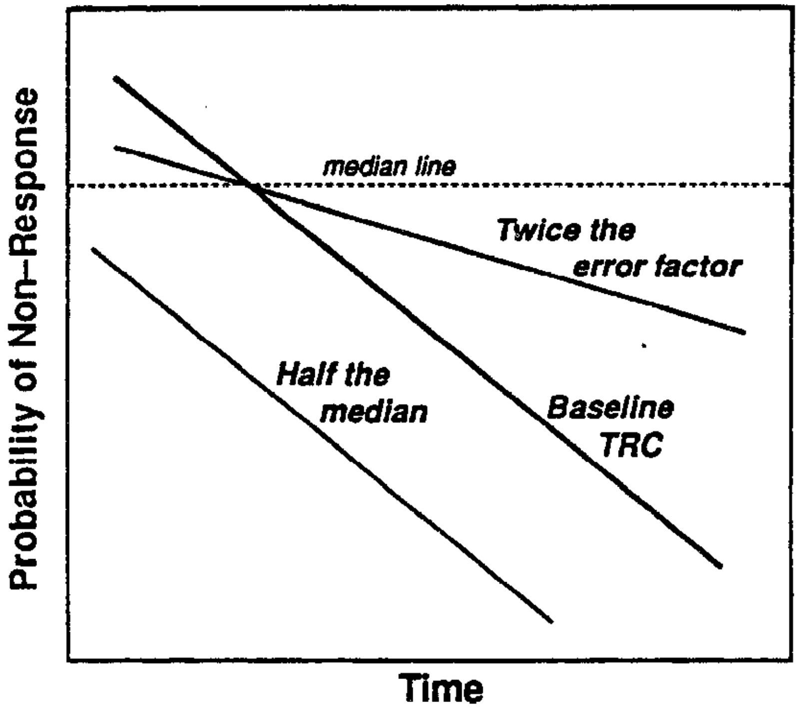


Figure 7. The effects on a baseline TRC when the median is halved and the error factor is doubled for human performance theoretical reasons.

As a last example, the timeline of diagnostic and decisional cues in an accident can provide another time–dependent measure of human reliability. The new emergency procedures required in all commercial nuclear power plants provide a redundancy to operator diagnosis in the following manner. The time frame of a nuclear plant accident is typically "forgiving". An accident must evolve, from a minor plant upset to more and more severely degraded conditions [this may be analogous to the buildup of military tensions prior to a war more than combat itself]. As conditions worsen, new cues arrive in a control room from the plant instrumentation. Thus, the operating crew has more than a single shot at successful diagnosis, although the later the correct diagnosis is made, the less likely the recovery may succeed.

Suppose an initial cue arrives at time, $t = 0$. Suppose further a second indication of impending undesired consequences arises at time, $t = T$. The crew may act on the initial cue, but if not, may still act on the second cue, if it is sufficiently leading. If the response density for the first cue is f_x and the second is f_y, then the time to successful response is the minimum of the two. Mathematically, the formulation for the successful response has a TRC of:

Situation with
Redundant Cues

$$\mathrm{TRC}[\,t\,] = \begin{cases} \mathrm{TRC_X}[\,t\,] & \text{when } t \leq T \\ \mathrm{TRC_X}[\,t\,] \times \mathrm{TRC_Y}[\,t-T\,] & \text{when } t \geq T \end{cases} \qquad (3)$$

In other words, the second TRC has the effect of a redundancy, i.e., another probability that lowers the first failure probability. Also, the "clock" on the second cue does not start until the cue arrives. Thus, although a redundancy factor, the second term is conditional on the timing of the situation and is not as large as a factor were the times assumed to be independent.

BURDEN VS. WORKLOAD

Military scientists are familiar with the concept of workload, both physical and mental. The transformation of this concept to a stochastic–time situational model results in what the nuclear plant industry is now calling burden.[14] Research into workload has historically had a problem attaching quantitative indicators to this sound psychological phenomenon. A typical measure was the number of tasks of some specified type per unit of time. In other words, workload is often considered the human analog to pace, the intrinsic rate of an ongoing event. Since experiments show that some people can reliably operate under quicker paces than others, it was decided that workload is a subjective indicator, i.e., depends on the subject, his skills, training, and knowledge [among almost any other influences].

Subjective indicators are resorted to in HRA commonly enough, but only as a last recourse. The stochastic framework of TRCs, however, allows a twist to be introduced to workload analysis that deterministic or subjectivist methods may miss. Table 2 assumes that a task consists of two or three subtasks, which must be performed in sequential order. The pace concept of workload will usually imply that the performance of three subtasks over a period of time is more workload than the performance of two. However, the subtasks may have different intrinsic response times. Then the sequential task response time has a density found by using what is called a convolution operator, and the resulting task TRC is:

Situation of Sequential Tasks

$$\mathrm{TRC}_{\mathrm{task}}[\,t\,] = \int_t^{\infty}\int_0^{t} f_X(t-x)\,g_Y(x)\,dx\,dt\,. \tag{4}$$

The convolution of three densities is found similarly.

For illustrative purposes, exponential densities are used in Table 2. The convolution of exponential densities has several useful properties, two of which are:

1. If one mean time is much longer than the others, then the TRC of the convolution acts approximately like the TRC of the single density with the largest mean.
2. As available time gets longer, the TRC of a convolution acts approximately like the TRC of the single density with the largest mean.

Note that the convolution of two or more exponential densities are not exponential densities, but are linear sums of the underlying densities.

Table 2 shows the difference between the concept of burden and the concept of workload. In the first case, the third density is just an additional task that must be completed and thus, the idea of worlkoad and burden agree—more tasks per time means less reliability. However, if the response to the additional task is quicker than the other two, the additional task adds to the measure of workload but does not substantially increase burden. This is the second case in Table 2. Finally, if the task consists of three fast subtasks vs. two slow ones, then the result, in the third case, is that the workload is greater [intuitively] for the task with three subtasks whereas the burden is *less* for the three subtasks, when stochastic times are considered.

Thus, again by solely considering the stochastic properties of human reliability, a human factors situation can be better modeled than traditionally. If indeed skilled behavior is "faster" than rule–based behavior, in a TRC sense, then training a task to the skill level can "cure" some problems associated with workload.

Table 3 shows that the choice of the type of density can be significant.[14] The lognormal density is characterized by two parameters, the median and error factor, and the exponential by one, the mean. To compare the effects of choosing between the density types, the lognormal density needs to be "translated" into an exponential density. One way to do

Table 2

The Variable Stochastic Effects of an Extra Task

Time, min	*2 TRCs*	*3 TRCs*
Just Adds a Third Task — <1, 1.5> vs. <1, 1.5, 2>		
5	0.09	0.4
10	0.004	0.04
30	0.000000006	0.000002
Adds a Fast Third Task — <1, 1.5> vs. <1, 1.5, 0.5>		
5	0.09	0.1
10	0.004	0.006
30	0.000000006	0.000000009
Three Fast Tasks vs. Two Slow Tasks — <1, 2> vs. <1, 1.5, 0.5>		
5	0.2	0.1
10	0.01	0.006
30	0.0000006	0.000000009

Note — the exponential model is used to ease the necessary calculations, with given mean response times

Model for two TRCs
with unequal means, m_i,

$$\mathrm{TRC}[t] = \sum_{i=1}^{2} \left[\frac{m_i}{m_i - m_{i-1}} e^{-t/m_i} \right]$$

Model for three TRCs
with unequal means, m_i,

$$\mathrm{TRC}[t] = \sum_{i=1}^{3} \left[\frac{m_i^2}{(m_i - m_{i-1})(m_i - m_{i+1})} e^{-t/m_i} \right]$$

this is to preserve the mean. Since the two second lognormal densities of Table 3 both have a mean of 5, a comparison of the resulting convolutions of TRCs will demonstrate the effects that a second characterizing parameter will have on the results. Table 3 shows that the high error factor, but low median curve, <1.88, 10> "hangs up"in probability as time increases, while the low error factor, high median curve, <4, 3> does not. [Note that these two densities have a common mean of 5.] This means that if median and error factor are used to reflect different human factors, then the lognormal family of densities is more sensitive than the exponential family to changes in these human factors and is more useful for this reason.

OTHER HUMAN FACTORS

Crew reliability may correlate to circadian effects—recently several plant crews were found asleep on the night shift [nuclear plant crews are rotated through the plant's shifts]. A recent plant trip was caused because a vital control was unprotected and a painter, painting the wall on which the control was located, inadvertently painted over the control switch. Several wrong train or wrong unit errors have occurred because of poor lighting or inadequate equipment labeling. The number of human factors that can contribute to undesired consequences of some significance is unlimited. Analogously to the MIL–STD approach to hardware reliability, an HRA practitioner typically assigns "nominal" reliability parameters to a situation and then adjusts these parameters by an index subjectively obtained that reflects the influences of the influential human factors out of the myriad of others.

Table 3
The Choice of TRC Type is Significant

Lognormal TRCs	(1) Convolution of <4, 3.2> and <4, 3> (2) Convolution of <4, 3.2> and <1.88, 10>	[mean 5.1] [mean 5.0]
Exponential TRCs	The convolution of TRC with mean 5.1 and with mean 5.0	

Time min	*Lognormal TRCs* <4, 3.2>*<4, 3>	*Lognormal TRCs* <4, 3.2>*<1.88, 10>	*Exponential TRCs*[1]
5	0.88	0.70	0.74
10	0.41	0.32	0.41
20	0.056	0.093	0.096
30	0.0093	0.041	0.019
60	0.00017	0.012	0.00009

---

[1] The mean is preserved when transforming the lognormal TRCs into exponential TRCs. Since the mean for the second pair of lognormal TRCs is 5.0 in both cases, there is only a single exponential convolution.

We have seen how this is done for a TRC system. However, the NRC is beginning to become more interested in "human factoring" the PRA and HRA process. Systems engineers invented the stochastic–time situational modeling framework that the TRC concept evolved into. However, this transfusion of the more traditional, and in particular, more mechanism–oriented perspective is overdue and welcomed. The mix of disciplines may even go so far as to fundamentally change the modeling framework. But even if not, human reliability requires an interdisciplinary approach.

WHERE TO GO NEXT?

Time will tell whether the apparent analogies between the human performance of accident management in nuclear plants and military combat can lead to a cross–fertilization of methods. However, it seems plausible that an interchange of technology can only benefit understanding in both areas, if only in the knowledge that different human environments require different human performance modeling.

The probabilistic aspects of stochastic–time situational modeling naturally extends traditional human performance and human factors analysis methods, while potentially solving some problems that the more deterministic analyses cannot. The details of the TRC approach are still only speculative, but thanks to the tight coupling of TRCs to simulators, they are at least in principle empirically testable. There will always be uncertainties associated with any modeling approach and this probabilistic framework can account for these stochastic sources as well.

REFERENCES

1. A. D. Swain & H. E. Guttmann, *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Applications*, NUREG/CR–1278, USNRC, Washington DC, August 1983.

2. I. L. Janis & L. Mann, *Decision Making*. NY: Free Press, 1977.

3. Virginia Power/Institute of Nuclear Power Operations, *Virginia Power Human Performance Evaluation System Program Report*, October 1987.

4. E. M. Dougherty, *Distribution Theory Applied to PRA*, SAIC-NY/OR-11/88-1, Science Applications International Corporation, November 1988.

5. E. M. Dougherty & J. R. Fragola, *Human Reliability Analysis: A Systems Engineering Approach with Applications to Nuclear Power Plants.* NY: John Wiley, Inc., 1988.

6. T. F. Bott, et al., *Criteria for Safety-Related NPP Operator Actions: Initial BWR Simulator Exercises*, NUREG/CR-2535, USNRC, Washington DC, November 1982.

7. A. N. Beare, et al., *Criteria for Safety-Related NPP Operator Actions: Initial PWR Simulator Exercises*, NUREG/CR-1908, USNRC, Washington DC, September 1981.

8. R. E. Hall, J. R. Fragola, & J. Wreathall, *Post Event Human Decision Errors: Operator Actions Trees/Time Reliability Correlation*, NUREG/CR-3010, USNRC, Washington DC, November 1982.

9. J. G. Wohl, "Maintainability Prediction Revised: Diagnostic Behavior, System Complexity, and Repair Time", *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-12, no. 3, May/June 1982, pp. 241-250.

10. E. Hollnagel, O. M. Pedersen, & J. Rasmussen, *Notes on Human Performance Analysis*, RISØ -M-2285, RISØ National Laboratory, Denmark, June 1981.

11. Private communication with D. Worledge, Electric Power Research Institute project manager of their simulator studies.

12. D. E. Embrey, *The Use of Performance Shaping Factors and Quantified Expert Judgment in the Evaluation of Human Reliability: An Initial Appraisal*, NUREG/CR-2986, USNRC, Washington DC, 1983.

13. D. D. Woods, et al., *Models of Cognitive Behavior in Nuclear Power Plant Personnel*, NUREG/CR-4532, USNRC, Washington DC, July 1986.

14. J. R. Fragola, "Identification of Potential Operator Burden Through Systematic Assessment of Actual Plant Transients, *IFAC Conference on 3rd Man-Machine Systems*, Oulu, Finland, June 14-16, 1988.

SESSION II: HUMAN PERFORMANCE MODELS AND APPLICATIONS

SESSION CHAIR: Michael Strub, Chief, US Army Research Institute Field Unit, Fort Bliss

The presentations in this session illustrate the major benefits to be gained from including human performance in combat models as well as significant obstacles to be overcome in linking human and combat models. The session begins with an overview of human performance models in which many models are examined which were not developed in the context of combat simulation. However, certain of these models such as the crew performance models might be used offline to generate human performance parameters which would become input to the combat models. It is clear from the review that recent developments in network modeling tools appear well suited to combat model applications. The presentation summarizes the findings of a NATO Research Study Group on "Modelling of Human Operator Behavior in Weapon Systems".

The second presentation on "A Human Performance Consultant System and Some Applications" describes the development of a human performance expert system from a FORTRAN module of an air defense model to a stand-alone PROLOG Human Performance Consultant System. The presentation describes the reasons for the system's evolution, as well as its applications in air defense and air-to-ground mission modeling.

The third presentation is an update on a major Department of the Army program called Combined Arms in a Nuclear and Chemical Environment (CANE). The CANE program was established to provide measured data and determine how well combat and support units can perform their missions in extended operations where nuclear and chemical weapons are employed. The presentation includes a short overview of the program, the results and lessons learned to date, and future directions to address key Army areas of concern.

The next three presentations deal with a network modeling tool called Micro SAINT. The presentation on "Task Network Modeling Constructs as Applied to Modeling Human Performance in Combat" provides an overview and discussion of the concept of task network modeling and includes near term research and development suggestions such as embedding task network models within some existing combat models and developing human performance shaping functions from existing human performance data. In the next presentation, Micro SAINT was used to link crew performance directly to system performance. Two models were developed and used over a wide range of critical task error rates to illustrate how average performance time may degrade very slowly in spite of errors. The final presentation in the session describes the development of Micro SAINT simulations of the Close-In Weapon System loading operation with a three man crew. It is a case study on the use of Micro SAINT to model a multi-man shipboard tactical operation. It includes major issues which surfaced in the attempt to model human behavior and performance.

AN OVERVIEW OF HUMAN PERFORMANCE MODELS AND POTENTIAL APPLICATIONS TO COMBAT SIMULATION

Grant R. McMillan
Armstrong Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, OH 45433-6573

Edward A. Martin
Aeronautical Systems Division
Wright-Patterson Air Force Base, OH 45433-6573

INTRODUCTION

Human factors engineering has evolved out of early work in applied experimental psychology through the design handbook era and is becoming more interdisciplinary with strong technological influences from computer and information science, operations research, and systems simulation and modelling. Human factors/ergonomics models have the potential for representing human performance in ways which are compatible with both operations research and systems engineering. Despite this, human factors models are not widely used, either by human engineering specialists, engineers, or systems designers.

In response to this, a Research Study Group (RSG) on "Modelling of Human Operator Behaviour in Weapon Systems" was established by Panel 8 of the NATO Defence Research Group to consolidate the available knowledge, to stimulate information exchange and cooperative research, to foster the practical application of modelling research, and to provide a bridge between models and approaches adopted by engineers and behavioral scientists.

Acting on these terms of reference, the RSG:

(1) reviewed current human performance models which are in use;
(2) investigated the development of micro-computer based models of human performance;
(3) conducted a technology demonstration workshop which included working demonstrations of typical models and technical papers on their application.

The presentations and discussions of the workshop are being published in a book [1] which provides an overview of the state of the art of human performance modelling. The review of current human performance models is being published as a NATO report [2] in the form of a directory. It will provide potential users with brief reviews of over 50 models reported in a standard format.

The work of the RSG focused on system design applications of models in the following seven categories: Task Allocation and Workload Prediction, Single Task Models, Multi-Task Models, Multi-Operator Models, Biomechanics and Workspace Design, Training and Skill Retention Models, and Network Modelling Tools. The scope was largely limited to models of operator and maintainer performance, with an emphasis on models that permit some type of computer-based simulation of the man-machine system.

Most of the models reviewed were normative. That is, they represent an ideal operator, or are based on a theory of what should happen. Many of the models could be manipulated to represent non-ideal performance, however, and as such might be applicable in combat effectiveness simulations. Some are also related to human response to operational/battle stress and fatigue, which have recently become a subject of interest to Panel 7 (Defence Applications of Operational Research) of the Defence Research Group. Such models contain terms that permit one to represent the difference between idealized and real operators, the effects of environmental and battle stressors, the differences between ideal and combat effectiveness measures, or the differences between individual and group performance.

This paper summarizes the findings of the RSG and suggests some ways in which human performance models might be utilized by operations research analysts. The following section provides an overview of the state of the art in each of the modelling areas addressed by the RSG. The last section provides several examples of how these tools might be used to enhance combat simulations.

THE STATE OF THE ART IN HUMAN PERFORMANCE MODELLING

Task Allocation and Workload Prediction

These tools address the problem of allocating functions to the human or to the machine. Since the goal of this process is to make the best use of the abilities of each of these "components", models which suggest optimal allocations would be most useful. However, no such models exist. As a result, this stage is usually an iterative cycle of function allocation ... task analysis ... and workload prediction to evaluate the designer's proposed solutions. The techniques are more akin to analysis aids, or structured analysis procedures than to true models, and are designed to assist the user in this evaluation process.

The RSG found only a single model which directly evaluates allocation of functions, and it has not had much use. Several models are available for analyzing/predicting human operator workload, and this continues to be a very active area of development. Essentially three approaches are currently being pursued.

The first approach, time line analysis, compares the time required to complete assigned tasks to the time available. This approach is well developed and commonly used in the design of new weapon systems [3]. Such techniques typically assume that operators perform their assigned tasks in a fixed serial sequence, or the techniques require separate analyses of alternative sequences.

Since it is well known that operators will alter the sequence of task performance to manage their workload, some current techniques are using modelling languages such as SAINT (Systems Analysis of Integrated Networks of Tasks) to simulate this probabilistic behavior [4]. This approach tends to produce a less rigid estimate of human capabilities.

A third approach is termed the attentional demand or multiple resource approach [5,6]. It recognizes the fact that some tasks can be accomplished essentially in parallel (low attentional demand), while other tasks must be done serially (high attentional demand). In effect, such models quantify workload in terms of attention required versus attentional resources available. They tend to give a more optimistic estimate of an operator's workload capacity.

Before workload can be computed, these approaches require a priori estimates of the time required to perform each task, or of the attentional demands of each task. These estimates may be obtained from experiments or from subject matter experts. The techniques appear to be useful tools in estimating the workload demands of various systems design options. The workload estimates must be considered relative rather than absolute values, and very little validation of the predictions has been conducted.

Single Task Models

These models allow behavioral predictions when an operator is performing an individual task, e.g. tracking a target, monitoring auditory signals, or making a discrete movement to a control. Because of the strong constraints of the task environment, these models contain some of the most formalized mathematical structures reviewed by the RSG. Many of these models are able to make precise, highly detailed predictions about the performance elements of their tasks. They typically allow predictions about the effects of molecular equipment characteristics (e.g. control gains, control spacing, display resolution, signal/noise ratios) on human performance. They are highly valid in their domain of application. On the other hand, most of these models have very limited domains. They are not general models of listening, of reading displays, or of movement.

Two of the more promising simple models are Fitts' Law for predicting discrete movement or target acquisition times [7], and an auditory detection model developed for the design of helicopter warning signals by Rood, Patterson, and Lower [8]. The well known manual control models, such as McRuer's Crossover Model [9] and the Optimal Control Model [10], have been validated for the prediction of tracking performance. They are being used for research and for the design/analysis of vehicle control systems. The manual control models have been used to model the effects of acceleration, vibration, control/display degradation, heat stress, and the encumbrance of chemical protective gear (See reference 11, for an example).

Multi-Task Models

In all realistic environments, humans have multiple task demands competing for their limited behavioral resources. Multi-task models attempt to predict human performance (not just workload) in these complex task environments. The majority of techniques reviewed by the RSG are models of display monitoring and decision making in a multi-task environment. Some include submodels for rather elaborate control actions based on decisions, while for others the control is essentially limited to selection from a limited set of established procedures.

Most of these models have been minimally evaluated for validity or generality. In experimental settings, some of these models have made predictions that would be useful to designers of supervisory control or command/control systems. For example, a monitoring and decision making model, DEMON [12], was used to analyze human control of multiple remotely piloted vehicles (RPVs). This application involved: (a) detection of whether any of the RPVs had exceeded desired navigation tolerances, (b) deciding which of the RPVs to monitor at a given time and whether to take corrective action, and (c) RPV control in terms of navigation correction, control hand-off to a terminal phase operator, or execution of a "pop-up". Although the model predictions were reasonable and interesting from a design perspective, no funding was available for validation.

The two techniques which seem to have received the most use in system design are single operator forms of the Siegel-Wolf model [13], and the Human Operator Simulator (HOS, reference 14). These applications have involved military problems such as carrier landing performance, air refueling, missile launch, and P-3 patrol aircraft sensor operation. The RSG found few readily usable models of human cognition and problem solving. This is largely because the available tools are conceptual/verbal models, have not been computerized, and have tended to remain in the academic and research communities.

Multi-Operator Models

This category of models adds a significant level of complexity: communication and interaction among operators. Despite this complexity, these models have almost all been applied to significant real-world design or evaluation problems. Some of them, such as multi-operator forms of the Siegel-Wolf model [2], have been highly successful in a number of applications. Most of them were developed under contracts to military agencies, and were designed to simulate the performance of crews performing specific military missions. Most of these models have an operations research flavor to them. They are not designed to address molecular equipment characteristics, but rather focus on procedural issues, task organization, task assignments among operators, required crew size, the effects of fatigue or time pressure, etc. In general, these models attempt to predict global system performance measures such as the probability of mission success or the time required to accomplish the mission. They almost never represent the mechanisms of human performance.

The models typically require the user to develop rather detailed descriptions of all tasks to be performed by the crew. This includes descriptions of task criticality, permissible task sequences, average task completion times and variances, and the assignment of tasks to individual crew members. Because of their task analytic nature, operator loading is the key element driving the predictions of these models. Loading is usually defined as the time required to complete all tasks versus the time available. In most cases, these models run in a Monte Carlo, or iterative, fashion to allow sufficient sampling from task performance distributions to make stable statistical predictions. Success or failure of the mission does not depend on the probability of accomplishing any single task, but on whether or not the crew completes all essential tasks in the required time. Thus, each task has an effect on mission success, but not necessarily a primary one.

Some of these models allow users to evaluate the effects of performance modulation factors such as fatigue and operator skill level. Typically, these effects are implemented by modifying the means and/or variances of task completion times. In some cases these functions must be provided by the user, while in others, such as the Siegel-Wolf model, generalized functions are provided.

While the user input requirements are significant, the data can often be obtained at a sufficient level of accuracy from subject matter experts. For complex systems, the resulting models are usually quite large and complex. These models are rarely, if ever, formally validated by comparing model predictions to the results from actual field trials. Rather, these models tend to be "exercised" and their results evaluated for reasonableness. If the results seem reasonable to the user, design or procedural decisions will often be based upon them.

Biomechanics and Workspace Design

This category is also rich in models developed specifically for system design applications, but at a molecular level. This review encompasses two principal categories of modelling:

(1) anthropometric models which predict the ability of an operator of a given physical size, to work within a given space, to reach specific controls, and to see specific displays

(2) biomechanical models which predict human materials handling capabilities

The anthropometric models are computerized versions of traditional drawing board, manikin, and mock-up approaches. As a result, they offer the ability to readily change workstation dimensions and characteristics, to represent individual operator body dimensions, and to evaluate the fit, reach, and vision envelopes of a wide range of human populations if appropriate data bases are available.

Anthropometric models have several common shortcomings. Most have only been validated over limited ranges of reaches and fits. Most permit only one or two operator postures, e.g. sitting erect or sitting slumped, and do not represent the effects of postural changes on reach envelopes. Effects of clothing or other restraints are often not modelled. Finally, most models are incompatible with other Computer Aided Design programs and systems. At the present time, they appear to be useful tools for experts in anthropometry, but can easily mislead the novice.

Biomechanical models which attempt to describe the human body as a mechanical, load bearing device have a long history. Much of the work in this area has focused on defining human tolerance limits to vibration and acceleration/deceleration stress. Such tolerance models were not reviewed by the RSG. Techniques to model performance effects of these stressors have also been developed, and are summarized in the McMillan, et al. report [2]. In many cases, these performance models are based on existing single task models with the vibration or acceleration stress represented as a disturbance to visual perception or motor control.

Another type of biomechanical model forms the bulk of the techniques reviewed. These models predict human lifting capacity in materials handling situations. Although this may seem like a restricted domain, reviews of physically demanding tasks in a variety of trades show that most involve lifting.

The limitations of these models are in some ways similar to those of the anthropometric models. Generally, they assume a limited range of lifting postures and geometries, no mechanical aids, smooth symmetrical lifts, good floor contact, and so on. Of course, these assumptions are violated in many operational settings. Nevertheless, these models do appear to be getting increased use.

Training and Skill Retention Models

This area of human performance modelling has a long and rich history. Much of the model building has been done by psychologists as part of their theory development and testing process. As a result, most of these models are of a qualitative, descriptive nature and were never really intended for system design applications. However, the increasing sophistication of systems in which man must perform has produced a growing demand to improve training devices and procedures, to forecast their effectiveness during the design process, and to quantitatively evaluate their effectiveness during the system life cycle.

Perhaps the earliest quantitative attempts to describe and predict learning involved fitting the training data with "fixed-form" equations. A variety of such equations have been utilized and their relative merits debated. Some success has been achieved in using such models to predict production rates in industrial settings as workers learn a new manufacturing procedure [15].

The qualitative models reviewed by the RSG address a broad range of issues. Some focus on the stages of skill acquisition and have been used in the overall organization and sequencing of training curricula. Some attempt to categorize human information processing skills and to develop general guidelines for training these skills. Others attempt to predict the effectiveness of proposed training devices on the basis of the difficulty of acquiring certain skills and the importance of these skills to operational tasks. These and other qualitative models appear to be increasingly used in the training system design process, and are a significant area of model development activity.

Network Modelling Tools

This category includes special computer languages developed for the purpose of simulating man-machine systems. There is an important tie between these tools and the Multi-Operator Models. Many of the latter models have been developed using simulation languages such as SAINT, SLAM (Simulation Language for Alternative Modelling), or Micro SAINT [2]. Our theoretical foundations are weak in the area of group/interactive behavior. As a result, it is difficult to develop formal mathematical models of crew performance. In general, the theoretical constructs are not available for developing equations that describe the performance of multi-operator systems.

On the other hand, such systems can often be simulated as task networks which represent the sequence of activities in the system. In setting up such a network, the user must specify for each task: (a) the predecessor tasks that must be completed before the task in question can begin, (b) the statistical characteristics of the task in question, and (c) the branching to other tasks to be performed upon task completion. The statistical characteristics of a task may include task duration distributions, task criticality, operator speed, operator accuracy, etc. Different operators may be responsible for different tasks, and probabilistic branching among tasks may be used to represent interactions among operators.

These modelling tools thus provide a flexible structure for simulating the performance of multi-task/operator systems. With this flexibility comes the requirement to provide a large body of data about the system's structure and task characteristics. In many cases, such data can be obtained by interviewing subject matter experts. Validation of such large-scale models is very difficult, and any specific model is not likely to generalize to other systems. However, the same constraints apply when one attempts to predict the performance of such systems using man-in-the-loop simulations.

RELEVANCE TO COMBAT SIMULATION

The need for including the effects of human performance in combat models was clearly expressed by Van Nostrand [16]. In that report, Van Nostrand first reviewed human factors and combat data to determine which performance effects are most likely to modify combat outcome. Second, she analyzed some representative combat models to determine where these kinds of human factors data could be incorporated. The results of this analysis are summarized in Table 1, which is taken directly from the report. The table divides the human factors data requirements into two broad categories: those for which data exist, and those for which none were found. Within these two categories, Van Nostrand notes where combat model pergorithms (personnel performance algorithms) are available to incorporate the effects.

TABLE 1. Human Performance Data Needs (From Van Nostrand, 1986)

SOME DATA EXIST, OR WILL IN THE NEAR FUTURE

1. SOME COMBAT MODELS HAVE APPROPRIATE PERGORITHMS:

RATE OF FIRE
PROBABILITY OF DETECTION
PROBABILITY OF HIT
IDENTIFICATION RANGE
VISIBILITY

2. PERGORITHMS DO NOT EXIST:

PERCENT WHO FIRE
TOXIC SUBSTANCE
COMBAT FATIGUE
SLEEP LOSS
HEAT
COLD
NEW GUY FACTOR
SUPPRESSION

3. WILL NOT CHANGE RESULTS:

FEW EFFECTS EXPECTED

HUMIDITY
NOISE
VIBRATION
CONFINEMENT
ALTITUDE
CROWDING
ISOLATION

DO NOT CHANGE DURING COMBAT

MENTAL ABILITY
% GNP TO MILITARY
GNP / CAPITA

NO DATA FOUND

4. MODELS HAVE REQUIREMENTS FOR:

COMBAT WORTH
BREAKPOINT
MOVEMENT RATES
TARGET SELECTION
POSTURE DETERMINATION
DAMAGE ASSESSMENT TIME

5. PERGORITHMS DO NOT EXIST:

FEAR
DEPRIVATION (E.G. FOOD, WATER)
COMMUNICATION
COURAGE
MORALE
COHESION
NATIONAL DIFFERENCES IN:
    RESISTANCE TO DISEASE
    WILLINGNESS TO FIGHT
    TRAINING LEVELS
    LEADERSHIP

SURPRISE
INTELLIGENCE
COMMAND AND CONTROL
DECISIONMAKING
AGGREGATION TECHNIQUES

A brief look at combat models suggests that attempting to incorporate human performance models directly is likely to be unsuccessful. These two types of model were developed for different purposes, deal with human performance at very different levels of resolution, and generally have quite different structures. Nevertheless, it does appear reasonable to use human performance models for off-line generation of statistical distributions and performance shaping functions that can be incorporated in combat models (Figure 1). In considering these suggestions, the reader should remain aware of model validation issues. Most models have not been adequately validated. The extensions proposed here would require further efforts to ensure that the model predictions have sufficient accuracy for the intended application.

Currently available models might be used to generate performance functions for several items in Category 1 of Table 1. Although not directly reviewed by the RSG, models are available to predict target detection as a function of target size, target-background contrast, etc. [17]. Models of target tracking performance (e.g. Crossover Model or Optimal Control Model) can be used as inputs to weapon ballistics/flyout models and thus generate probability of hit functions. Visibility effects on target detection can be estimated by adjusting target-background contrast levels in visual detection models. Similarly, visibility effects on weapons aiming and tracking can be represented by changing well-understood parameters in the manual tracking models (Figure 2).

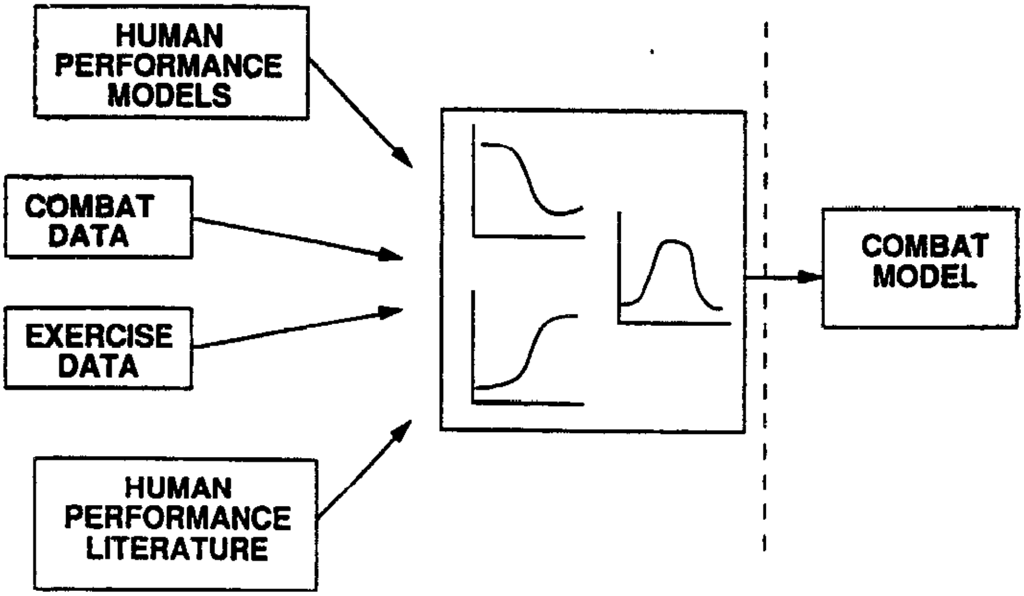


Figure 1. Application of human performance models to generate statistical distributions and performance shaping functions for use in combat models.

In Category 2, human performance models also have potential applications. Van Nostrand [16] describes the use of the PERFECT (Performance Effectiveness of Combat Troops) model to estimate the effects of continuous operations on small unit performance. As mentioned in the previous section of this paper, manual control models have been used to predict the effects of heat stress and encumbrance produced by chemical protective gear (Figure 2). In addition, models based on the Siegel-Wolf technique have been used to estimate the effects of time stress on multi-task performance and might be extended to incorporate combat fatigue, sleep loss, and other effects [2].

An important factor neglected in most combat models is the adaptation and learning that replacements ("new guys") must undergo before reaching full combat effectiveness. While it is likely that much of this represents adaptation to the fear and stress of combat, the learning effects might be described with some of the skill acquisition models reviewed above. Researchers have found that fairly simple functions, such as first order time constants, can be used to describe skill learning in an industrial environment [15]. It is highly probable that similar functions can be applied to simple combat-related skills, to predict changes in unit proficiency with practice. While these functions could capture a range of learning effects for such skills, they may not be adequate to describe learning or forgetting of the more complex skills critical to combat outcome.

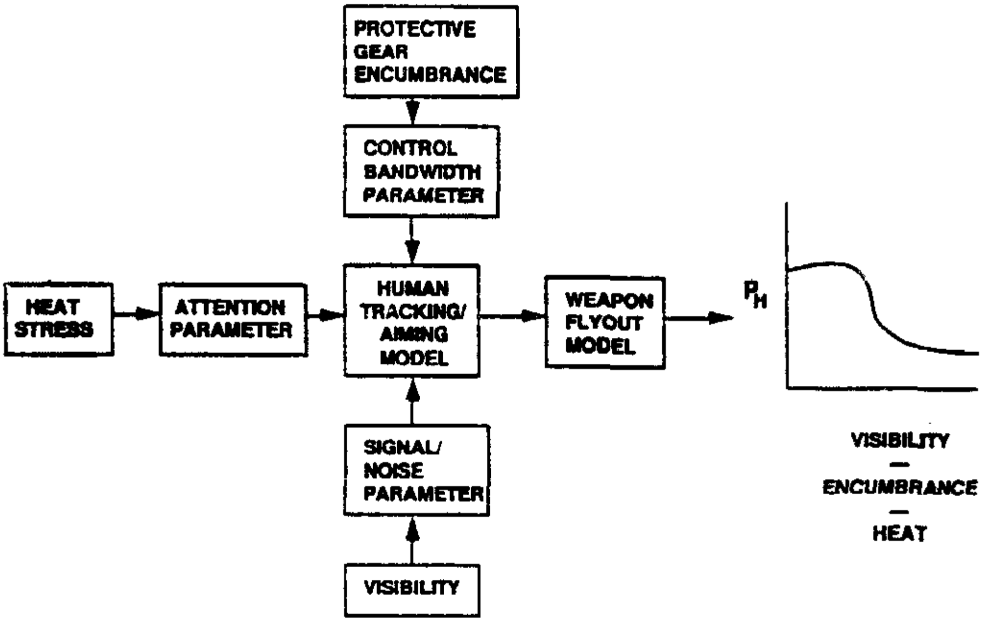


Figure 2. Use of a human tracking/aiming model to generate estimates of weapon system probability of hit under several combat stressor conditions. The stressors are represented by manipulating the specified parameters of the tracking model.

It does not appear that human performance models can address many of the Category 4 requirements. In the area of target selection, some relevant basic work has been done. While the authors are aware of no models that will directly predict human judgments of target worth, there are models that predict how humans will select targets if the worth is known. The Dynamic Decision Model of Pattipati, Kleinman, and Ephrath [18] was able to predict operator selection of targets based on target worth and the time required to process the target. That is, the model predicted the tradeoff humans make between the payoff for processing a target versus the resources required to process it. This tradeoff is undoubtedly a key element often neglected in combat simulations.

In Category 5 it also appears that current human performance models have minimal direct application. Nevertheless, advances in integrated multi-task models offer a great deal of promise, especially for use in lower-level combat models. For example, the Procedure-oriented Crew Model (PROCRU) is a comprehensive model covering a range of operator activities including monitoring, situation assessment, decision making, communication, and discrete and continuous control [19]. PROCRU is an outgrowth of the DEMON model discussed previously. Although the overall integrated PROCRU model has not been experimentally validated, many of its constituent components -- such as the continuous information processing submodel and some single-task submodels -- have been well validated. As with any such comprehensive model, potential interactions among submodels leave the validity of the overall integrated model open to question, regardless of the validity of the submodels.

Operations analysts should also review models such as Metacrew [20], which is a computer simulation of the Joint Surveillance Target Attack Radar System (JSTARS) Ground Station Module. It was developed to provide a time-efficient tool that could examine the relationship between human performance variables and overall systems performance under a wide range of battlefield conditions. The model simulates the crew's target processing and decision making tasks and the normal sequential flow of these tasks.

Components of the model include: (a) the operator behavioral model, (b) the personnel resources model, (c) the battlefield scenario, (d) the commander's guidance, and (e) the output performance data file. The model allows manipulation of interesting variables such as the effect of the commander's guidance on mission emphasis and crew workload, and the effects of crew size, operator task assignments, and operator skill on system throughput. While Metacrew is unlikely to be directly used in a combat model, it clearly has the potential to generate functional relationships which represent communication, command decision making, and personnel resource effects on the performance of a small sensor data processing unit. Alternatively the Metacrew model might be modified to represent other types of intelligence units or to estimate the impact of other human factors variables.

This leads to the principal point of this paper. Many of the human performance models reviewed by the RSG require modification to be useful, but form an excellent starting point to generate the types of data required by combat models. The potential of network modelling tools such as Micro

SAINT [4], and analysis and modelling systems such as GENSAW II [21] is in this same vein. They provide the analyst with tools to generate the required functional relationships.

CONCLUSIONS

Existing human performance models will be most useful in an off-line mode to generate statistical distributions and performance shaping functions that are compatible with combat models.

Human performance models generally have not been developed in the context of combat simulation. However, the model parameters often allow representation of factors important to combat outcome.

The availability of network modelling tools provides a means to readily implement system simulations. Human performance modelers are currently exploiting these tools, and are developing models that have application to combat simulations.

Implementation or modification of human performance models requires specific knowledge and experience. It is unlikely that operations research analysts can effectively exploit available human performance models without the assistance of experienced performance modelers.

REFERENCES

1. McMillan, G.R., Beevis, D., Salas, E., Strub, M., Sutton, R. & van Breda., L. (Eds.). (in press). Applications of Human Performance Models to System Design. New York: Plenum.

2. McMillan, G.R., Beevis, D., Salas, E., Stein, W., Strub, M. & Sutton, R. (in press). A Directory of Human Performance Models for System Design. Brussels, Belgium: NATO Headquarters, Defence Research Section.

3. Parks, D.L. & Boucek, G.P. (in press). Workload prediction, diagnosis, and continuing challenges. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

4. Laughery, K.R. (in press). Micro SAINT - a tool for modeling human performance in systems. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

5. Aldrich, T.B., Szabo, S.M. & Bierbaum, C.R. (in press). The development and application of models to predict operator workload during system design. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

6. North, R.A. & Riley, V.A. (in press). W/INDEX: a predictive model of operator workload. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

7. Jagacinski, R.J. (in press). Target acquisition: performance measures, process models, and design implications. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

8. Rood. G.M., Patterson, R.D. & Lower, M.C. (in press). Modeling of auditory masked thresholds in humans. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

9. Allen, R.W., McRuer, D.T. & Thompson, P.M. (in press). Dynamic systems analysis programs with classical and optimal control applications of human performance models. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

10. Levison, W.H. (in press). The optimal control model for manually controlled systems. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

11. McMillan, G.R., Cody, W.J. & Mills, R.G. (1983). Laboratory studies of aircrew chemical protective ensemble: effects on pilots' performance. In Sustained Intensive Air Operations: Physiological and Performance Aspects (AGARD-CP-338). Paris, France: Advisory Group for Aerospace Research and Development.

12. Muralidharan, R. & Baron, S. (1979). Demon: a human operator model for decision making, monitoring, and control. Proceedings of the Fifteenth Annual Conference on Manual Control (AFFDL-TR-79-3134). Wright-Patterson Air Force Base, OH: Air Force Flight Dynamics Laboratory.

13. Siegel, A.I. & Wolf, J.J. (1969). Man-machine Simulation Models: Psychosocial and Performance Interaction. New York: Wiley.

14. Harris, R., Iavecchia, H. & Dick, A.O. (in press). The Human Operator Simulator (HOS-IV). In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

15. Towill, D.R. (in press). Selecting learning curve models for human operator performance. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

16. Van Nostrand, S.J. (1986). Model Effectiveness as a Function of Personnel [ME=f(PER)] (CAA-SR-86-34). Bethesda, MD: US Army Concepts Analysis Agency. (DTIC No. AD-B109139)

17. Greening, C.P. (1976). Mathematical modeling of air-to-ground target acquisition. Human Factors, 18(2), 111-148.

18. Pattipati, K.R., Kleinman, D.L. & Ephrath, A.R. (1983). A dynamic decision model of human task selection performance. IEEE Transactions on Systems, Man, and Cybernetics, 13(3), 145-166.

19. Baron, S. & Corker, K. (in press). Engineering-based approaches to human performance modeling. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

20. Plocher, T. (in press). Metacrew: simulating operator teams in a sensor data processing system. In G.R. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton & L. van Breda (Eds.), Applications of Human Performance Models to System Design. New York: Plenum.

21, Mills. R.G., Meyer, E.G., & Maher, F.A. (1988). A user-assisted Generic Systems Analyst Workstation (GENSAW): a technical paper. Proceedings of the National Aerospace and Electronics Conference. Dayton, OH.

A HUMAN PERFORMANCE CONSULTANT SYSTEM AND SOME APPLICATIONS

Valerie J. Gawron, David J. Travale, and Jeanette G. Neal
Arvin/Calspan

This paper will describe the evolution of a human performance expert system from a FORTRAN module of an Air Defense Model to a stand-alone PROLOG Human Performance Consultant System. This paper will describe the reasons for the system's evolution, as well as its applications in air defense and air-to-ground mission modeling.

Step 1 - Model of Operator Performance in Air Defense Systems (MOPADS)

As systems become more complex and expensive, greater emphasis is being placed on evaluation of system performance during the design phase. In lieu of a hardware model, a growing number of system designers are using computer models to evaluate alternate system designs. A key element in determining system performance from these models is a consideration of the performance of the human in the system. An accurate representation of the human's capabilities and limitations in a system model can greatly enhance the model's utility and applicability to the real world. Recognizing this fact, the Army supported the development of a human performance model incorporated into a netted air defense system testbed. This effort, Model of Operator Performance in Air Defense Systems (MOPADS), began in 1981 and ended in 1983.

MOPADS is written in the SAINT (Duket, et al., 1978) simulation language. SAINT uses activity networks to represent operator/machine systems. The components of the networks are "task nodes" which represent tasks or activities and "branches" which define the method by which the activities are sequenced. The human performance model, contained within MOPADS, is a series of FORTRAN subroutines. These subroutines were developed in four successive tasks.

Task 1 - Develop a Skill Taxonomy. The performance of any task requires the operator to exercise one or more skills. Task performance can be moderated by calculating the effect of one or more independent variables on the performance of the skills necessary to execute that task. Towards this end, a response-defined skill taxonomy was developed and is presented in Table 1.

Table 1
MOPADS SKILL TAXONOMY

1. Detection - the ability to discover or become aware of a visual, auditory, tactile, olfactory, proprioceptive, or kinesthetic stimulus.
2. Fine Manipulative Ability - the ability to manipulate controls through a physical effort which requires sensitive movement and touch rather than physical strength.
3. General Physical Effort - the ability to perform tasks which require strength and reach more than sensitive control.
4. Gross Manipulative Ability - the ability to manipulate large controls through a physical effort which requires sensitive movement, touch, and physical strength.
5. Long-Term Memory of Sensory Information - the ability to recall sensory information from long-term memory, e.g., the meaning of an auditory alert.

6. Long-Term Memory - Symbolic - the ability to recall symbolic information in the long-term memory, e.g., the mathematical operations associated with +, -, /, and x.

7. Numeric Manipulation - the ability to estimate and perform mathematical calculations.

8. Probability Estimation - the ability to estimate the probability (or the chance) of events occurring.

9. Reaction - the ability to physically respond to a detected stimulus.

10. Recognition - the ability to identify a detected stimulus.

11. Short-Term Memory of Sensory Information - the ability to recall sensory information from the short term memory buffer, e.g., last value on a display being monitored.

12. Short-Term Memory of Symbolic Information - the ability to recall symbolic information from the short-term memory buffer, e.g., last position of a target on a display.

13. Team Coordination - the ability to organize and implement a team effort.

14. Time Estimation - the ability to estimate the time a moving body will take to travel a fixed distance.

15. Time Sharing - the ability to perform more than one task simultaneously.

16. Tracking - the ability to follow a moving target with a control, e.g., joystick.

Task 2 - Conduct a Literature Search. A computerized literature search (Laughery, 1981; 1982) was conducted to find human-performance data relating independent variables of interest to the skills identified in the taxonomy. As a result, three categories of independent variables were identified: 1) environmental variables, e.g., ambient temperature, 2) operator variables, e.g., operator's time on task, and 3) task variables, e.g., the modality of the target to be detected.

Task 3 - Curve Fit Human Performance Data. Data identified during the literature search which related to operator performance in air defense systems were then subjected to standard curve-fitting techniques to reduce the data to a single moderator function. Seven types of equations were used (see Table 2). The form of the equation which accounted for the most variance (greater correlation R^2) in the data was chosen for use in the model.

Table 2
TYPES OF EQUATIONS

SIMPLE REGRESSION ($y=mx+b$)
SIMPLE REGRESSION - EXPONENTS ($y=x^{m}+b$)
SIMPLE REGRESSION - INVERSE EXPONENTS ($y=x^{-m}+b$)
SIMPLE REGRESSION - SQUARED EXPONENTS ($y=x^{m^2}+b$)
SIMPLE REGRESSION - INVERSE SQUARED EXPONENTS ($y=x^{-m^2}+b$)
MULTIPLE REGRESSION ($y=m_1x_1+m_2x_2 \ldots +b$)
POLYNOMIAL REGRESSION ($y=mx^2+mx+b$)

Task 4 - Code Human-Performance Subroutines. Thirty-two human-performance subroutines were written: one calculating time to complete a task and one calculating probability of completing a task for each of the sixteen skills listed in Table 1. A simplified portion of one of these subroutines is presented in Table 3. Please note that some of the statements exceed the 72 column FORTRAN limit. This has been done to ease reading the code. T(NME) refers to the time for an observer to detect a target as calculated from equation number NME. In the following code, the effects of sixteen independent variables are being calculated: time on task, target types, horizontal range to target, observer's field-of-view, target background complexity, contrast ratio, target subtense, difference between nontarget and target diameters, lines on CRT, target color, number of background characters, search area, ambient noise level, target location, effective temperature, and number of operators on duty.

Table 3
A PORTION OF THE MOPADS SUBROUTINE USED TO CALCULATE TIME TO DETECT A TARGET

```
C**NPOS=THE OBSERVER-TO-TARGET POSITION
C
    GO TO (1,2,3,4)NPOS
C
C**THIS IS A GROUND-TO-GROUND, VISUAL DETECTION TASK
C**TOT=OPERATOR'S TIME-ON-TASK IN HOURS
C**TARGET=TARGET TYPE
C
    1 NME-NME+1
      NIV(NME)=2
C
C**TARGET IS A TANK
C
    IF (TARGET.EQ.11.0)T(NME)=(25.333-0.167*TOT)/60.0
C
C**TARGET IS A JEEP
C
    IF (TARGET.EQ.12.0)T(NME)=(28.000-0.500*TOT)/60.0
C
C**TARGET IS TROOPS
C
    IF (TARGET.EQ.13.0)T(NME)=(34.000-1.000*TOT)/60.0
C
C**XYARD=HORIZONTAL RANGE TO TARGET IN YARDS
C
    NME=NME+1
    NIV(NME)=2
C
C**TARGET IS A TANK
C
    IF (TARGET.EQ.11.0)T(NME)=2.115-1.605*EXP(-0.001*XYARD**2))/60.0
C
C**TARGET IS A JEEP
C
    IF (TARGET.EQ.12.0)T(NME)=(1.618-1.569*EXP(-0.0004*XYARD**2))/60.0
C
C**TARGET IS AN ARMORED PERSONNEL CARRIER
C
    IF (TARGET.EQ.14.0)T(NME)=(0.063+0.014*XYARD)/60.0
C
C**TARGET IS A TRUCK
```

```
C
      IF (TARGET.EQ.15.0)T(NME)=(0.042+0.013*XYARD)/60.0
C
C**TARGET IS A SOLDIER
C
      IF (TARGET.EQ.25.0)T(NME)=(1.581+0.0001*EXP(0.0007*XYARD**2))/60.0
C
C**CALL ABSTMH OR RELTMH TO DERIVE SKILL MODERATORS FOR
C     GROUND-TO-GROUND, VISUAL DETECTION TASKS
C
      IF (NME.EQ.0) THEN
         NME=1
         NIV(NME)=1
         T(NME)=DIST(1)
         CALL RELTMH(DIST,NIV,NME,T,XM)
      ELSE
         CALL ABSTMH(DIST,NIV,NME,T,XM)
      ENDIF
      GO TO 99
C
C**THIS IS AN AIR-TO-GROUND, VISUAL DETECTION TASK
C**FOV=THE OBSERVER'S FIELD-OF-VIEW IN DEGREES
C
      2 NME=NME+1
        NIV(NME)=1
        T(NME)=(332.250-249.106*EXP(-0.002*FOV**2))/60.0
C
C**COMPLEXITY=TARGET BACKGORUND COMPLEXITY
C**CONTRAST=TARGET/BACKGORUND CONTRAST RATIO
C
      NME=NME+1
      NIV(NME)=3
C
C**TARGET IS A TANK
C
      IF (TARGET.EQ.11.0)
         T(NME)=(-55.327+32.149*COMPLEXITY+32.538*CONTRAST)/60.0
C
C**TARGET IS AN ARMORED PERSONNEL CARRIER
C
      IF (TARGET.EQ.14.0)T(NME)=(81.303-2.300*COMPLEXITY-36.077*CONTRAST)/60.0
C
C**TARGET IS A TRUCK
C
      IF (TARGET.EQ.15.0)T(NME)=(-22.331+30.799*XO(5)-0.385*XO(6))/60.0
C
C**CALL ABSTMH TO DERIVE SKILL MODERATORS FOR AIR-TO-GROUND,
C     VISUAL DETECTION TASK
C
      CALL ABSTMH(DIST,NIV,NME,T,XM)
      GO TO 999
C
C**THIS IS A GROUND-TO-AIR, VISUAL DETECTION TASK
C**XMARC=ANGLE THE TARGET SUBTENDS IN MINUTES OF ARC
```

```
C
    3 NME=NME+1
      NIV(NME)=4
      T(NME)=(0.422+61.845*EXP(-0.105*XMARC**2))/60.0
C
C**CALL ABSTMH TO DERIVE SKILL MODERATORS FOR GROUND-TO-AIR,
C   VISUAL DETECTION TASKS
C
    CALL ABSTMH(DIST,NIV,NME,T,XM)
    GO TO 999
C
C**THIS TASK INVOLVES DETECTING TARGETS ON A DISPLAY
C**XINCH=DEIFFERENCE BETWEEN NONTARGET AND TARGET DIAMETERS
C   IN INCHES
C
    4 NME=NME+1
      NIV(NME)=2
      T(NME)=(1/(XINCH**2))/60.0
C
C**LINES=DISPLAY RESOLUTION IN LINES ON A CRT
C
    NME=NME+1
    NIV(NME)=4
    T(NME)=(105.324-2.579*XMARC-0.045*LINES)/60.0
C
C**COLOR:  1=WHITE, 2=TAN; 3=GREEN; 4=BLUE; 5=RED; 6=YELLOW
C**NUMBER=NUMBER OF BACKGROUND CHARACTERS
C
    IF (COLOR.LE.2.0) THEN
C
C**TARGET IS WHITE OR TAN
C
       NME=NME+1
       NIV(NME)=2
       T(NME)=(9.9042-2.491*COLOR+0.0004*NUMBER)/60.0
    ENDIF
C
C**XSINCH=SEARCH AREA IN SQUARE INCHES
C
    IF (COLOR.LE.2.0) THEN
       NME=NME+1
       NIV(NME)=3
       T(NME)=(-2.555+1.366*COLOR+0.020*XSINCH)/60.0
    ENDIF
C
    IF (COLOR.LE.5.0) THEN
       NME=NME+1
       NIV(NME)=1
       T(NME)=(0.594+4.715*EXP(-0.263*COLOR**2))/60.0
    ENDIF
C
    NME=NME+1
    NIV(NME)=1
    T(NME)=(6.000-4.848*EXP(-0.0002*NUMBER**2))/60.0
```

```
C
C**NOISE=AMBIENT NOISE LEVEL IN DB
C
      NME=NME+1
      NIV(NME)=2
      T(NME)=(0.834+0.017*NUMBER+0.0008*NOISE)/60.0
C
      NME=NME+1
      NIV(NME)=1
      T(NME)=(1.197+0.000748*LINES)/60.0
C
C**LOCATION=LOCATION OF TARGET OF DISPLAY
C
      NME=NME+1
      NIV(NME)=1
      IF LOCATION.EQ.1.0)T(NME)=0.027
      IF LOCATION.EQ.2.0)T(NME)=0.025
      IF LOCATION.EQ.3.0)T(NME)=0.025
      IF LOCATION.EQ.4.0)T(NME)=0.026
      IF LOCATION.EQ.5.0)T(NME)=0.024
      IF LOCATION.EQ.6.0)T(NME)=0.026
      IF LOCATION.EQ.7.0)T(NME)=0.032
      IF LOCATION.EQ.8.0)T(NME)=0.028
      IF LOCATION.EQ.9.0)T(NME)=0.030
C
C**ET=EFFECTIVE TEMPERATURE IN DEGREES CENTRIGRADE
C
      NME=NME+1
      NIV(NME=3
      T(NME)=(0.681+0.307*TOT+0.020*ET)/60.0
C
C**OPERATORS=NUMBER OF OPERATORS ON DUTY
C
          NME=NME+1
          NIV(NME)=2
          T(NME)=(1.210+0.098*OPERATORS+0.041*TOT)/60.0
C
C**CALL ABSTMH TO DERIVE SKILL MODERATORS FOR
C     DISPLAY TARGET DETECTION
C
      CALL ABSTM(DIST,NIV,NME,T,XM)
999   RETURN
      END
```

Two characteristics of the code are readily apparent: heavy use of IF and comment statements. The IF statements were used to meet one of the goals of MOPADS, specifically, to maximize the match between the conditions being simulated and the conditions under which field data were collected. This required the heavy use of branching in the code based on the current values of the environmental, operator, and task state variables. In the detection subroutine, the major braches were defined by target modality (auditory, visual, or both) and observer-to-target position (ground-to-ground, air-to-ground, ground-to-air, or at-a-display). The heavy use of comment cards were used to meet a second goal of MOPADS: to provide traceability of results to empirical data.

What is not obvious from the data is the method used to combine the results of multiple equations relevant to the current state being simulated. In these cases, the results were combined in a weighted equation to produce a single derived time. The weight for each calculated time was the number of independent variables represented in the moderator equation from which that particular time value was calculated. Further, the difference between the values of the baseline mean and the time calculated from moderator equation N was divided by N. (The equations were ordered from largest to smallest effect on the dependent variable.) This approach was based on the law of diminishing returns which states that the inclusion of more predictor variables in a regression equation will increase the total amount of variance accounted for by the equation; however, each new variable added will account for a lesser portion of the total variance than any of the variables which preceded it. This method was empirically evaluated in Step 4 of the Human evolution. A sample list of input variables in the detection time subroutine is provided in Table 4.

Table 4
INPUT VARIABLES FROM THE MOPADS SUBROUTINE USED TO CALCULATE TIME TO DETECT

Nontarget-target diameter
Lines on a CRT
Target Color
Number of background characters
Search area .
Noise level
Target location
Effective temperature
Number of operators on duty

Listings of each of the thirty-two skill subroutines are provided in Laughery and Gawron (1983) and can be incorporated into any combat simulation capable of supporting FORTRAN subroutines.

Step 2 - Human Utilization Model and Analytic Network (HUMAN)

At the end of the MOPADS effort it was clear that the human-performance model was a very powerful computer aided engineering tool for human factors engineers. It was also clear that it had three major problems: 1) the model was only accessible through the large MOPADS air defense system model, 2) users could not distinguish 16 separate skills, and 3) no estimates of human performance errors were calculated. To meet these deficiencies, HUMAN was developed as part of the Air Force's Cockpit Automation Technology (CAT) program. The purpose of the CAT program was to develop a methodology to evaluate aircrew performance in crew stations being designed. HUMAN's first application was in evaluating the tactical worth of eight cockpit automation technologies.

Description of HUMAN. HUMAN is a stand-alone model which models task performance of human beings under conditions prescribed by the user. Large tasks can be modeled by using the network capability of HUMAN. Specifically, complex tasks can be simulated by breaking the task into subtasks each corresponding to a node in the network. These nodes may be used in series to indicate sequential tasks or in parallel to model a time-shared task.

The conditions of each node in the network are defined by three state vectors. Vector 1 contains variables which describe the human's surroundings. Examples are type and concentration of chemical agent present and characteristics of collective protection shelters. Vector 2 contains variables which describe a human's physical capabilities and present condition. Examples might include amount of antidote present in the bloodstream and degree of sleep deprivation. Vector 3 contains variables which describe the type of task the human is asked to perform. Examples of these are number of displays to monitor and number of alternative responses possible. Using these vectors, the user can input values to define the environment in which the human is to be modeled.

After values are input into the vectors, HUMAN then requires the user to input the skill(s) the human will need to perform the desired task. A menu, consisting of ten skills (reduced from the 16 in MOPADS due to user comments) to which the user must assign weights, is presented. The total weight for all ten skills must sum to 1.00. Therefore, should a task require detection and tracking skills with equal importance, each of their assigned weights should equal 0.50. Skill definitions are given in Table 5. The outputs of the model are time to complete, the probability of completion, and the percentage of errors incurred while performing the task.

Table 5
HUMAN SKILL DEFINITIONS

1. Decision Making - the ability to choose between two or more alternatives.
2. Detection - the ability to discover or become aware of a visual, auditory, tactile, olfactory, proprioceptive, or kinesthetic stimulus.
3. Fine Manipulation - the ability to manipulate controls through a physical effort which requires sensitive movement and touch rather than physical strength.
4. Gross Manipulation - the ability to manipulate controls that require significant physical strength.
5. Numeric Manipulation - the ability to estimate and perform mathematical calculations.
6. Probability Estimation - the ability to predict the chance of an event occurring.
7. Recognition - the ability to identify a detected stimulus.
8. Team Coordination - the ability to organize and implement a team effort.
9. Time Estimation - the ability to predict how long it will take a moving body to travel a fixed distance.
10. Tracking - the ability to follow a moving target with a control, e.g., joystick.

Application of HUMAN. HUMAN's first case was used as part of a suite of models to evaluate the tactical worth of four air-to-ground and four air-to-air cockpit automation technologies. The evaluation process consisted of three consecutive tasks.

Task 1 - Develop Baseline Mission Timelines. Two timelines were developed: 1) an air-to-ground mission timeline derived from the Methodology for Penetration Evaluation (MPIRE) series of mission models (see Gawron, Travale, and Quinn, 1986) and 2) an air-to-air mission timeline derived from a McDonnell Douglas mission model (see Quinn, Gawron, Arbak, and Dike, 1986). Both timelines were the results of a baseline aircraft (i.e., without cockpit automation technology enhancements) penetrating a hostile territory. Both timelines provided time constraints for completing in-cockpit tasks as well as the conditions under which those tasks would be performed (e.g., g load and airspeed). Portions of each timeline are presented in Figures 1 and 2.

The mission timelines are presented here in chronological order. Mission segments and ground speeds are presented in coded form. For each segment, weather

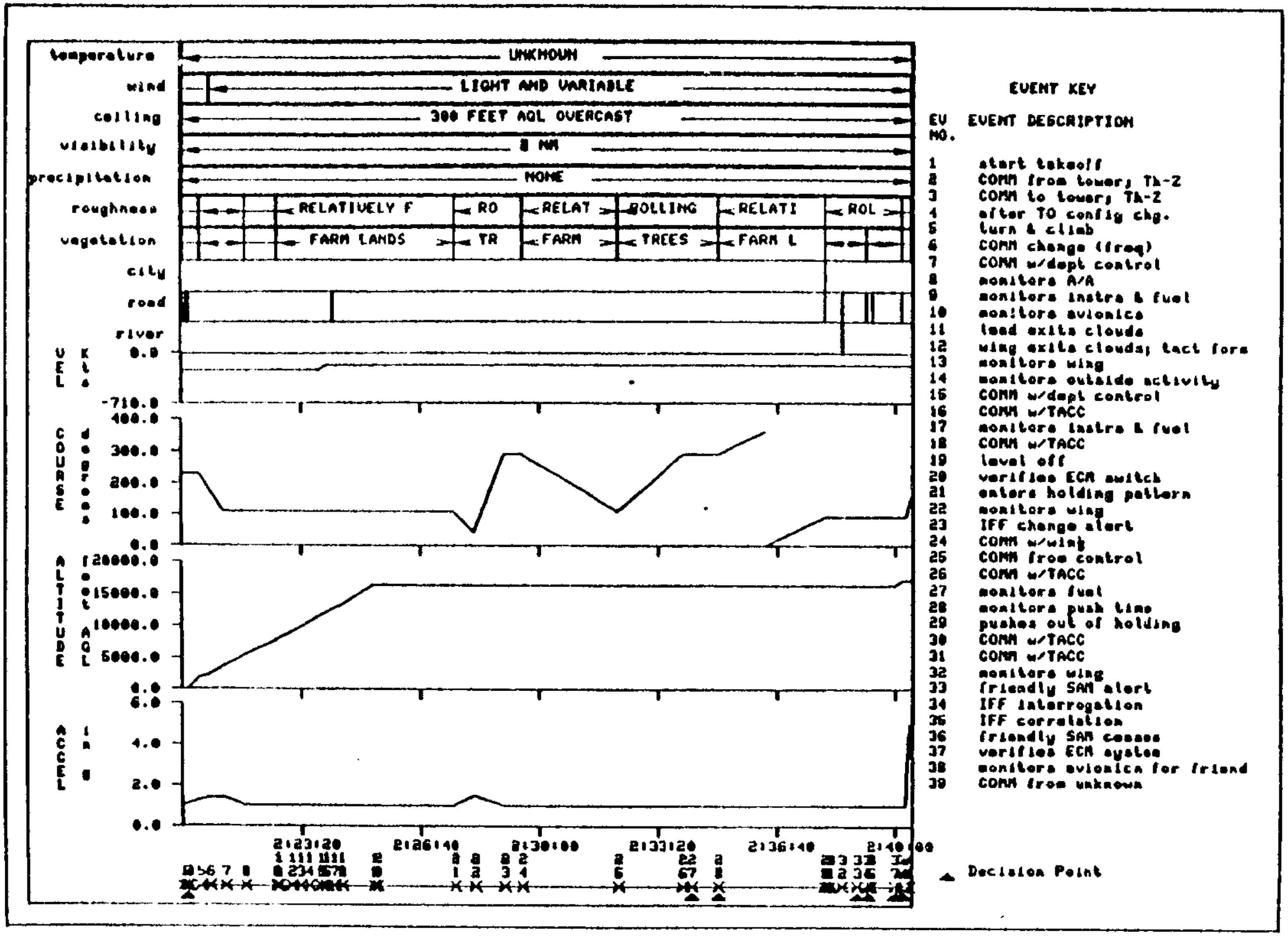
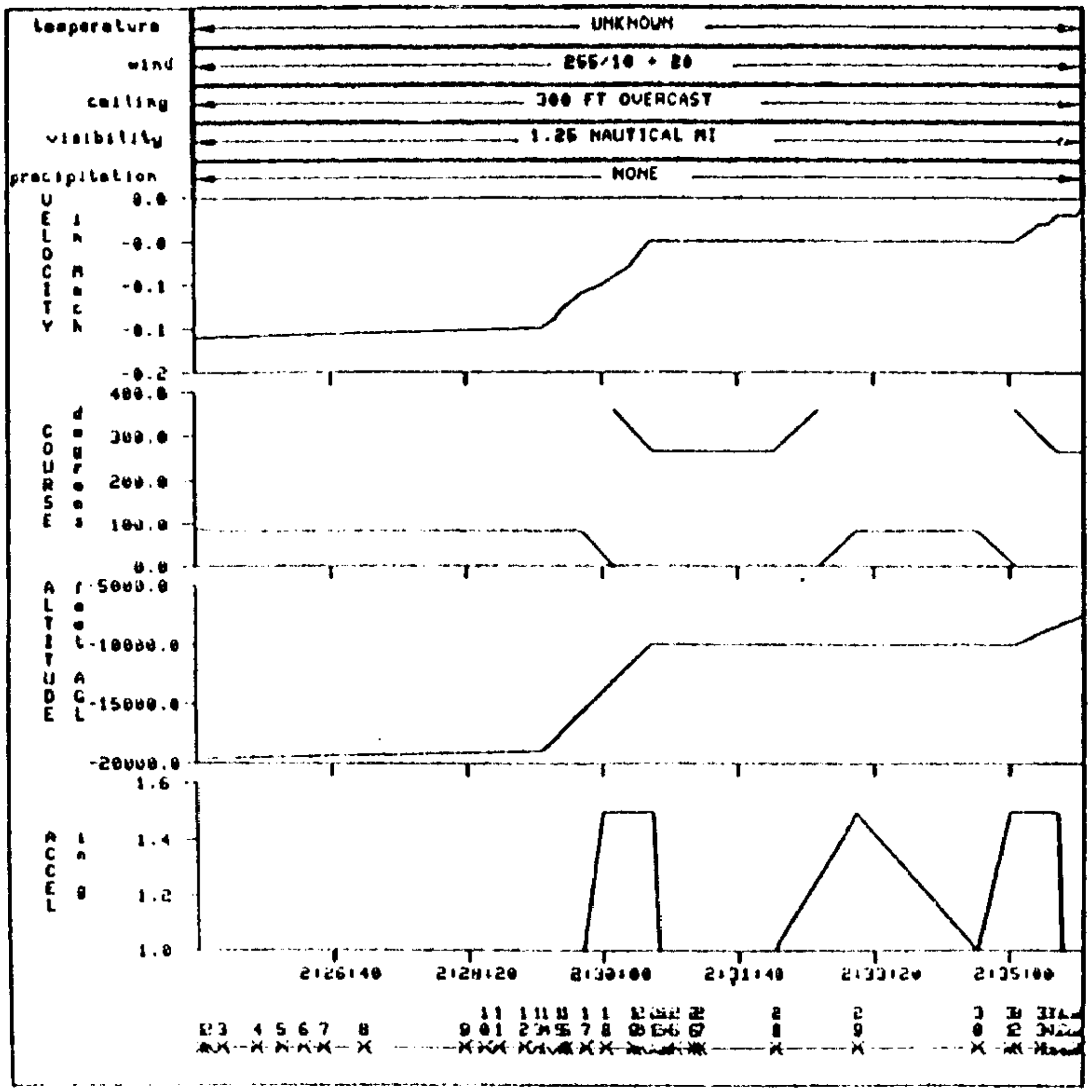


Figure 1 AIR-TO-GROUND BASELINE MISSION TIMELINE



EVENT KEY

EV NO.	EVENT DESCRIPTION
1	start TO; COMM freq chg; mon
2	after TO config chg; adj pit
3	turns left & continues climb
4	COMM from dept control
5	maintains 300 kts climb spee
6	Cog mon eng instrs & sysems
7	mon avionics displays for th
8	other flt member event only
9	scans flt instrs; cks fuel q
10	scans nav displays; verifies
11	exits clouds
12	COMM fm dept control; COMM f
13	verifies flt members on corr
14	monitors eng & flt instrs; c
15	observes data on battle mgmt
16	COMM from TACC regarding hol
17	enters left holding & contin
18	second wingman joins formati
19	performs navigation & fuel c
20	3rd wingman joins; rvws bat
21	levels off at FL 200; adjust
22	turns outbound in holding pa
23	second flight lead joins in
24	COMM from TACC regarding pro
25	first wingman of second flig
26	second wingman of second fli
27	third wingman of second flig
28	starts turn inbound in holdi
29	rolls out inbound
30	starts turn outbound in hold
31	COMM w/TACC re climb and flt
32	cks bat mgmt; climbs; enters
33	mon flt rle on avionics for
34	AAA/SAM indications from fri
35	rolls out outbound
36	verifies ECM mode
37	IFF interrogated
38	correlates interrogations w/
39	SAM alert; determines friend
40	IFF interrogated

▲ Decision Point

Figure 2 AIR-TO-GROUND BASELINE MISSION TIMELINE (air temperature, wind velocity, ceiling, visibility, and precipitation), terrain (roughness, vegetation, and presence of cities, roads, and rivers), aircraft altitude, speed, g load and course are given. Along the x-axis of the timeline, the time and event occurrence information are plotted. Times are coded against a base time of the beginning of the segment. Events are numbered. Decision nodes are indicated by an upward arrow. Actual event and decision descriptions are generally classified in the mission document and therefore are merely plotted in the timeline without description. In the baseline air-to-ground timeline: the aircraft spent 2490 seconds within missile tracking range, 25 missiles were launched against the aircraft, and all 25 reached successful intercept range.

Task 2 - Identifying Critical Events. A three-point rating scale was developed to identify the criticality of each event in both mission timelines. On this scale, criticality was defined in terms of the consequence of not performing the appropriate tasks in response to a given tactical event. A rating of one was assigned when no important effect was evident or the mission would be degraded only slightly. Two was given when the consequence was a shortened mission or damaged equipment. The highest level of criticality, three, was assigned when personnel injury or catastrophe (crash/death) would result.

Task 3 - Estimating Pilot Performance Using Each Cockpit Automation Technology. Using HUMAN's network capability, a task network of the pilot/cockpit-automation-technology interactions was developed for each level three critical event and each candidate cockpit automation technology.

The first air-to-ground technology was an azimuth release computation (ARC) system. ARC was developed to reduce pilot workload in the weapon delivery phase. It automatically tracks the target and maneuvers the aircraft.

The second air-to-ground technology evaluated was a Highly Accurate Navigation (HAN) system. HAN automatically updates position information. HUMAN determined that HAN had a minimal effect on reducing pilot workload since only one initial navigation system was made for the entire baseline air-to-ground timeline.

The third air-to-ground technology was automatic terrain following/terrain avoidance (ATF/TA) performance. HUMAN showed that ATF/TA significantly reduced workload during ingress and egress but had no effect on target acquisition or weapon delivery. The fourth air-to-ground technology was an automatic threat response/route planning (ATR/RP) system. HUMAN indicated that ATR/RP reduced pilot workload during ingress and egress.

Four air-to-air cockpit automation technologies were also evaluated using HUMAN. A helmet mounted display/sight had no effect on pilot workload. This may be due to the scenario which was a long-range encounter. The multi-sensor integration system for identifying foes also had little effect on pilot workload since, in the generated air-to-air baseline timeline, the identity of an aircraft was never in question. An automatic airborne target prioritization (AATP) system was similarly evaluated as not useful since the timeline indicated 1.00 second to judge the range and closing rate of 20 red fighters. Finally an automatic distribution of airborne target assignments (ADATA) system greatly reduced workload by substantially reducing the clutter on the pilot's radar display.

HUMAN is a stand-alone FORTRAN model currently hosted on microcomputers at Calspan and at the Armstrong Aerospace Medical Research Laboratory. It has a user-friendly interface and can be used to 1) generate data for human performance look up tables in combat models or 2) be incorporated into the combat models as a set of FORTRAN subroutines.

Step 3 Evaluation of Combining Algorithms

HUMAN's greatest strength is its use of empirical rather than the theoretical data. However, a problem was encountered when multiple independent data sets were relevant to one task. For example, a modeler is interested in the effects of luminance and font size on number of reading errors. Two data sets exist in the literature: one examining the effects of luminance, the other, font size. The data in the two sets were collected at different locations with different subjects and at different times in history. How can the two data sets be combined to address the designer's problem?

Four combining algorithms were developed. Algorithm one calculated a simple mean:

$$E_1 = \frac{\sum_{i=1}^{n} P_i}{n}$$

where E is the estimate of human performance; P, the performance predicted from one independent data set; and n, the number of independent data sets. This algorithm is simple to apply and straight-forward. There is also a theorem that states that the best predictor of performance in a distribution is the mean - if the distribution is normal or near-normal. Most of the data in the human-performance literature were not normally distributed, however, especially reaction times. Further, this algorithm fails to account for the numbers of observations in each data set. Data sets of 1 or 1000 observations are given equal weight.

To address this issue, a second algorithm was developed:

$$E_2 = \frac{\sum_{i=1}^{n} P_i \cdot k_i}{\sum_{i=1}^{n} k_i}$$

where k is the number of observations in each independent data set. This algorithm gives heaviest weight to those studies with the most observations, since these studies are most likely to have the greatest stability in performance estimates. However, variability is not addressed.

The third algorithm uses the standard deviation:

$$E_3 = \frac{\sum_{i=1}^{n} \frac{P_i}{SD_i}}{\sum_{i=1}^{n} \frac{1}{SD_i}}$$

where sd is the standard deviation in each of the independent data sets.

None of these algorithms, however, reflects the number of independent variables used to make the prediction. Therefore, the last algorithm uses the number of independent variables:

$$E_4 = \sum_{i=1}^{n} \frac{P_i \, IV_i}{\sum_{i=1}^{n} IV_i}$$

where IV is the number of independent variables manipulated in the data set. This algorithm is based on the law of diminishing returns which states that including more predictor variables in a regression equation will increase the total amount of variance in the data accounted for by the equation; however, each new variable added will account for a lesser portion of the total variance than any of the variables that preceded it.

These algorithms were then tested in two steps. In step one, two reaction-time experiments were conducted: one to evaluate the effect of the number of alternatives on reaction time; the second, signals per minute and number of displays being monitored. The four algorithms were used on the data from these two experiments to predict reaction time in the situation where all three independent variables are manipulated simultaneously. In step two of the test procedure, a third experiment was conducted. Subjects who had not participated in either Experiment One or Two performed a reaction-time task under the combined effects of all three independent variables. The predictions made from step one were compared to the actual empirical data collected in step two.

To summarize the results, the best predictor of the mean was an unweighted average of the means in the independent data sets; the best predictor of the standard deviation was an unweighted average of the standard deviations in the independent data sets. These results have been incorporated into HUMAN. Specifically, an unweighted average of the means is used to predict mean performance which results from the combined effects of the independent variables being analyzed in HUMAN. Similarly, an unweighted average of the standard deviations is used to predict the standard deviation of performance resulting from the combined effects of the independent variables being analyzed.

Step 4 **Human Performance Consultant System (Human)**

Although HUMAN was a useful tool for the CAT project, a number of problems were identified in attempting to develop it into a more general purpose system:

1) separation of data base and knowledge base: since a human performance data base was used to store source data and while HUMAN was used for the knowledge base, updating and maintaining both systems was difficult;
2) computational limits: since only the three dependent variables were computed (i.e., time to complete the task, probability of completing the task, and percent errors), important studies (which recorded other dependent measures) could not be included;
3) difficulty of updating: each new study required that an entire FORTRAN subroutine be coded or recoded. This was not only time consuming but would result in the program growing out of bounds. This procedure also limited expansion to users capable of programming in FORTRAN (or at least to those with a FORTRAN programmer available).
4) inaccessibility of the source code: the user was not able to query the system to identify what "rules" or what studies were being used in the analysis.
5) lack of simultaneous data base inquiry while modeling: since the data base was separate from the procedurally represented knowledge base, a user had to quit the modeling system to query the data base.
6) all variables had to be set: regardless of what task a user wished to model, values had to be assigned to all items referenced in the three vectors. By the end of model development, there were over one hundred values that had to be assigned.
7) numerous default values: while default values may be set, a user still needs to keep in mind that default values for variables which may not be relevant to the current issue could influence the outcome.

The overall goal of the redesign of HUMAN was to provide a flexible tool for human factors engineers to use in assessing human performance in most any domain. Such a system must remain flexible enough to contain both general-world knowledge and domain-specific knowledge. It must do this but avoid becoming too domain-specific as was the case with HUMAN. Achieving this balance was a difficult task even with modern AI techniques and programming tools.

The best alternative, considering the goals and requirements of the system, was a design that uses the same data for both traditional data base activities (queries, etc.), and for the knowledge base. This would allow users to browse through the same data that the expert system component used to make judgments and computations. Since such a system would be manipulating data to make expert judgments, it would be more an "intelligent data base" than an expert system.

An integrated design has several distinct advantages, the most important being consistency of data and ease of maintainability. A new study, added as a new record to the data base, would be automatically available to the expert system. Likewise changes would be immediately available to both systems. The lack of redundancy that this design provides also ensures that both the data base and knowledge base are always using the same information. The importance of this will become more critical as the system grows. It would eventually become impossible to ensure consistency of data if the two are separated. With separate systems, the potential exists for system inferences to be made with different knowledge from that which the user sees in the data base. This would most certainly confuse the user and cause a loss of faith in the system. Figure 3 is a conceptualization of how this integrated design differs from the design used for the human performance data base and HUMAN. The new system was defined in terms of a central control module and six major sub modules. These modules are illustrated in Figure 4. The new system was named The Human Performance Consultant System (Human).

The central control module is the main entry point of the program as well as the main control module. In general, it loads the data base, calls various screens and menus, and activates main goals. This feature allows modules simply to be added to the central control without concern for causing side effects to other parts of the program.

The user interface module defines the screens and menus used to allow user selection. While it is easy to define screens, it is difficult to change existing screen definitions. As such, care should be taken to assure that the screens are as final as possible before putting them into place. Defining new screens will not be an ongoing process but is done only during system development. A sample screen is presented in Figure 5.

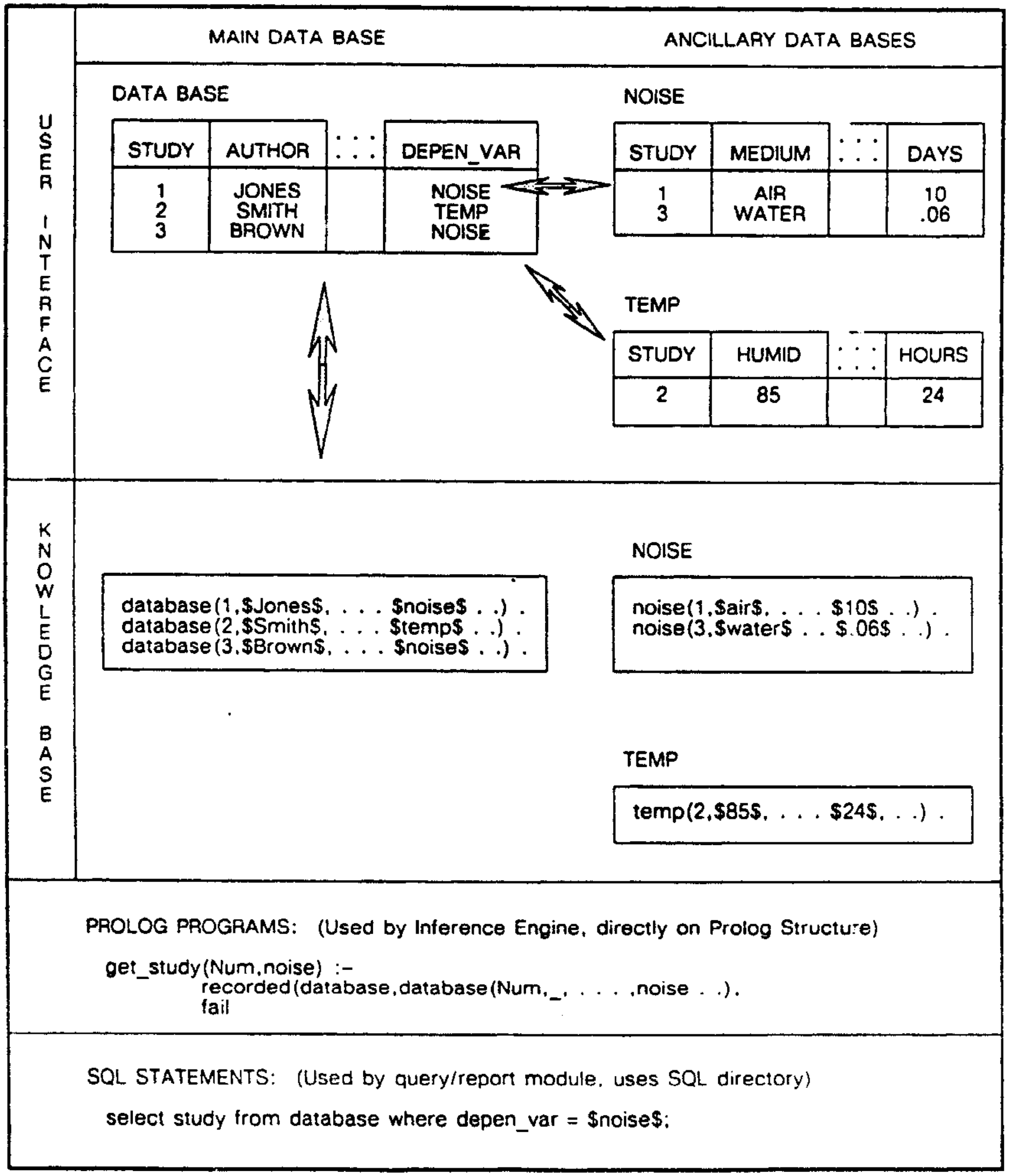
The Data Base Definition module defines the relational data base tables used by the system. Once defined, a table is saved as part of the system environment. As such, this process is repeated only when data bases are redefined or new ones are added.

The Data Base Processes module contains traditional data base functions that add, edit, and delete records. It also contains query definitions and goals that activate queries. If there are any reports defined for the system, such as a listing of all studies, the processes to produce them would also be included here.

The Modeling Module is the process that most closely resembles HUMAN. It is the intelligent component that carries out processes for modeling user-defined scenarios and predicting expected outcomes.

The knowledge acquisition part of the system includes all processes related to the system gaining new knowledge. It will rely heavily on the Data Base Processes module and will serve to coordinate use of processes in that module. It determines and does what is needed to facilitate the addition of new studies or information.

With this design, the data base itself becomes the most important part of the system. It is in the data base that all knowledge is represented. Other parts of the system will manipulate this knowledge to accomplish desired results. This manipulation will range from simple update and queries to "intelligent" processing with information retrieved from the data base.



DATA BASE

STUDY	AUTHOR	. . .	DEPEN_VAR
1	JONES		NOISE
2	SMITH		TEMP
3	BROWN		NOISE

NOISE

STUDY	MEDIUM	. . .	DAYS
1	AIR		10
3	WATER		.06

TEMP

STUDY	HUMID	. . .	HOURS
2	85		24

```
database(1,$Jones$, . . . $noise$ . .) .
database(2,$Smith$, . . . $temp$ . .) .
database(3,$Brown$, . . . $noise$ . .) .
```

NOISE

```
noise(1,$air$, . . . $10$ . .) .
noise(3,$water$ . . $.06$ . .) .
```

TEMP

```
temp(2,$85$, . . . $24$, . .) .
```

PROLOG PROGRAMS: (Used by Inference Engine, directly on Prolog Structure)

```
get_study(Num,noise) :-
          recorded(database,database(Num,_, . . . ,noise . .).
          fail
```

SQL STATEMENTS: (Used by query/report module, uses SQL directory)

```
select study from database where depen_var = $noise$;
```

Figure 3 HUMAN INTEGRATED DATA REPRESENTATION

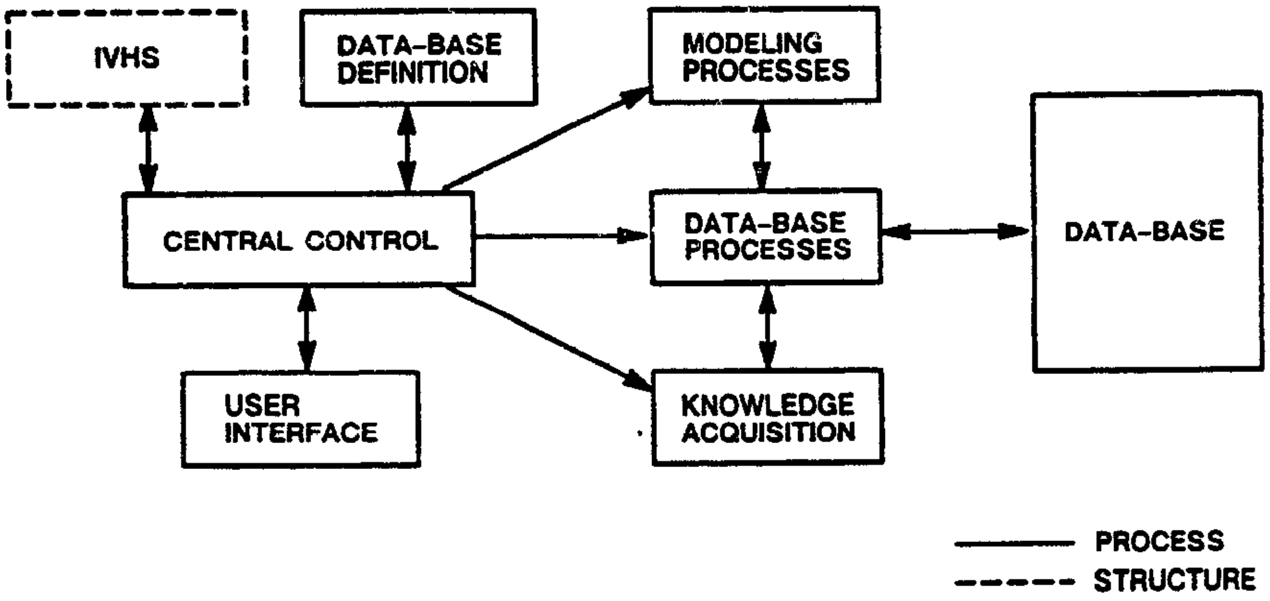
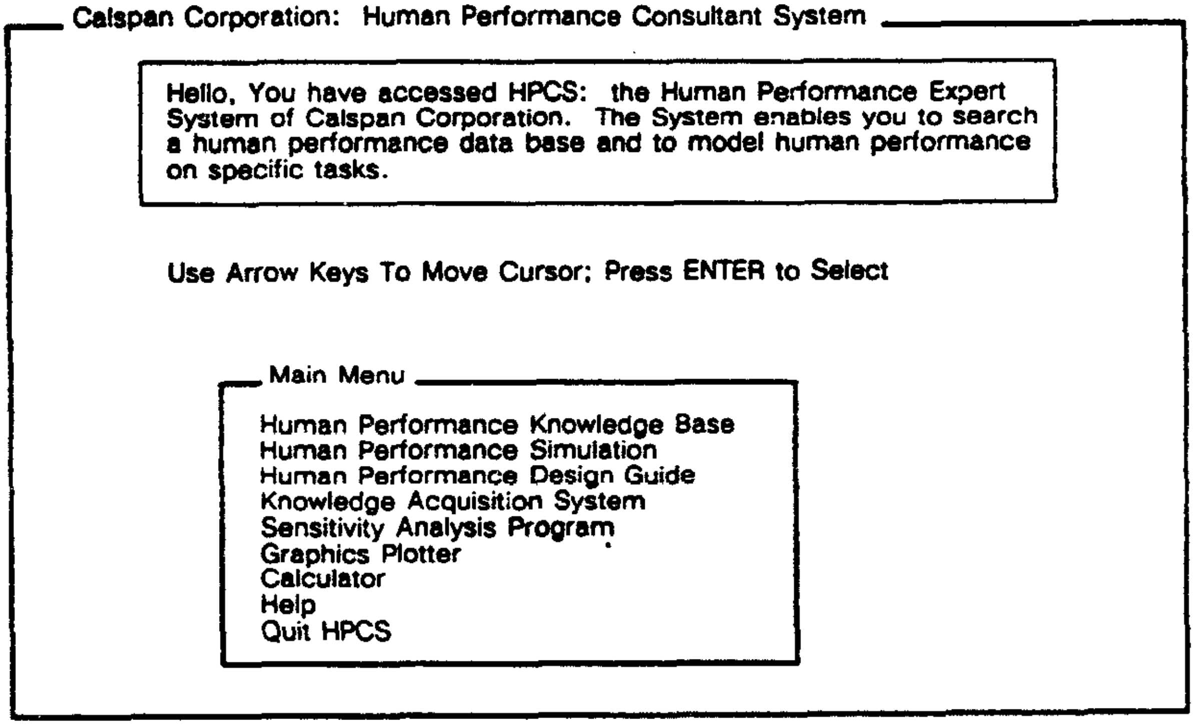


Figure 4 HUMAN MAIN MODULES



Calspan Corporation: Human Performance Consultant System

Hello, You have accessed HPCS: the Human Performance Expert System of Calspan Corporation. The System enables you to search a human performance data base and to model human performance on specific tasks.

Use Arrow Keys To Move Cursor; Press ENTER to Select

Figure 5 SAMPLE SCREEN

An assessment of the kinds of information required by the system revealed that there were basically two different types of information to represent: 1) that kind of data that is common to all studies (i.e., author, title) and 2) data that are specific to a particular classification of studies. Note development of the IVHS is described in Step 6.

The first kind of data can be determined by human factors engineers based on what information was contained in the human performance data base and what kinds of things they thought would be useful in the data base. A needs analysis was done and new fields for the data base were recommended (Pugliese, 1986).

The second classification of data is much more complex since it must be a well defined taxonomy that can be used to classify all studies of interest in the field of Human Factors. It was decided that an Independent Variable Hierarchy Structure (IVHS) that had been developed over several years at Calspan would provide the general template for this taxonomy (Domminesey, 1986; Howard, 1986). Note development of the IVHS is described in Step 5.

This need to represent two types of data led to using two logically separate but related data bases. This avoids the need to have records with unused fields and provides a convenient way of organizing the data base along the lines of the established IVHS. One main data base is used (see Table 6), establishing a unique study-number for each record. For all the other fields that will make up the main data base system, the field names presented on the screen to the user will be identical to those used in programming. The data that are common to different classes of studies, as determined by the IVHS, are represented in separate, but related data bases, referred to as "Ancillary Data Bases".

Each record in the main data base has an "assocvar" field. The value of this field classifies the study as one of the elements in the IVHS. Each classification, in turn, has an associated group of attributes. These attributes are the parts of the classification to which values can be assigned. For example, a study describing the effect of noise level on performance might be assigned the IVHS class of "noise". This class then has attributes such as decibel level, frequency, and so on. These are the independent variables of the study and are the variables with which the user is most likely concerned.

We can make use of this structure by using the information contained in the assocvar field to indicate which ancillary data base contains the attributes of the independent variables from a study which must be given values. The study number, common to both data bases, tells which record in the ancillary data base relates to that study. With this design, we have separated the general information (main data base) from the information that is specific to a class of studies (ancillary data bases). We now can have studies representing a whole range of topics without having irrelevant fields in our main data base.

The ancillary data bases are also represented as relational data base tables. For convenience and understandability, each ancillary table is given the name of the classification itself. For example, the classification for "noise" is represented in a table named "noise" whose fields are the assigned attributes (independent variables) associated with noise. The structure for these ancillary data bases is given by the IVHS. An example of a data base representing studies on noise is presented in Table 7.

Human is currently hosted on an IBM XT at Calspan headquarters and is available for use in combat modeling.

Step 5 Human's IVHS

The importance of Human's IVHS in streamlining user queries has been described in the previous step. Its development is described in this final evolutionary step. Human's IVHS was developed in a three task process: 1) review existing taxonomies, 2) add independent variables used in Human, and 3) remove redundancy

Table 6
HUMAN MAIN DATA BASE FIELDS

FIELD NAME	LENGTH	DESCRIPTION
study_number	integer	Unique number that identifies study
title	160 char	Title of study
author	80 char	Name of author(s) of study
reference	80 char	Source of study (publication)
keywords	80 char	Words to aid search
abstract1	240 char	A summary of the study and findings
performing_agency	80 char	Agency where research was performed
sponsoring_agency	80 char	Agency funding research
number_subjects	10 char	Number of subjects in experiment
payment_received	40 char	Compensation (if any) to subjects
subject_description	160 char	Brief description of subject population
subject_age	40 char	Age range of subjects
subject_gender	40 char	Gender mix of subjects
environment	160 char	Environment in which experiment was performed
control	160 char	Control devices (if any) used
display	160 char	Displays (if any) used
task	80 char	Brief description of task performed
interstimulus_interval	240 char	Time between stimulus presentation in seconds
trial_duration	20 char	Length of each trial period in minutes
number-of_trials	10 char	Number of trials per subject
inter_rest_interval	20 char	Length of rest between trials in minutes
training	40 char	Experience or training level of subjects
days_on_task	20 char	Number of continuous days on task
page_number	40 char	Page where data are best summarized
data_format	40 char	Format of data in study (table, etc.)
experiment_number	40 char	Number of experiment referenced (if multiple)
dependent_variable	80 char	Dependent variable of study
independent_variable_1	80 char	First independent variable of study
limits_1	80 char	Range of values of first independent variable
number_of_stimuli	10 char	Number of stimuli per trial
independent_variable_2	80 char	Second independent variable of study
limits_2	80 char	Range of values of second independent variable
mean_expression	80 char	Arithmetic expression to predict mean
mean_fit	80 char	Curve fit R-squared value for mean scores
standard_deviation_expression	80 char	Arithmetic expression to predict standard deviation
standard_deviation_fit	80 char	Curve fit R-squared value for standard deviation scores
distribution	80 char	Curve distribution
ancillary_data_base_name	40 char	Name of ancillary data base

Table 7
NOISE ANCILLARY DATA BASE FIELDS

FIELD NAME	LENGTH	DESCRIPTION
study_number	integer	Unique number that identifies study
medium	20 char	Substance through which sound traveled
range_dba	40 char	Range of loudness studied in dBA
range_hz	40 char	Range of frequencies tested in Hz
duration	40 char	Duration of stimulus presented in minutes

and ambiguity. This process and the resultant taxonomy are described in further detail in the following sections.

Task 1 - Reviewing existing taxonomies. There were over 30 existing taxonomies. A description of their strengths and weaknesses follows. Miller (1962) described a taxonomy based on the behavior associated with performance of a task. A similar but more detailed taxonomy was developed by Berliner, Angell, and Shearer (1964, cited in Fleishman and Quaintance, 1984). Some of the task characteristics in both taxonomies were covert behaviors, for example, "short-term retention" in Miller's taxonomy and "analyzes" in the Berliner, et al. taxonomy. Such behaviors can be inferred but not directly observed. Further, neither taxonomy made provision for incorporating the effect of the environment on behavior.

In contrast, the Fitts taxonomy (Fitts, 1965) is based on elements required for task performance, including the task environment. The Fitts taxonomy is simple but lacks depth, i.e., it names only the elements in the top level of the hierarchy. Also, it places heavy emphasis on the human in the system (for example, Fitts assigned three elements to the operator) in contrast to other system elements. Finally, stimuli (e.g., letters, tones, and colors) were not included in the taxonomy (although they are implied in the display branch) and yet stimulus characteristics have a major influence on human performance. Blanchard (1973) compensated for Fitts's lack of depth by developing a very detailed taxonomy for visual and auditory displays.

Hoffman, Imwold, and Koller (1983) revised Fitts's (1965) taxonomy based on differential learning data. Their taxonomy was extremely limited but did identify the importance of interrelationships among elements. In independent work, Willis (1961) developed an input-output taxonomy which provided depth for a part of Fitts's taxonomy but not for others. Chambers (1973) expanded the environment element.

Harrow's (1972) taxonomy was restricted to the psychomotor domain. This taxonomy is extremely cumbersome to use and requires considerable learning - thus making it a poor candidate to structure an expert system that would be used extensively by people with little or no background in motor movement. But elements of Harrow's taxonomy could be used to expand Fitts's original framework, especially in the area of perception.

Hindmarch (1980) developed a taxonomy to classify the effects of psychoactive drugs. It was very similar to Fitts' taxonomy but had greater depth. It is that depth, however, that has created greater ambiguity. For example, what is the difference between "perception" and "detection" or between "processing" and "integration"? Further, Hindmarch's breakdown of the Central Nervous System (CNS) was cursory. Gagne (1974) provided a more detailed breakdown.

Shingledecker, Crabtree, and Acton (1982) developed a taxonomy for use in identifying workload in a fighter aircraft. It includes very specific tasks, e.g., "flight decision assessment". Meyer, Laveson, Pape, and Edwards (1978) developed a different taxonomy for the same environment. Carter (1986) developed another special purpose taxonomy; this one for computer functions. One of the most detailed taxonomies to date was developed by Meister (cited in Fleishman and Quaintance, 1984). His taxonomy allows comparison of tasks on multiple dimensions. These dimensions are easily distinguishable even to the nonsophisticated user. McCormick (1979) added a category "Relationships with other persons" since he felt this was critical in describing a task. Fine (1974) broke this element down even further.

A trilogy of taxonomies was reported in Fleishman, Teichner, and Stephenson (1970). The first (Theologus, Romashko, and Fleishman, 1969) is a taxonomy of human abilities. The second (Farina and Wheaton, 1970, cited in Fleishman, Teichner, and Stephenson, 1970) is a taxonomy of tasks. The last (Teichner, 1970) is a taxonomy of tasks based on information theory.

The above taxonomies were reviewed using Companion and Corso's (1982) set of criteria. These criteria state that a taxonomy: 1) must simplify task description, 2) should be generalizable, 3) must employ user-compatible terms, 4) must be complete and internally consistent, 5) must be compatible with the theory or system being analyzed, 6) must provide a method for predicting performance, 7) must have practical utility, 8) must be cost-effective, 9) must provide a framework for data, 10) should account for interactions of task characteristics and operator performance, and 11) should be applicable to all system levels. After this review, the taxonomies were combined into a single taxonomy by overlaying related elements. Specifically, whenever the same or synonymous words appeared in two existing taxonomies, they were used as the point of connection between the two taxonomies.

Task 2 - add the independent variables used in HUMAN. Further, variables that were not manipulated in a single experiment but rather described differences between studies were added to Human's IVHS. These included type of environment, gender, and age.

Task 3 - Streamline the IVHS. Redundant variables were removed, variables with ambiguous names were renamed, and variables with rarely-used names were renamed to match the most common name (i.e., used by the most studies). Finally, the resulting IVHS was reviewed by human factors engineers for clarity and completeness. On the basis of their review, new variables were added. The complete IVHS is presented in Table 8.

Human's current IVHS has 3 major branches: environment, subject, and task. These branches match the traditional breakdown of experimental variables used in human-factors textbooks. Further, each branch has both a type and attribute leaf. This structure is standard in most AI systems. Finally, the depth of the hierarchy varies tremendously between variables, for example, 3 levels of age and 8 levels for tracking tasks. This difference in depth reflects the diversity of specificity in human-factors experiments. Human's IVHS can be used to identify and categorize those independent variables that affect human performance in combat.

SUMMARY

The evolution of Human has generated six distinct tools to help the combat modeler.

1) A set of 32 FORTRAN subroutines which predict human performance in an air defense system. The complete listings of these subroutines are given in Laughery and Gawron (1983).
2) A stand-alone FORTRAN model which predicts human performance in air defense systems and in tactical fighter aircraft. The system is available from the Air Force Armstrong Aerospace Medical Research Laboratory.
3) An empirically validated algorithm to combine independent sets of human performance data:

$$E = \frac{\sum_{i=1}^{n} P_i}{n}$$

where E is the estimate of human performance; P, the performance predicted from one independent data set; and n, the number of independent data sets.

4) A human performance consultant system, Human, which predicts performance. A prototype version of Human is available for use at Calspan.
5) A hierarchy of independent variables that affect human performance. This IVHS is presented in Table 8.

And Human evolution continues . . .

Table 8 HUMAN'S IVHS

- 1 Environment
  - 1.1 Type
    - 1.1.1 Laboratory
    - 1.1.2 Office
    - 1.1.3 Outer Space
  - 1.2 Attributes
    - 1.2.1 Acceleration
    - 1.2.2 Confinement
    - 1.2.3 Contaminants/Toxicants
    - 1.2.4 Day/Night Cycles
    - 1.2.5 Electricity
    - 1.2.6 Isolation
    - 1.2.7 Lighting
      - 1.2.7.1 Type
        - 1.2.7.1.1 Flourescent
        - 1.2.7.1.2 Incandescent
        - 1.2.7.1.3 Sunlight
      - 1.2.7.2 Attributes
        - 1.2.7.2.1 Luminance in Foot Lamberts
    - 1.2.8 Magnetism
    - 1.2.9 Noise
      - 1.2.9.1 Duration
        - 1.2.9.1.1 Continuous
        - 1.2.9.1.2 Impulsive
        - 1.2.9.1.3 Intermittent
        - 1.2.9.1.4 Single
      - 1.2.9.2 Frequency
        - 1.2.9.2.1 Constant
        - 1.2.9.2.2 Variable
      - 1.2.9.3 Intensity
        - 1.2.9.3.1 Constant
        - 1.2.9.3.2 Variable
      - 1.2.9.4 Medium
        - 1.2.9.4.1 Atmosphere
        - 1.2.9.4.2 Communication
        - 1.2.9.4.3 Hydrosphere
      - 1.2.9.5 Range
        - 1.2.9.5.1 Infrasonic
        - 1.2.9.5.2 Sonic
        - 1.2.9.5.3 Ultrasonic
      - 1.2.9.6 Spectrum
        - 1.2.9.6.1 Broad Band
        - 1.2.9.6.2 Narrow Band
        - 1.2.9.6.3 Pure

- 1.2.10 Pressure
    - 1.2.10.1 Ambient Vapor Pressure in MB
    - 1.2.10.2 Gravity
- 1.2.11 Radiation
    - 1.2.11.1 Infrared
    - 1.2.11.2 Microwave
    - 1.2.11.3 Radio Frequency
    - 1.2.11.4 Ultraviolet
    - 1.3.11.5 Visible
    - 1.3.11.6 X-Ray
- 1.2.12 Reduced/Zero Gravity
- 1.2.13 Terrain
- 1.2.14 Thermal
    - 1.2.14.1 Ambient Dry Bulb Temperature in Degrees Celsius
    - 1.2.14.2 Humidity
- 1.2.15 Vibration

- 2 Subject
    - 2.1 Physical Characteristics
        - 2.1.1 Age
        - 2.1.2 Effector
            - 2.1.2.1 Feet
                - 2.1.2.1.1 Agility
                - 2.1.2.1.2 Dominance
                - 2.1.2.1.3 Lift Strength
            - 2.1.2.2 Hands
                - 2.1.2.2.1 Dominance
                - 2.1.2.2.2 Flexibility
                - 2.1.2.2.3 Grip Strength
            - 2.1.2.3 Voice
        - 2.1.3 Fatigue
        - 2.1.4 Gender
            - 2.1.4.1 Female
            - 2.1.4.2 Male
        - 2.1.5 Height in Cm
        - 2.1.6 Limbs
            - 2.1.6.1 Legs
                - 2.1.6.1.1 Endurance
                - 2.1.6.1.2 Strength
            - 2.1.6.2 Arms
                - 2.1.6.2.1 Length
        - 2.1.7 Weight in Kg
    - 2.2 Mental State
        - 2.2.1 Attention Span
        - 2.2.2 Drugs
            - 2.2.2.1 Type
            - 2.2.2.2 Attributes

            - 2.2.2.2.1 Dosage
            - 2.2.2.2.2 Number of Days Since Last Taken
            - 2.2.2.2.3 Number of Days Taken
    - 2.2.3 Memory
        - 2.2.3.1 Long Term
            - 2.2.3.1.1 Number of Times Has Done Task Before
        - 2.2.3.2 Short Term
            - 2.2.3.2.1 Number of Items Stored
    - 2.2.4 Personality Trait
        - 2.2.4.1 Perceived Probability of Success
    - 2.2.5 Sleep in Hours
    - 2.2.6 Work Schedule
        - 2.2.6.1 Days on Duty
        - 2.2.6.2 Rest Periods
            - 2.2.6.2.1 Duration
            - 2.2.6.2.2 Frequency
- 2.3 Senses
    - 2.3.1 Auditory
        - 2.3.1.1 Acuity
        - 2.3.1.2 Biaural
        - 2.3.1.3 Monaural
        - 2.3.1.4 Tone Perception
    - 2.3.2 Olfactory
    - 2.3.3 Tactual
    - 2.3.4 Vision
        - 2.3.4.1 Accommodation
        - 2.3.4.2 Acuity
        - 2.3.4.3 Binocular
        - 2.3.4.4 Color Perception
        - 2.3.4.5 Convergence
        - 2.3.4.6 Monocular

3 Task

- 3.1 Control Device
    - 3.1.1 Type
        - 3.1.1.1 Knobs
        - 3.1.1.2 Levers
        - 3.1.1.3 Pedal
        - 3.1.1.4 Pushbutton
        - 3.1.1.5 Switch
            - 3.1.1.5.1 Rocker
            - 3.1.1.5.2 Rotary Selector
            - 3.1.1.5.3 Toggle
        - 3.1.1.6 Track Ball
        - 3.1.1.7 Touch Device
            - 3.1.1.7.1 Keyboard
                - 3.1.1.7.1.1 Membrane

        - 3.1.1.7.1.2 Teletype
      - 3.1.1.7.2 Light Pen
      - 3.1.1.7.3 Pointer
      - 3.1.1.7.4 Touch Panel
      - 3.1.1.7.5 Touch Screen
    - 3.1.1.8 Voice Activated
    - 3.1.1.9 Wheels
      - 3.1.1.9.1 Steering Wheels
      - 3.1.1.9.2 Thumb Wheels
  - 3.1.2 Attributes
    - 3.1.2.1 Number of Positions
    - 3.1.2.2 Size
    - 3.1.2.3 Type Damping
    - 3.1.2.4 Type Feedback
- 3.2 Display Device
  - 3.2.1 Type
    - 3.2.1.1 Auditory Displays
      - 3.2.1.1.1 Electromechanical
        - 3.2.1.1.1.1 Bells
        - 3.2.1.1.1.2 Buzzers
        - 3.2.1.1.1.3 Horns
        - 3.2.1.1.1.4 Sirens
      - 3.2.1.1.2 Electronic
        - 3.2.1.1.2.1 Electronic Tones and Signals
        - 3.2.1.1.2.2 Recorded Signals Directions
    - 3.2.1.2 Visual
      - 3.2.1.2.1 CRT Alphanumeric-Pictorial Displays
        - 3.2.1.2.1.1 Computer Output Displays
        - 3.2.1.2.1.2 Infrared Sensor Displays
        - 3.2.1.2.1.3 Low-Light-Level TV Displays
        - 3.2.1.2.1.4 Television Output Displays
      - 3.2.1.2.2 CRT Electronic Parameter Displays
        - 3.2.1.2.2.1 Analog Computer Output Displays
        - 3.2.1.2.2.2 Bargraph Displays
        - 3.2.1.2.2.3 Waveform Displays
      - 3.2.1.2.3 CRT Spatial Relation Displays
        - 3.2.1.2.3.1 Radar Displays
        - 3.2.1.2.3.2 Sonar Displays
      - 3.2.1.2.4 Hard Copy Readout Displays
        - 3.2.1.2.4.1 Plotters
        - 3.2.1.2.4.2 Printers
        - 3.2.1.2.4.3 Recorders
      - 3.2.1.2.5 Indicator Lights (Transilluminated)
        - 3.2.1.2.5.1 Lighted Pushbutton Displays
        - 3.2.1.2.5.2 Multiple Status
        - 3.2.1.2.5.3 Single Status

3.2.1.2.6 Light Emitting Diode (LED)
3.2.1.2.7 Liquid Crystal Displays (LCD)
3.2.1.2.8 Mechanical
3.2.1.2.9 Projection
3.2.1.2.10 Random-Access Digital Readouts
3.2.1.2.10.1 Back-Lighted Belt Displays
3.2.1.2.10.2 Cold Cathode Tubes
3.2.1.2.10.3 Edge-Lighted Plates
3.2.1.2.10.4 Light-Emitting Diode Displays
3.2.1.2.10.5 Projection Readouts
3.2.1.2.10.6 Segmented Matrices
3.2.1.2.11 Scalar Displays
3.2.1.2.11.1 Fixed Pointer, Moving Scale
3.2.1.2.11.2 Moving Pointer, Fixed Scale
3.2.1.2.12 Sequential Access Digital Readouts
3.2.1.2.12.1 Electromechanical Drum Counters
3.2.1.2.12.2 Flag Counters
3.2.1.2.13 Status Displays
3.2.1.2.13.1 Large Screen Displays
3.2.1.2.13.2 Map Displays
3.2.1.2.13.3 Matrix Boards
3.2.1.2.13.4 Plot Boards
3.2.1.2.13.5 Projected Displays
3.2.2 Attributes
3.2.2.1 Size
3.2.2.1.1 Diameter in Cm
3.2.2.1.2 Height in Cm
3.2.2.1.3 Width in Cm
3.2.2.2 Viewing Conditions
3.2.2.2.1 Collimation
3.2.2.2.2 Distance of Operator to Display in Cm
3.2.2.2.3 Magnification
3.2.2.2.4 Ocular Design
3.2.2.2.4.1 Binocular
3.2.2.2.4.2 Dichoptic
3.2.2.2.4.3 Monocular Left Eye
3.2.2.2.4.4 Monocular Right Eye
3.2.2.2.5 Resolution
3.2.2.2.6 Visual Angle/Field of View in Degrees
3.3 Machine
3.3.1 Computer
3.3.1.1 Mainframe
3.3.1.2 Personal
3.3.2 Vehicles
3.3.2.1 Aircraft
3.3.2.1.1 Helicopter

      - 3.3.2.1.2 Jet
      - 3.3.2.1.3 Propeller
    - 3.3.2.2 Motorized Ground Vehicle
      - 3.3.2.2.1 Car
      - 3.3.2.2.2 Half Track
      - 3.3.2.2.3 Jeep
      - 3.3.2.2.4 Tank
      - 3.3.2.2.5 Truck
    - 3.3.2.3 Ship
      - 3.3.2.3.1 Aircraft Carrier
      - 3.3.2.3.2 Destroyer
      - 3.3.2.3.3 Submarine
    - 3.3.2.4 Spacecraft
  - 3.3.3 Weapon
- 3.4 Stimulus
  - 3.4.1 Type
    - 3.4.1.1 Auditory
    - 3.4.1.2 Kinesthetic
    - 3.4.1.3 Visual
      - 3.4.1.3.1 Alphanumeric
      - 3.4.1.3.2 Graph
  - 3.4.2 Attributes
    - 3.4.2.1 Background
      - 3.4.2.1.1 Complexity
      - 3.4.2.1.2 Contrast
      - 3.4.2.1.3 Number of Background Characters
    - 3.4.2.2 Characteristics
      - 3.4.2.2.1 Alphanumerics
      - 3.4.2.2.2 Changing/Moving Stimulus
      - 3.4.2.2.3 Coded Stimulus
      - 3.4.2.2.4 Conspicuity
      - 3.4.2.2.5 Raw Stimulus
      - 3.4.2.2.6 Static Stimulus
    - 3.4.2.3 Color
    - 3.4.2.4 Duration
      - 3.4.2.4.1 Continuous
      - 3.4.2.4.2 Intermittent
        - 3.4.2.4.2.1 Probability
        - 3.4.2.4.2.2 Rate
      - 3.4.2.4.3 Single
    - 3.4.2.5 Information Presented
      - 3.4.2.5.1 Content
      - 3.4.2.5.2 Qualitative
      - 3.4.2.5.3 Quantitative
    - 3.4.2.6 Location on Display
      - 3.4.2.6.1 Center

- 3.4.2.6.2 Lower Left
- 3.4.2.6.3 Lower Middle
- 3.5.2.6.4 Lower Right
- 3.4.2.6.5 Middle Left
- 3.4.2.6.6 Middle Right
- 3.4.2.6.7 Upper Left
- 3.4.2.6.8 Upper Middle
- 3.4.2.6.9 Upper Right
- 3.4.2.6.10 Predictability of Location
- 3.4.2.7 Mechanism
    - 3.4.2.7.1 Directly Viewed Event
    - 3.4.2.7.2 Display
    - 3.4.2.7.3 Written Material
- 3.4.2.8 Number
    - 3.4.2.8.1 Multiple
    - 3.4.2.8.2 Single
- 3.4.2.9 Range of Values
- 3.4.2.10 Relative Movement
    - 3.4.2.10.1 Observer and Target at Rest
    - 3.4.2.10.2 Observer and Target in Motion
    - 3.4.2.10.3 Observer in Motion, Target at Rest
    - 3.4.2.10.4 Observer at Rest, Target in Motion
- 3.4.2.11 Relative Position of Observer
    - 3.4.2.11.1 Horizontal Range in Km
    - 3.4.2.11.2 Offset in Km
    - 3.4.2.11.3 Positions
        - 3.4.2.11.3.1 Air-to-Air
        - 3.4.2.11.3.2 Air-to-Ground
        - 3.4.2.11.3.3 At a Display
        - 3.4.2.11.3.4 Ground-to-Air
        - 3.4.2.11.3.5 Ground-to-Ground
- 3.4.2.12 Size/Amplitude
    - 3.4.2.12.1 Signal-to-Noise Ratio

- 3.5 Task
    - 3.5.1 Type
        - 3.5.1.1 Communication
            - 3.5.1.1.1 Type
                - 3.5.1.1.1.1 Advise
                - 3.5.1.1.1.2 Answer
                - 3.5.1.1.1.3 Communicate
                    - 3.5.1.1.1.3.1 Job-Related
                    - 3.5.1.1.1.3.2 Public-Related
                - 3.5.1.1.1.4 Comprehend
                - 3.5.1.1.1.5 Coordinate
                - 3.5.1.1.1.6 Direct
                - 3.5.1.1.1.7 Indicate

            - 3.5.1.1.1.8 Inform
            - 3.5.1.1.1.9 Instruct
            - 3.5.1.1.1.10 Request
            - 3.5.1.1.1.11 Supervise
            - 3.5.1.1.1.12 Transmit
        - 3.5.1.1.2 Attributes
            - 3.5.1.1.2.1 Oral
            - 3.5.1.1.2.2 Written
    - 3.5.1.2 Meditation
        - 3.5.1.2.1 Type
            - 3.5.1.2.1.1 Information Processing
                - 3.5.1.2.1.1.1 Categorize
                - 3.5.1.2.1.1.2 Calculate
                - 3.5.1.2.1.1.3 Code
                - 3.5.1.2.1.1.4 Compute
                - 3.5.1.2.1.1.5 Interpolate
                - 3.5.1.2.1.1.6 Itemize
                - 3.5.1.2.1.1.7 Learn
                - 3.5.1.2.1.1.8 Tabulate
                - 3.5.1.2.1.1.9 Translate
            - 3.5.1.2.1.2 Problem Solving and Decision Making
                - 3.5.1.2.1.2.1 Analyze
                - 3.5.1.2.2.2.2 Deduce
                - 3.5.1.2.2.2.3 Induce
                - 3.5.1.2.1.2.4 Calculate
                - 3.5.1.2.1.2.5 Choose
                    - 3.5.1.2.1.2.5.1 Choose from Known Alternatives
                    - 3.5.1.2.1.2.5.2 Choose from Unknown Alternatives
                    - 3.5.1.2.1.2.5.3 Choose from Unspecified Alternatives
                - 3.5.1.2.1.2.6 Compare
                    - 3.5.1.2.1.2.6.1 Order
                - 3.5.1.2.1.2.7 Compute
                - 3.5.1.2.1.2.8 Estimate
                - 3.5.1.2.1.2.9 Integrate
                - 3.5.1.2.1.2.10 Plan
                - 3.5.1.2.1.2.11 Supervise
            - 3.5.1.2.1.3 Recall
                - 3.5.1.2.1.3.1 Recall Facts
                - 3.5.1.2.1.3.2 Recall Principles

- 3.5.1.2.1.3.3 Recall Procedures
- 3.5.1.2.1.3.4 Timeshare

- 3.5.1.2.2 Attributes
  - 3.5.1.2.2.1 Complexity
  - 3.5.1.2.2.2 Difficulty
- 3.5.1.3 Motor Processes
  - 3.5.1.3.1 Type
    - 3.5.1.3.1.1 Complex–Continuous
      - 3.5.1.3.1.1.1 Adjust
      - 3.5.1.3.1.1.2 Align
      - 3.5.1.3.1.1.3 Insert Object
      - 3.5.1.3.1.1.4 Regulate
      - 3.5.1.3.1.1.5 Remove Object
      - 3.5.1.3.1.1.6 Synchronize
      - 3.5.1.3.1.1.7 Track
        - 3.5.1.3.1.1.7.1 Visual Tracking Only
        - 3.5.1.3.1.1.7.2 Visual Tracking Plus Position Plotting
      - 3.5.1.3.1.1.8 Type Message on Keyboard
      - 3.5.1.3.1.1.9 Write
    - 3.5.1.3.1.2 Compound
    - 3.5.1.3.1.3 Reflex
      - 3.5.1.3.1.3.1 Intersegmental
      - 3.5.1.3.1.3.2 Segmental
      - 3.5.1.3.1.3.3 Suprasegmental
    - 3.5.1.3.1.4 Simple–Discrete
      - 3.5.1.3.1.4.1 Activate
      - 3.5.1.3.1.4.2 Close
      - 3.5.1.3.1.4.3 Connect
      - 3.5.1.3.1.4.4 Disconnect
      - 3.5.1.3.1.4.5 Join
      - 3.5.1.3.1.4.6 Move
      - 3.5.1.3.1.4.7 Press
      - 3.5.1.3.1.4.8 Set
      - 3.5.1.3.1.4.9 Turn Single Rotary Control
  - 3.5.1.3.2 Attributes
    - 3.5.1.3.2.1 Ballistic
    - 3.5.1.3.2.2 Continuous
    - 3.5.1.3.2.3 Coordinated
    - 3.5.1.3.2.4 Fine
    - 3.5.1.3.2.5 Gross
    - 3.5.1.3.2.6 Repetitive

- 3.5.1.3.2.7 Serial
- 3.5.1.3.2.8 Static
- 3.5.1.4 Perceptual Processing
  - 3.5.1.4.1 Searching for and Receiving Information
    - 3.5.1.4.1.1 Detect
      - 3.5.1.4.1.1.1 Detect Nonverbal Cues
      - 3.5.1.4.1.1.2 Detect Verbal Cues
    - 3.5.1.4.1.2 Inspect
    - 3.5.1.4.1.3 Observe
    - 3.5.1.4.1.4 Read
    - 3.5.1.4.1.5 Receive
    - 3.5.1.4.1.6 Scan
    - 3.5.1.4.1.7 Survey
  - 3.5.1.4.2 Identifying Objects, Actions, Events
    - 3.5.1.4.2.1 Discriminate
      - 3.5.1.4.2.1.1 Discriminate Auditory Cues
      - 3.5.1.4.2.1.2 Discriminate Kinetic Cues
      - 3.5.1.4.2.1.3 Discriminate Tactile Cues
      - 3.5.1.4.2.1.4 Discriminate Nonverbal Cues
      - 3.5.1.4.2.1.5 Discriminate Verbal Cues
      - 3.5.1.4.2.1.6 Discriminate Visual Cues
    - 3.5.1.4.2.2 Identify
      - 3.5.1.4.2.2.1 Identify Nonverbal Cues
      - 3.5.1.4.2.2.2 Identify Verbal Cues
    - 3.5.1.4.2.3 Recognize
      - 3.5.1.4.2.3.1 Recognize Nonverbal Cues
      - 3.5.1.4.2.3.2 Recognize Verbal Cues

- 3.5.2 Attributes
  - 3.5.2.1 Amount of Labor Required
  - 3.5.2.2 Complexity
  - 3.5.2.3 Degree of Response Chaining
  - 3.5.2.4 Difficulty
  - 3.5.2.5 Knowledge of Results
  - 3.5.2.6 Output

3.5.2.7 Pacing
3.5.2.8 Precision
3.5.2.9 Repetitiveness
3.5.2.10 Skill Demands
3.5.2.11 Simultaneity of Responses
3.5.2.12 Task Autonomy

REFERENCES

1. BEIDEMAN, L.R., GOMER, F.E., and LEVINE, S.H. (1980). Dynamic target acquisition: Empirical models of operator performance (Report TR-80-1177). Bolling Air Force Base, D.C.: Air Force Office of Scientific Research.
2. BLANCHARD, R.E. (April, 1973). Requirements, concept and specification for a Navy human performance data store (Report No. 102-2). Santa Monica, CA: Behaviormetrics.
3. CARTER, J.A., JR. (1986). A taxonomy of user-oriented functions. International Journal of Man-Machine Studies, 24, 195-292.
4. CHAMBERS, A.N. (1973). Development of a taxonomy of human performance: A heuristic model of the development of classification systems. JSAS Catalog of Selected Documents in Psychology, 3, 24-25.
5. COMPANION, M.A. and CORSO, G.M. (1982). Task taxonomies: a general review and evaluation. International Journal of Man-Machine Studies, 17, 459-472.
6. DATE, C.J. (1982). A Guide To DB2. Addison-Wesley.
7. DATE, C.J. (1985). An Introduction to Database Systems (Third Edition). Addison-Wesley.
8. DUKET, S.D., et al. Documentation for the SAINT simulation program. Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory, AMRL-TR-77-63, June 1978.
9. DOMMINESEY, M. (1986). The Hierarchy of Independent Variables for the Human Performance Model - HUMAN (Internship Report). State University of New York at Buffalo.
10. FINE, S.A. (1974). Functional job analysis: An approach to a technology for manpower planning. Personnel Journal, 53, 813-818.
11. FITTS, P.M. (1965). Factors in complex skill training. In R. Glaser (Ed.) Training research and education. New York: Wiley & Sons.
12. FLEISHMAN, E.A. and QUAINTANCE, M.K. (1984). Taxonomies of human performance. New York: Academic Press.
13. FLEISHMAN, E.A., TEICHNER, W., and STEPHENSON, R.W. (1970). Development of a taxonomy of human performance (AIR-726 -1/70-TPR2). Washington, D.C.: Advanced Research Projects Agency.
14. GAGNE, R.M. (1974). Essentials of learning for instruction. New York: Holt, Rinehart, and Winston.
15. GAWRON, V.J. (1987). Evaluation of Algorithms For Combining Independent Data Sets In A Human Performance Expert System. (Master's Thesis). State University of New York at Buffalo.
16. HARROW, A.J. (1972). A taxonomy of the psychomotor domain. New York: David McKay Company.
17. HINDMARCH, I. (1980). Psychomotor function and psychoactive drugs. Journal of Clinical Pharmacology, 10, 189-209.
18. HOFFMAN, S.J., IMWOLD, C.H., and KOLLER, J.A. (1983). Accuracy and prediction in throwing: A taxonomic analysis of children's performance. Research Quarterly for Exercise and Sport, 54, 33-40.
19. HOWARD, M. (1986). Design of the Prototype of Calspan's Expert System (Internship Report). State University of New York at Buffalo.
20. HOWELL, W.C., JOHNSTON, W.A., and GOLDSTEIN, I.L. (1966). Complex monitoring and its relation to the classical problems of vigilance. Organizational Behavior and Human Performance, 1, 129-150.
21. ITZER, R. (1985) User-System Interface Analysis of the Human Performance Model - HUMAN. (Internship Report). State University of New York at Buffalo.
22. KOWALSKI, R. (1979). Logic for problem solving. New York: American Elsevier.

23. LAUGHERY, K.R. and DITZIAN, J.L. (1982, September). MOPADS literature review. (MOPADS 5.1). Fort Bliss, TX: Army Research Institute.

24. LAUGHERY, K.R. and GAWRON, V.G. (1983, June). MOPADS skill moderator function subroutines. (MOPADS 5.10 DF). Fort Bliss, TX: Army Research Institute.

25. MCCORMICK, E.J. (1979). Job analysis: Methods and applications. New York: American Management Association.

26. MEYER, R. P., LAVESON, J.I., PAPE, G.L., and EDWARDS, B. (1978, September). Development and application of a task taxonomy for tactical flying. (AFHRL-TR-78-42). Brooks Air Force Base, TX: Human Resources Laboratory.

27. MILLER, R.B. (1962). Task description and analysis. In R.M. Gagne (Ed.) Psychological principles in system development. New York: Holt, Rhinehart, & Winston, Chapter 6.

28. PUGLIESE, A. (1986). *Design of Data Base for the Human Expert System* (Internship Report) State University of New York at Buffalo.

29. QUINN, T.J., GAWRON, V.J., CODY, W.J., ARBAK, C.J., & ROUSE, W.B. (1986, May). Cockpit Automation Technology (CAT) Final Report (AAMRL-TR-86). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

30. ROUSELL, P. (1975). Prolog: Manuel de Reference et d'Utilisation. Group d'Intelligence Artificielle, Universite de Marseille-Luminy.

31. SHINGLEDECKER, C.A., CRABTREE, M.S. & ACTON, W.H. (1982). Standardized tests for the evaluation and classification of workload metrics. Proceedings of the Human Factors Society - 26th Annual Meeting, 648-651.

32. TEICHNER, W.H. (1970). An information-theoretic approach to task classification. In E.A. Fleishman, W. Teichner, and R.W. Stephenson, Development of a taxonomy of human performance (AIR-726-1/70-TPR2). Washington, D.C.: Advanced Research Projects Agency.

33. THEOLOGUS, G.C., ROMASHKO, T., and FLEISHMAN, E.A. (1969). Classification in terms of human abilities. In E.A. Fleishman, W. Teichner, & R.W. Stephenson, Development of a taxonomy of human performance (AIR-726-1/70-TPR2). Washington, D.C.: Advanced Research Projects Agency.

34. TRAVALE, D. and GAWRON, V.J. (1986). Estimating Operator Performance In NBC Conditions, Proceedings of the 10th Symposium of Psychology in the Department of Defense, 148-151.

35. WILLIS, M.P. (1961). Deriving training device implications from learning theory principles (NAVTRADEVCEN 784-1). Port Washington, NY: USNTDC.

ACKNOWLEDGMENT

The MOPADS portion of this work was supported by the U.S. Army Research Institute (contract number MDA903-81-C-AA06). Pritsker & Associates, Inc. was the prime contractor; ARVIN/CALSPAN, the subcontractor. The views, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

The CAT portion of this work was supported by the U.S. Air Force Armstrong Aerospace Medical Research Laboratory (contract number F33615-84-R-0500). McDonnell Douglas Corporation was the prime contractor; ARVIN/CALSPAN, the subcontractor.

The evaluation of the combining algorithms and development of the prototype Human were funded by ARVIN/CALSPAN through IR&D 5502 and supervision of industrial engineering and computer science interns. The computer science internship included a master's project to develop software for the prototype of Human.

DISCUSSION OF "HUMAN PERFORMANCE MODELS AND POTENTIAL APPLICATIONS TO COMBAT SIMULATION"

by G. McMillan and E. Martin

and

"A HUMAN PERFORMANCE CONSULTANT SYSTEM"

by V. Gawron, D. Travale and J. Neal

DISCUSSANT: Norman Lane, Essex Corporation

Both of the preceding papers are summaries of very wide-ranging complex efforts. McMillan/Martin is an overview of a distillation of many meetings and technical sessions growing out of a NATO Research Study Group on operator modeling. Their paper presents a "taxonomic" outline of an extensive reference book emerging from the work of that group, and describes most of the main families of models that are in use today. Gawron and her coauthors provide an evolutionary history of a program that began as a straightforward modeling effort and emerged six years later as a sophisticated "expert consultant" for use by system designers. Both papers deal with mainstream examples of what are commonly referred to as "human factors engineering" (HFE) or ergonomic models. Such models are characterized largely by a focus on estimating the performance of an individual operator, often at a considerable level of detail and at a molecular task level, with an occasional excursion into group/team performance.

The HFE community is one of at least three identifiable communities that have been concerned, some for a long time, some more recently, about the impact of humans on the outcomes of systems under other-than-nominal conditions. A second community is obviously that of operations research, at least in its interests in combat models. The focus here is largely on estimate of system performance and outcomes of opposing forces at a level of aggregation usually well above the individual operator or task. A third community is biomedical research. Although some defense medical R&D is integral with HFE and other behaviorally-oriented research, there remains a significant community whose interests have ranged more toward understanding and defining the physiological effects of factors which degrade performance, rather than the direct estimation of performance changes as a result of those factors. Within this community as well, there is heightened concern about translating research data on human reactions to stressors into functional statements that can be used for operator and system modeling.

It is notable that this is the third meeting in the last three months dealing specifically with the theme of accounting for human capabilities and limitations in models that estimate more global outcomes. One was sponsored largely by the Department of Defense HFE community, one was more heavily biomedical in orientation, and the present one is cast in an operations research framework. These are encouraging convergences. We should not yet, however, be too comfortable with the idea that closing the gaps among these different disciplines is a simple thing to do. As yet, we are all still looking at different parts of the elephant and pronouncing his shape in our own image. We lack a convenient common language across communities. There has historically been only limited cross-training of practitioners (this is changing as HFE and operations research degrees share core curricula at many departments). The measures of effectiveness and performance used by the disciplines are often taken at different levels within the system. As the McMillan/Martin paper noted, HFE models, despite some powerful capabilities, have typically been restricted to the operator or team level, and to the performance of relatively small task sets. Their inputs and functions are not usually in a form convenient to incorporate in operations research models, and operations research models, in turn, rarely have architectures which can easily handle the complexity of full-blown human performance models. Medical and behavioral researchers are often reluctant to simplify their detailed data on stressor effects into the functional forms that allow them to be used in larger-scale models. There are some notable exceptions to the above generalities, but by and large they are fair statements.

Because of these divergences in emphasis and architecture, it will not be simple to completely resolve the differences among today's generation of models. The solution lies in the next generations of models and gaming efforts that are now under development. A heightened awareness in each community of what the others can provide, the forms in which information should be structured, and the limitations and appropriate use of that information is best obtained from the interactions in forums such as this one. Beyond that, there is a distinct need for joint model development efforts that take deliberate aim at the common threads within disciplines. There will likely be funding, budgeting and franchise confusions in such efforts that would challenge the wisdom of Solomon, but they remain the best hope, and are worth going after.

MODELING THE IMPACT OF CHEMICAL OPERATIONS
ON UNIT COMBAT EFFECTIVENESS

Col. Walton Phillips (retired), ORI, Inc.
LTC Jack Lombardi, U.S. Army Chemical School
Roger C. Eyler, Brunswick Corporation, Defense Division

Chemical weapons and the associated chemical agents present a threat to our troops through both the inhalation route and the percutaneous route. As a result, in order to provide full protection to the soldier he must utilize a full protective ensemble.

This ensemble includes the mask, hood, gloves, overgarment and boots (Figure 1). The ensemble is known as the mission oriented protective posture 4 (MOPP 4). The protective equipment fully encapsulates the body and, as a result, greatly inhibits individual performance in a number of ways. The eye pieces in the mask limit peripheral vision and make such tasks as sighting a weapon very awkard. The mouth piece degrades speech while the hood inhibits hearing. The mask has very efficient filters and, as a result, causes breathing resistance. The gloves are bulky and degrade those task activities that require the use of the hands. Physiologically, the total encapsulation results in an insulating effect that retards body heat dissipation.

Anyone who has ever donned the complete ensemble can surely attest that the foregoing effects produce reduced overall performance, particularly in activities associated with all types of military operations. In addition, there may be a profound psychological effect resulting from being confined in the suit. Finally, certain body needs such as eating, drinking, and other body needs must be taken care of periodically. Interest in the military impact of this protective ensemble has been existent for a long period of time. Early efforts in this regard was done by natick labs, but the first serious attempts at modeling the impact on combat operations began in the early 1960's. Over the years, numerous human factor type tests have been conducted. Also some limited troop tests have been done by the Army. Also, at the same time, numerous mathematical models have evolved which attempted to simulate the effect of losses in individual performance (as a result of the protective ensemble) upon a wide variety of combat operations. The data base that was generated was scattered over a wide variety of sources as were the mathematical models that utilized this data. No central, coordinated activity existed for obtaining validated data on combat degradation.

Recognizing this problem, the Chemical School at Fort McClellan, Alabama initiated a study titled "The Force Multiplier Study" (1) in 1982 (Figure 2). The study conducted an extensive literature review and systematically compiled a summary of pertinent data sources. The study also reviewed the mathematical models that were then employed to evaluate a variety of chemical operations. Based upon the best data and models then available, a first cut was made at evaluating the force multiplying effect of chemical operations on combined arms fighting powers. Finally, data deficiencies and data gaps were isolated and identified.

An extensive literature search produced about 700 documents which contained data or mathematical models applicable to the chemical operations problem (Figure 3). Almost all of the basic data was generated related to individual degradation. A few troop tests conducted at CEDEC and in Panama yielded some limited operational data. In no case was combat data developed that could be used to evaluate unit combat power as it pertains to the success or failure of a particular, two sided combat or combat support operation. In almost all cases, the data was too limited to establish the statistical validity of the resulting data.

A survey of the mathematical modeling effort (Figure 4) revealed that an evaluation of models had taken place over a period of 15 years. As the models were evolved or as new models were formulated, the demand for basic data grew. However, due to the fact that the data base was very limited, many modelers incorporated their own interpretation of the combat effects of chemical operations (most of which could not be validated) into their computer simulations. The models covered a (Figure 5) hierarchy of operations ranging from the Goldman model which characterized the results of heat stress on individual efficiency through small unit operations (company size, battalion size) up to theater operations as characterized by the Tacwar model. The data base and mathematical models were being used in the conduct of very important studies to support munition requirements and defensive system requirements such as alarms, individual protective equipment, collective protection and decontamination requirements. The fact that these study activities were being used to assist in important R & D and procurement decisions would dictate that analytical validity be the best that is available. Unfortunately, the limited data base prohibited conclusive model validation. This resulted because the data base consisted, mainly, of individual performance of a variety of tasks dealing with visual, audible, manual dexterity, and work rate efficiency.

The bottom line was that very little was known (Figure 6) about how to combine individual performance data with military unit operations to provide valid output data on maneuver, close combat, combat support, and combat service support effectiveness, or the resulting degradation thereof. The overall indications were that a tremendous loss on unit combat effectiveness would result from chemical operations. In particular, the synergistic effects of multiple impairments of individual performance activities upon small unit operations, which in turn impacted upon larger unit operations, was total unknown.

While data gaps existed in the data base over almost all aspects of chemical operations, the major needs were found to be centered upon combined arms operations of at least company size or greater. Since chemical operations affect the enemy forces as well as the friendly forces, two sided tests and experiments were desired. Thus, it was found that improved analysis of military requirements for chemical operations could only be obtained from combined arms troop experiments and tests. In addition, to establish the validity of the results of these tests, extensive planning must be done to provide comprehensive data collection for statistical analysis. The current Army program - CANE - specifically addresses this major and important problem area.

I will provide you with an update on a major department of the Army program called combined arms in a nuclear and chemical environment or better known by its acronym - CANE. I will give you a short overview of the program, the results and lessons learned to date, and where we are going in order to provide the Army the answers to vital questions in this area.

CANE got its start in 1980 when the Army working group, the Chemical SPR and Defense Science Board separately confirmed an urgent need to determine unknowns in NBC operations and recommended that a large scale FDTE be conducted to fill the data gaps.

The chemical action plan identified the critical deficiencies in the CW area and tasked various agencies with actions to address them (Figure 7). This is the purpose of CANE.

The key words are measured data and units. Up until the time of the CANE series of evaluations we had little or no data on units. Selected information on individuals in MOPP 4 (e.g., turning wrenches, heat stress, etc.) existed. Almost all of the unit data was subjective; no statistically sound data to support congressional and senior Army leadership decisions, training, doctrine, force structure and materiel (Figure 8).

The CANE issues were developed from input from each of the TRADOC service schools, FORSCOM, USAREUR, AMC and the medical community. These were consolidated by the Chemical School into 92 issues over six functional areas. At the time the Army went to the mission area analysis, TRADOC directed that the issues be aligned with the MAA's (Figure 9).

The 13th issue area shown here is the impact of chemical weapons on our doctrine, training, organizations and materiel based on the analysis of the other issue areas. All of the issue areas, with the exception of close combat, light, have been addressed to varying degrees. The CANE testing of CCL is scheduled for 1990 (Figure 10).

If the FDTE was to be conducted as one test, in order to meet the original timelines as conceived in 1981, it would have required the resources of over a division, an area as large as the NTC and $60M. Not only would it have required too many resources, but the Army was not able to conduct a test of this magnitude. As a result it was decided to conduct the test in three phases (Figure 11).

Phase I, the squad/platoon test was conducted in the spring of 1983 at FT Hunter Liggett by CDEC. I will show you the lessons learned from this test shortly.

Phase II, the CO/BN test could not be done in one exercise with the limited troop and testing resources available. The FORSCOM and TRADOC commanders personally decided to conduct Phase II as subtests.

Phase IIA, the company team test was conducted in the spring of 1985 at Ft. Hood, Texas. Classified results have been published and are provided to foreign governments through channels under agreements between our countries.

In order to provide the information to support these issues, it was decided that the tests supporting the CANE program would have the following characteristics (Figure 12).

By using a chemical scenario where there is no threat of NUC or Chem attack and then conducting the same scenario, with the same unit, same terrain, etc., in a NUC/Chem environment we can get a difference or "Delta" that is statistically significant. The baseline and NCE's are reversed with different units to capture any learning curve effects. The 12 hours in MOPP 4 on one day is the longest the surgeon general would allow at the temperatures expected at that time of year. During the CANE IIB we tested soldiers for up to 6-7 hours in MOPP 4 before significant events in order to fix where operational effectiveness hits its downward trend.

The scenarios were also 96 hours long in CANE IIB to allow for the additional planning and execution times for a battalion task force.

Figure 13 sketches the analytical methodology for the CANE program considered throughout the testing process; in planning, test development, execution, data processing and evaluation.

The issues, issue elements, and measure indicators in the independent evaluation plan (IEP) developed by the proponent schools were used by the tester to develop the measures of effectiveness and subsequently the pattern of analysis (PA). Our data evaluation panel reviewed the tester's plan and provided input to ensure consistency between the pattern of analysis and the independent evaluation plan.

The tester and our team were in constant communication throughout the test design process. Revisions were made in either the test design or analytical approach as approach.

We reviewed and commented on all of the data collection instruments used by the tester to ensure that the output from the evaluators would provide the correct data points. Then we ensured that the data file to be provided by the test to us for analysis was in a format that we could readily use and was consistent with the collection plan.

Based on our experience from CANE I and CANE IIA, we instituted quality control procedures that further enhanced the reliability of the data and reduced the "anomalies" that had to be resolved during the analysis.

The analytical approach and statistical procedures we used were reviewed by the analytical community. Detailed reviews were conducted by Mr. Walt Hollis and other senior ORSA's in the Army.

To date we have collected extensive and unique data on the effects on the force while we are fighting in a conflict where chemical weapons are used (Figure 14).

In order to answer the critical questions, the issue elements measures of effectiveness, test design and test scenario were all developed with the appropriate schools and centers.

The magnitude of the effort from CANE IIB alone reflects our attempt to make the most efficient use of the valuable troop resources the Army provided.

At the direction of the commandant of the Chemical School, a systematic method of evaluation CANE data was developed and implemented (Figure 15). His guidance was to insure that test results were analyzed from a combined statistical and military operational viewpoint. He stated also that if degradation did occur when a task was performed in the NCE, why was there degradation, and what are some suggested fixes? A panel was established consisting of experts on chemical warfare and mathematical analysis. A panel is tailored for each specific issue area (i.e., CCH, FS, CSS) by including panel members that are military experts in the area being evaluated. The success of the panel process is dependent on thorough and continuous coordination between the mathematical analysts and the military operational analysts from the design of the test to the completion of the IER. This joint effort began with the section of the family of models to be used to evaluate the data. A military analyst and a mathematical analyst were assigned for each issue area. They were responsible for the coordination between their respective groups during the initial statistical analysis. When the data was evaluated by the panel, these two analysts became ad hoc members of the panel.

Each variable is analyzed individually. They are also grouped into functional areas and an analysis of the groupings (functional area) is also conducted (Figure 16). For example, a group of individual items in C&C showed only small % differences in the chemical environment. When the data evaluation team had the analyst run them in groupings of R outline and scenario driven tasks we found a significant deficiency in tasks that require leader decision and reaction to changing situations.

The statistical analysis methodology was reviewed and concurred with by the analytical community, including a detailed review by Mr. Hollis, the DUSA(OR).

The output from this process is a summary evaluation report and lessons learned that are presented in a useful operational context. They provide not only the deficiency, but also the evaluation teams determination of why the deficiency occurred.

We have developed and distributed these products from CANE I and IIA and are about 75% through the process for CANE IIB.

The CANE program has provided, and will continue to provide, data to improve the chemical defensive posture of the U.S. Army. Shown are some of the results of the three tests that have been completed (Figure 17).

To insure the data is integrated into training, doctrine, force structure, models and materiel requirements, TRADOC has developed and is implementing a CANE implementation program.

The Chemical School has developed a systematic and comprehensive procedure for ensuring that NBC deficiencies in doctrine, training, organization and materiel are identified, solutions developed and action taken to correct these deficiencies. Implementation of the plan requires active participation of the proponent schools and centers whose products are affected by the deficiency (Figure 18).

This is a schematic of the key features of the decision process to determine and implement the solutions. The heart of developing the solutions comes from the evaluation panel which consists of personnel from the integrating center(s), proponent schools, doctrine, training and combat developments. Their task is to develop operationally effective solutions to the problem. Panel members are required to have the military experience and subject matter expertise necessary to address the problem. The panel reviews and evaluates the deficiencies in detail and then determines which fixes or combination of fixes are required to solve the problem. The impact of the corrective action on the force are considered from the beginning and throughout the development of the fixes.

References

(1) "Chemical Force Multiplier and Suppression Effects Study," Volume III, Appendix C "Degradation Data," Final Report, October, 1982, Secret NOFORN AD Number AD-C03592iL (over 200 references).

(2) Note - Data on CANE can be obtained (with proper clearances and need to know) by contacting Walton Phillips at AC 205-238-8132.

Basic Chemical Defense Ensemble - MOPP-4
(Mask, Head, Gloves, Overgarment, Boots)

ENSEMBLE DEGRADATION

- Visual
- Audible
- Respiratory
- Dexterity
- Heat Dissipation
- Psychological
- Body Needs
  - ✓ Drinking
  - ✓ Eating
  - ✓ Elimination

FORCE MULTIPLIER STUDY

- Literature Review
- Assemble Degradation Data
- Review Mathematical Models
- Evaluate (based on available data) Force Multiplying Effects
- Identify Data Deficiencies and Data Gaps

DATA BASE

- About 700 Documents with Applicable Data
- Data Very Limited: Sparse Statistical Basis
- Most Data On Individual Performance
  - ✓ Visual
  - ✓ Audible
  - ✓ Manual Dexterity
  - ✓ Heat Stress
- Only A Few Troop Tests:
  Combat Effectiveness Not Truly Measured

MATHEMATICAL MODELS

- Numerous Models Existed
  - ✓ Examples:
    - Goldman Model (Heat Stress)
    - AURA
    - Combat Rates Models
    - IOLOG/AMORE
    - TACWAR
  - ✓ Over 20 Models Identified Which Were Used in Chemical Analyses
  - ✓ All Attempted To Use Individual Performance Data To Assess Unit Performance

MATHEMATICAL MODELS

- Evolution Of Modeling Efforts
  - ✓ 1967 - Addendum Study To Mandrake Root
  - ✓ 1974 - Degrad Model
  - ✓ 1975 - IDA Model
  - ✓ 1977 - Calspan Model
  - ✓ 1978 - AMAF Model
  - ✓ 1979 - Scores Model
  - ✓ 1980 - SAI Logistical Model
  - ✓ 1982 - BRL AURA Model
  - ✓ Present - Army Model Improvement Program

CLOSING THE DATA GAPS

- **Major Unknowns:** Synergistic Effects
  Overall Unit Combat Power Loss

- **Major Needs:**
  Combined Arms Unit Operations
  Two Sided Tests
  Extensive Data Collection
  Statistical Validity

ARMY CHEMICAL ACTION PLAN

- Deficiency - insufficient large unit data to
  - Base concepts
  - Develop doctrine
  - Verify force structure

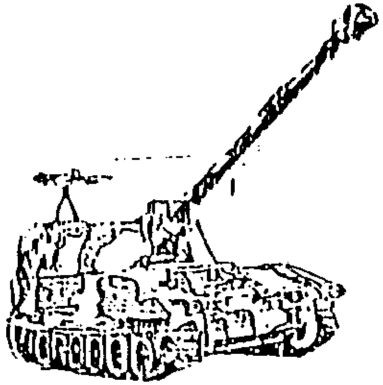
- TASKS
  - Conduct FDTE to measure effects in severe sustained integrated battlefield environment
  - Identify other planned tests or exercises which incorporate successful concepts
  - Determine requirements for follow-up studies

- ACTION: TRADOC, FORSCOM, HSC, OTEA, DAMO

FIGURE 7

PURPOSE

TO PROVIDE MEASURED DATA AND DETERMINE HOW WELL COMBAT AND SUPPORT UNITS CAN PERFORM THEIR MISSIONS IN EXTENDED OPERATIONS WHERE NUCLEAR AND CHEMICAL WEAPONS ARE EMPLOYED



ISSUE DEVELOPMENT

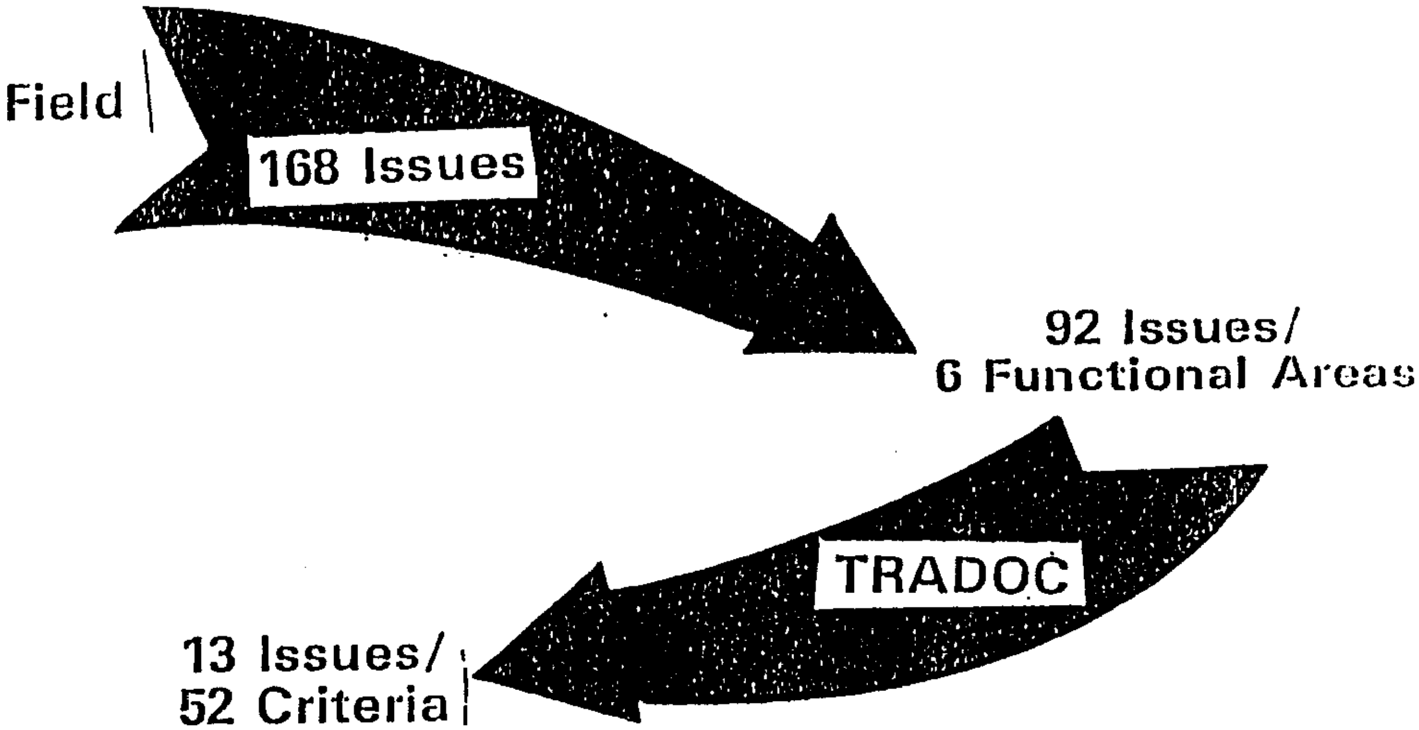


FIGURE 9

CANE ISSUES

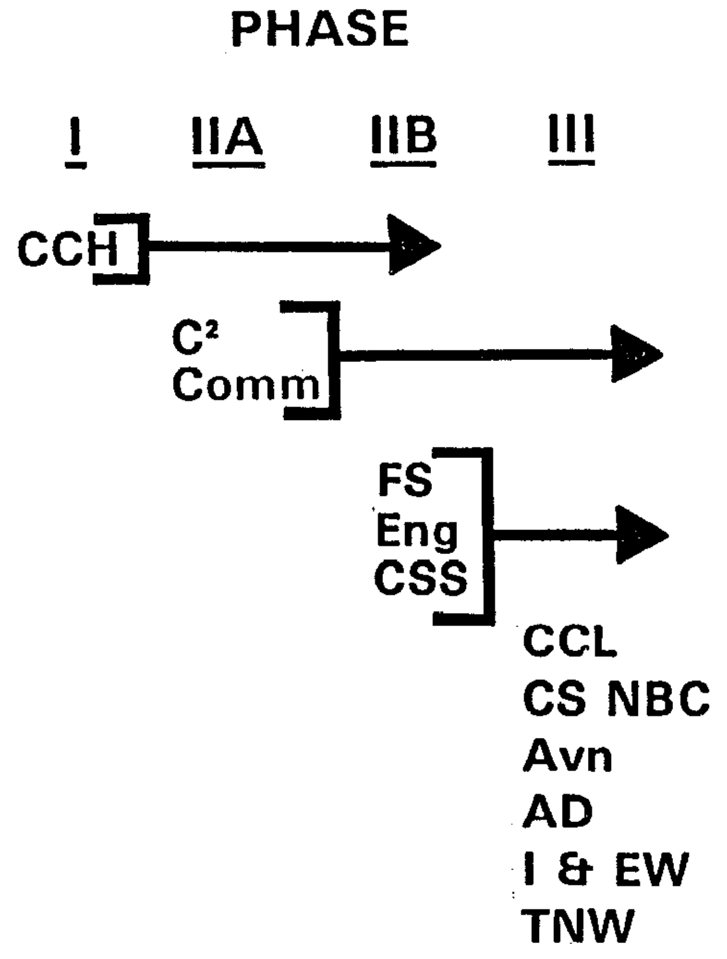
During sustained operations of the Combined Arms Force, what are the effects of the nuclear and chemical environment on:

ISSUE

1. Command and Control
2. Communication
3. Close Combat, Light
4. Close Combat, Heavy
5. Fire Support
6. Air Defense
7. Aviation
8. Intelligence and Electronic Warfare
9. Combat Support, Engineering and Mine Warfare
10. Combat Support, NBC
11. Combat Service Support
12. Battlefield Tactical Nuclear Warfare
13. Force Development (DOC, ORG, TNG, Materiel)

FIGURE 10

ISSUE EMPHASIS BY PHASE



PHASE

ISSUE

Close Combat Heavy

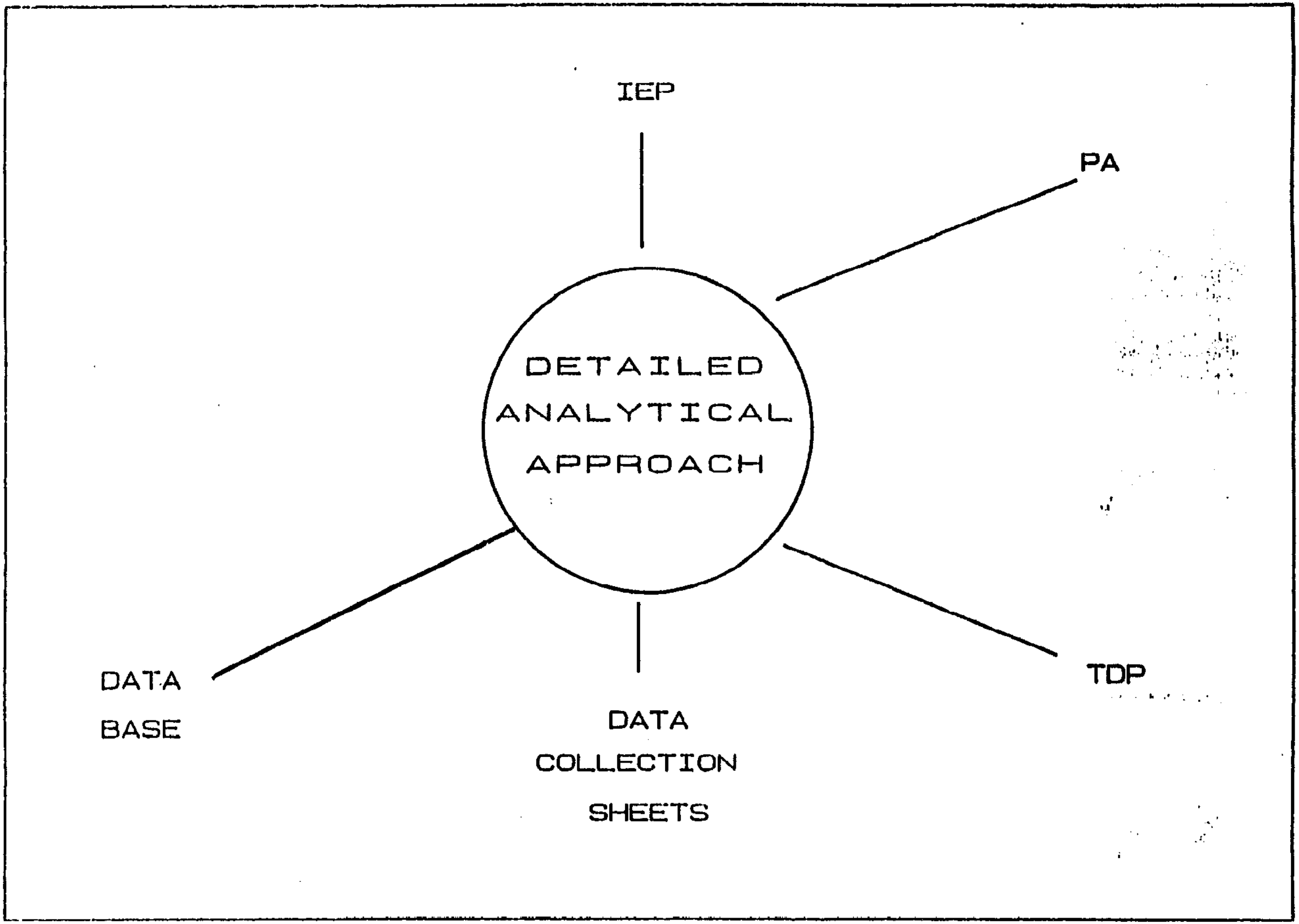
Command Control
Communications

Fire Support
Engineering & Mine Warfare
Combat Service Support
Close Combat Light
Combat Support, NBC
Aviation
Air Defense
Intelligence & EW
Tactical Nuclear Warfare

6392 0124/5789

CANE CHARACTERISTICS

- European scenario
- Force-on-force, near real time casualty assessment, high resolution field test
- Baseline and nuclear/chemical environments
- Minimum of 72 hours continuous operations
- Minimum of 12 hours continuous operations in MOPP 4
- Combat, Combat Support, and Combat Service Support Operations


IEP
PA
DETAILED ANALYTICAL APPROACH
DATA BASE
DATA COLLECTION SHEETS
TDP

CANE TEST DATA

- ISSUE AREAS
  - C², COMMO, CCH, FS, AIR DEF, ENG, CSS, MED, NBC
- TYPE OF DATA
  - INSTRUMENTED, OBJECTIVE, SUBJECTIVE
- NUMBER OF VARIABLES (TASKS) EVALUATED
  - 2900
- NUMBER OF DATA POINTS
  - 7.3 MILLION

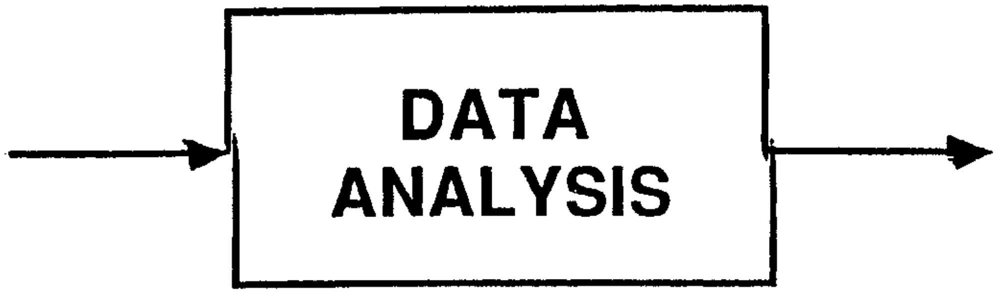
MILITARY PANEL REVIEW

To evaluate, using mathematical and professional judgments, the CANE IIA data and to determine:

- Which measurements and groups of measurements show a definite difference, which show a definite sameness, and which are inconclusive in comparing sustained unit operations in the NCE with those in Baseline.

- The cause(s) of differences noted.

- Those measurements which require additional testing.



DATA ANALYSIS

- Analysis by Issue Area
  - Each variable analyzed
  - Variables are grouped and analyzed
- Analysis Includes
  - Statistical considerations (computer models)
  - Operational considerations
- Data Evaluation Team
  - SME from proponent school
  - Chemical and test SME
  - Statistical analyst
- Output
  - Major deficiency
  - Why the deficiency occurred

FIGURE 16

SUMMARY

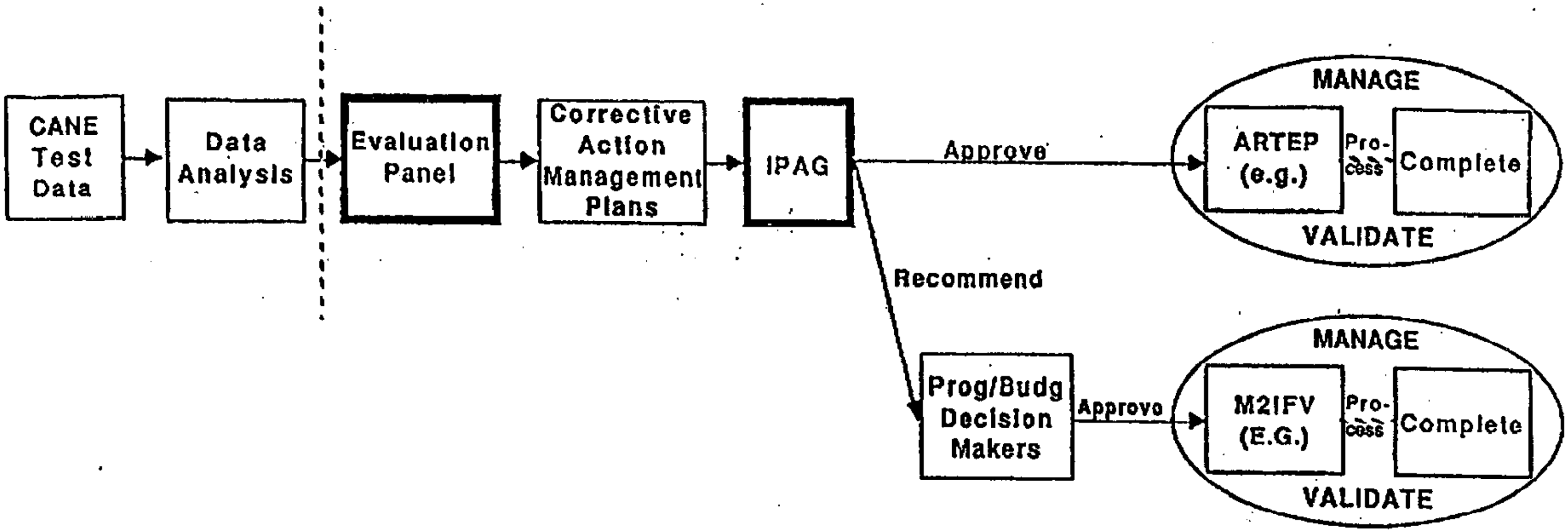
CANE I
- PROVIDED USEFUL SMALL UNIT DATA
- USED TO SUPPORT DOCTRINE, MODELING, ARMY PROGRAMS

CANE IIA
- PROVIDED SIGNIFICANT CO TEAM AND COMMAND AND CONTROL DATA
- REFINED SCOPE OF FUTURE TESTING

CANE IIB
- TEST OF BN TF COMPLETE
- ANALYSIS ON-GOING.

FOE
- CLOSE COMBAT LIGHT. TEST PLANNING UNDERWAY

FIGURE 17



CANE Implementation Program Process

FIGURE 18

DISCUSSION OF "MODELING THE IMPACT OF CHEMICAL OPERATIONS ON UNIT COMBAT EFFECTIVENESS"

by W. Phillips, J. Lombardi and R. Eyler

DISCUSSANT: Alan W. Longshore, US Army Chemical School

In conjunction with the Army Model Improvement Program (AMIP), the U.S. Army Chemical School's Chemical Model Improvement Program (CMIP) is looking at Army combat and training models to insure that chemical warfare is represented as accurately as possible. The main object is not to create new models but to examine current chemical inputs, provide realistic input data, and when necessary, construct chemical modules for models. This is currently underway with the Vector-In-Commander (VIC) model.

The Combined Arms in a Nuclear/Chemical Environment (CANE) series of tests is viewed as a primary data source for model degradation factors. The CANE Phase IIB test conducted in March, 1988 was the largest force-on-force field test ever conducted by the U.S. Army. Not only will these test data provide primary data for lower resolution models but, when examined from a higher resolution perspective, contain valuable information pertaining to command and control and communication throughout the command structure. In addition, these tests are conducted as training exercises; therefore, no volunteers are used and players do not remove themselves from play. These resulting data contain the human factor component as each player group was tested over two consecutive 96-hour scenarios.

The CANE program currently has an AR 5-5 study looking at models to use in test planning, and then, following the test, going back to see if the test results can be used to improve chemical representation in the model. This study is now focusing on CASTFOREM to aid in test planning for the upcoming Close Combat (Light Infantry) Test scheduled for the fall of 1990.

It should be pointed out that CANE also can provide extensive data on operations under non-chemical/nuclear battlefield conditions. After all, half the test was conducted under these conditions of conventional warfare. CANE data are able to address modeling needs such as additional time required to accomplish particular tasks at the unit level when under threat of nuclear/chemical warfare and in Mission Oriented Protective Posture 4 (MOPP 4).

There is a related program that is perhaps more directly related to human factors modeling. That is the Physiological and Psychological Effects of the Nuclear, Biological, and Chemical Environment and Sustained Operations on Systems in Combat (P^2NBC^2) program. The focus of this program is on individual and crew level performance degradation.

What must now be addressed in the near future is the construction of a common database so that modelers will know where to find degradation data, what kinds of data are available, where they came from, and that these data have been examined and are sound both from a statistical and operational point of view. This effort should involve data and people from many different activities: the Chemical School, AMSAA, BRL, TRAC and others. The modeling community has taken steps that will assist by adopting a Common Methodology. This should help drive the database structure. A major portion of the effort involved in this undertaking will be resummarization and reformatting of test data.

One aspect of the CANE analysis program which Mr. Phillips alluded to is the Panel Review Process. It is this concept that may have great payoffs for modeling as well. Military experts have observed the conduct of field tests and have, piece by piece, examined the collected data. Their value in interpreting these data for use in models cannot be overstated. The key to the success of the CANE program, I feel, has been this combination of statistical and analytic expertise with strong operational insights provided by military experts with many years of field experience. I feel that a similar relationship in the modeling community could only provide better models, which are based on better data, and whose results would be more readily accepted by users.

TASK NETWORK MODELING CONSTRUCTS AS APPLIED TO MODELING HUMAN PERFORMANCE IN COMBAT

By K. Ronald Laughery, PhD.
Micro Analysis and Design
3300 Mitchell Lane, Suite 175
Boulder, CO 80301
(303) 442-6947

INTRODUCTION

Over the past few decades, computer modeling of the combat environment has become an increasingly accepted method to help in devising doctrine and identifying weaknesses in NATO mission capabilities. However, one of the clear shortcomings of these combat models has been the extent to which factors affecting human performance have been considered. Even the relatively low level combat models (e.g., division level and below) have virtually ignored the human component. In these models, humans usually do not become fatigued, they continue to perform their tasks normally after a chemical attack (if they are not fully incapacitated), and, in general, they are either working at normal capacity or not working at all. In fact, we know that human performance is not simply "zero or one," but that it varies as a function of a large number of potential stressors. Few field commanders would propose that the battle outcome is unaffected by the level of performance of his soldiers. It is, therefore, obvious that, somehow, the variability of human performance associated with internal and external stressors should be considered in combat models.

In fairness to the combat modeling community, the human engineering community has consistently offered little help as to how human performance variability should be considered. Rather than offering quantitative mechanisms which could be embedded within the models, human engineers have instead tended towards bemoaning the complexity of the problem. While the data and techniques of the human engineering community are less than perfect for the inclusion of human factors into combat models, there are data and techniques that are clearly applicable. This paper will focus on one technique, task network modeling, as well as specific variations that could be applied towards the inclusion of human performance into combat models.

The remainder of this paper is separated into three sections. The next section provides an overview and discussion of the concept of task network modeling. Several models will be presented to provide the reader with a better understanding of the concepts. The following section presents three mechanisms whereby task network modeling could be used by combat models as a means of incorporating human performance variability. The last section will present some ideas for near term research and development. One of the key points of this paper is that the technology and data for the approaches discussed in this paper do not require long term research programs. The technology and data required to implement these approaches is basically available today.

TASK NETWORK MODELING - CONCEPTS

Task network modeling of human performance is a technique that has been under development over the past ten years and which has received fairly widespread use over the past two or three years. Essentially, the performance of an individual performing a function (e.g., driving a tank) is decomposed into a series of subfunctions which are then subsequently decomposed into tasks. This is, in human engineering terms, the task analysis. The sequence of tasks is defined by constructing a task network. An example to illustrate this concept is shown in Figure 1 which presents a series of tasks for a human fishing. This model provides a fairly intuitive introduction to the concepts of task network modeling. However, the basic concept of a network model can be expanded to include several humans as well as equipment and the environment. For example, Figure 2 presents the overview of a model of M60 tank crews in which each of the four crewmembers has his own separate subnetwork.

The level of system decomposition (i.e., how finely we decompose the tasks) and the amount of the system which is simulated depends on the particular problem. For example, in a study to analyze operator workload in a helicopter crew, we constructed autonomous networks for the operators, the aircraft, and the threat environment. While the networks associated with the humans' tasks were far more detailed than those for the helicopter and threat environment, we were able to capture enough of the critical elements of the helicopter and environment to permit a sound study of closed-loop human performance.

While the task networks in a model may be independent, performance of the tasks can be interrelated through shared variables. Once the network is defined, the modeler must determine what variables are relevant to the modeling problem and how those variables are affected by tasks in the networks. The relationships among different components of the system (which are represented by different segments of the network) can then communicate through these shared variables. For example, when a helicopter pilot initiates a pop-up activity, the variables associated with operator controls would be changed by that task to reflect the increase in power applied to the engine. These new values would then indicate to the helicopter portion of the model that it must start executing tasks associated with increasing the aircraft's altitude which is represented by another variable. Once the altitude is above the threshold required for the threats to observe and begin firing, the threat portion of the model begins executing tasks associated with shooting at the aircraft. Representing task sequencing through a network and interrelationships among tasks through the changing values of variables associated with tasks forms the foundation of all network models.

Of course, the strength of task network modeling is that the dynamic aspects of task networks can be simulated on a computer. I have been extensively involved in the development of a task network simulation language, Micro SAINT. This paper will not dwell on the features of Micro SAINT, but let us use the Micro SAINT menu for defining a "task" within a task network to clarify all of the information required to build a task network simulation. Figure 3 presents the menu within Micro SAINT for defining a task. If a user provides the information required on this menu for each task in the network, he will have defined virtually all of a task network model. Let us define these menu items briefly.

1. **Task Name** - any name used to identify the task

2. **Task type** - Tasks may be decomposed into subnetworks by changing the designation of task type to "network" and then decomposing the task into subtasks.

3. **Upper Network** - Since there may be numerous hierarchically embedded networks, this lets you know to which task this network "belongs."

Figure 1

TASK NETWORK FOR FISHING

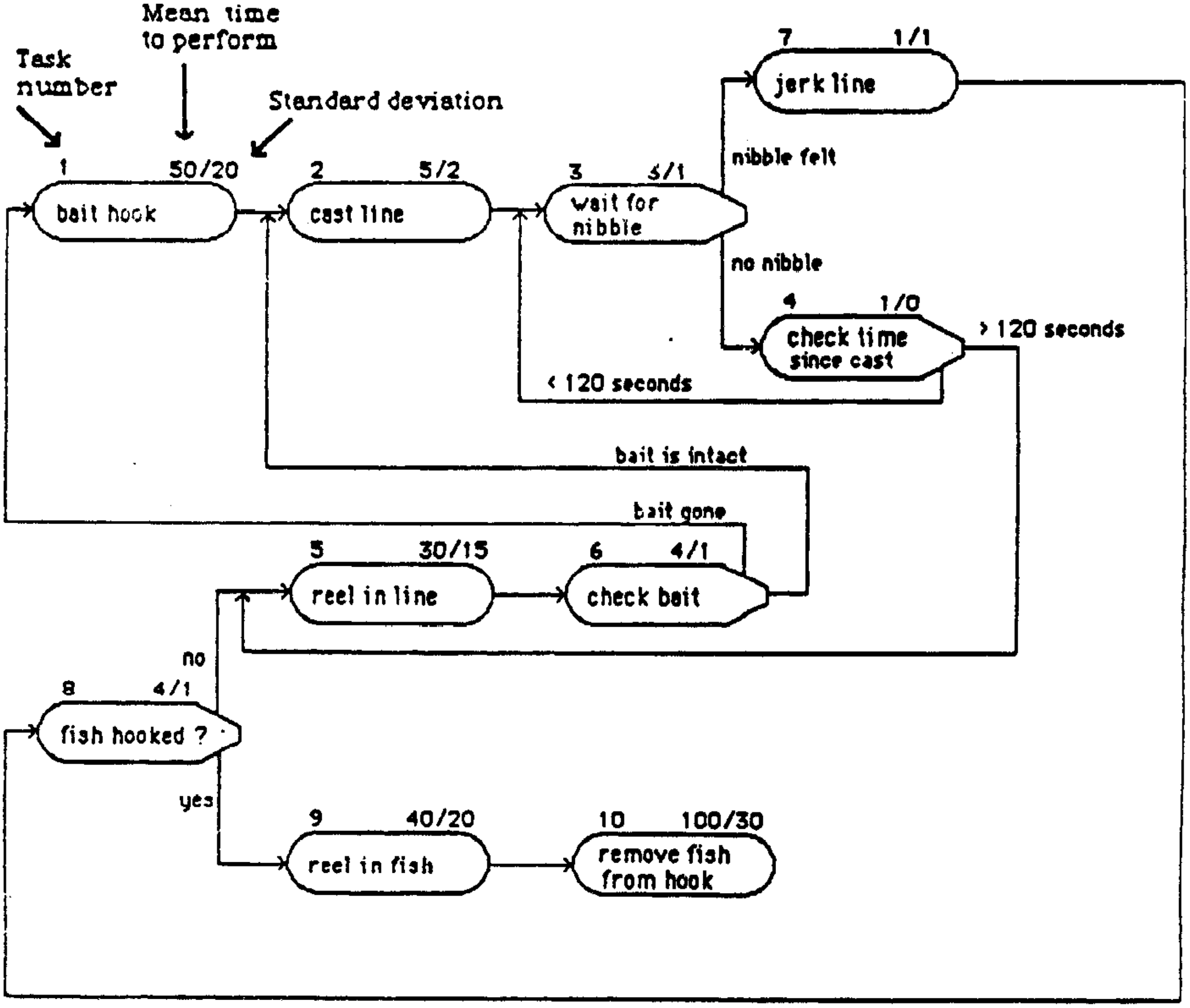


Figure 2

TASK NETWORK MODEL FOR M60 TANK CREW

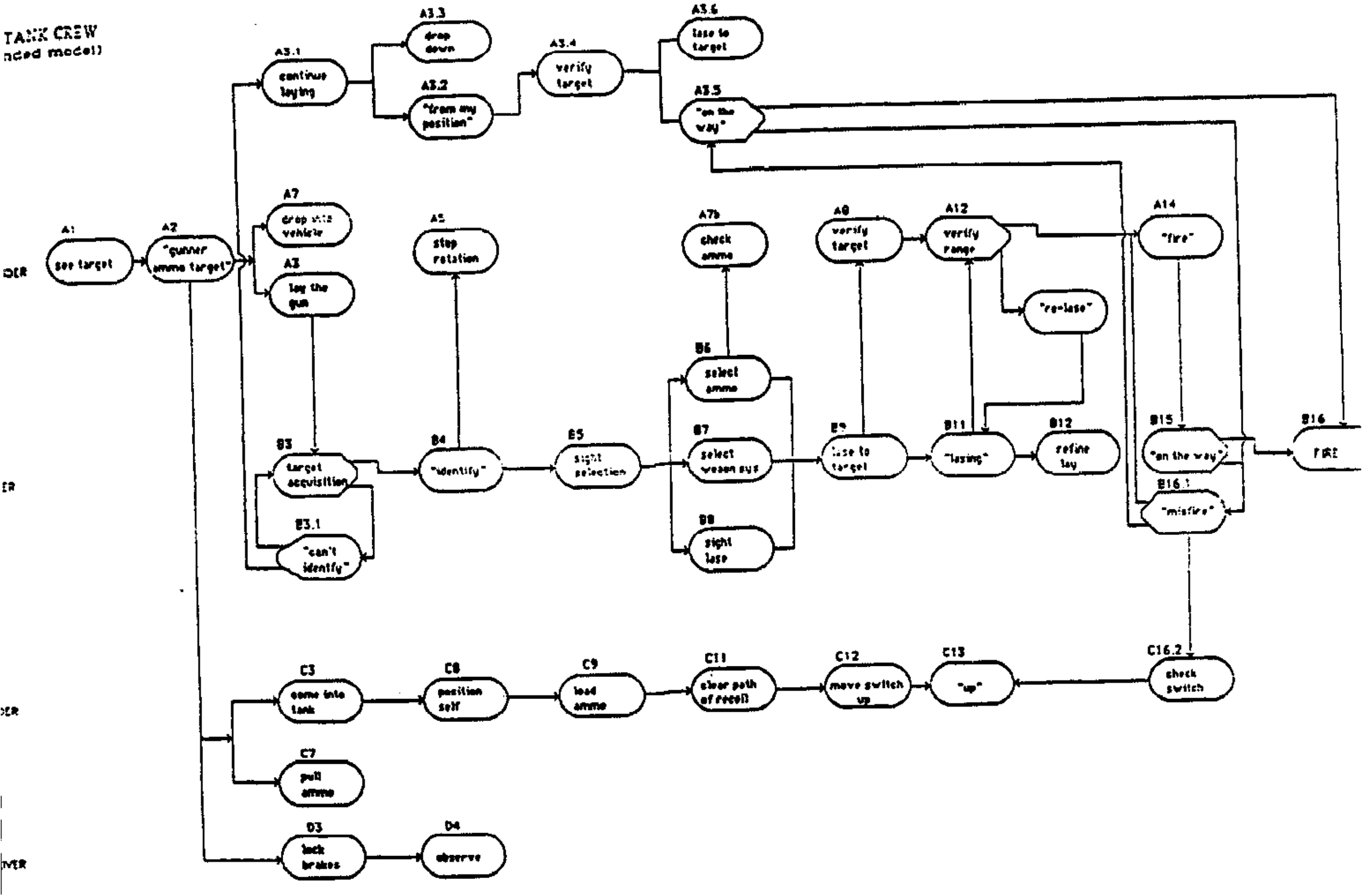


Figure 3

MICRO SAINT "MODIFY TASK" MENU

```
Task Number:  updatepr
(1)  Name:  update pressure valu                      (2)  Type:  Task
(3)  Upper Network:  plant  plant
(4)  Release Condition:  1;
(5)  Time Distribution Type:  Normal
(6)  Mean Time:  10;
(7)  Standard Deviation:  0;
(8)  Task's Beginning Effect:
(9)  Task's Ending Effect:
pressure = 1 - (2.71 ^ (-1 * ((clock*clock)/30000)));
if (clock - topen) > 10 then pressure = pressure - (2.71 ^ -1 *
     (clock-topen-10)/500);
(10) Decision Type:  Tactical
     Following Task/Network:      Tactical Expression:
      Number:        Name:
(11)  openvalve      open v   (12)  pressure > setpoint;
(13)  updatepr       update   (14)  pressure < setpoint;
(15)                          (16)
(17)                          (18)
(19)                          (20)
(21)                          (22)
(23)                          (24)
```

4. **Release condition** - This condition (if one is defined) will hold up task execution until the condition is met. For example, a condition stating that task won't start before another unit is available to assist may be represented by a release condition such as the following:

   unit == 1;

   The value of the variable "unit" would equal zero until a task is completed in which the operator becomes available. This task would wait for that condition to be true (which would probably occur as a result of another task's completion) before beginning execution.

   Another example would be if a task's commencement was contingent upon the availability of a piece of equipment. In this case, we could set a release condition such as the following:

   equipment > 0;

   With the variable "equipment" representing the number of pieces of the required equipment which are not currently in use. This task would then not commence until the value of the variable equipment was equal to 1.

5. **Time distribution type** - Micro SAINT will conduct monte carlo simulations with task performance times sampled from a distribution as defined by this option (e.g., normal, beta, exponential).

6. **Mean time** - This parameter defines average task performance time for this task.

7. **Standard deviation** - Standard deviation of task performance time.

8. **Task's beginning effect** - This field permits the definition of how the system will change as a result of the commencement of this task. For example, if this task used a unit that other tasks might need, we could set the following condition to show that the unit was currently unavailable while this task was being performed:

   unit = 1;

   Then, a release condition such as that shown in item # 3 above would suspend the commencement of another task requiring this unit until the value of unit was changed to a value of 0 indicating that the unit was no longer busy. In a similar manner, the availability of equipment in the example in item # 3 above would be reduced by a task beginning effect such as the following:

   equipment = equipment - 1;

   with this expression indicating that, while this task is active there, is one less piece of equipment available.

   Beginning effects are one key way in which tasks are inter[illegible]ted.

9. **Task's ending effect** - This field permits the definition of how the system will change as a result of the completion of this task. From the previous example, when this task was complete and the unit became available, we could set the

ending effect as follows:

unit = 0;

at which point the task using this as a release condition would begin. Likewise, when a task completes and a piece of equipment becomes available we could use the expression

equipment = equipment + 1;

indicating that an additional piece of equipment was now free for use.

Ending effects are another key way in which tasks can be interrelated.

10. **Decision type** - The decision type defines what happens at the completion of this task. There are several decision types including 1) single (always follow this task with the same task), 2) probabilistic (begin one of several tasks based on a probabilistic branch), 3) tactical (begin one of several tasks based on the branch with the highest "value"), 4) multiple (begin several tasks at the completion of this task) and 5) last (don't begin any task at the completion of this task). From these branch types, any branching logic can be developed.

11-24. **Following tasks/branch weights** - The odd numbered fields represent the tasks which may follow the completion of this task . The even numbered fields represent the weights associated with each branch. The values can be numbers, expressions, or complicated algorithms defining the probability (for probabilistic branches) or value (for tactical branches).

A final note on this menu is that values that appear on the screen can be not only numbers but also algebraic expressions, logical expressions, or groups of algebraic and logical expressions which would, essentially, form a subroutine.

As stated before, the process of constructing a task network model is primarily one of filling in the above information for each task in the network.

If the reader were to conclude that task network modeling is a straightforward concept which is a logical extension of task and systems analysis, he would be right. Task network modeling is a logical extension to task analysis - an evolution, not a revolution. It does, however, greatly increase the power of task analysis in that the ability to simulate a task network with a computer permits prediction of human performance rather than simply the description of human performance that task analysis provides.

THE CONCEPTS OF TASK NETWORK MODELING AS RELATED TO INCLUDING HUMAN PERFORMANCE VARIABILITY IN COMBAT MODELS

Task network modeling provides a sound method for modeling human performance in systems. It builds off of existing task analysis technology and it provides a modeling framework consistent with many other types of system modeling technologies. Additionally, as the title of this paper indicates, I propose that task network modeling provides a perspective on how to incorporate human performance variability into combat models. At a minimum, it provides a conceptual framework and, hopefully, it provides a set of existing technologies, data, and existing models for immediate consideration.

I propose that there are three fundamental approaches to incorporating human performance into combat models which are consistent with task network modeling constructs. These approaches are titled as follows:

1. Develop and include task networks of human performance within combat models
2. Develop performance shaping functions to dynamically manipulate human performance parameters during a combat model run.
3. Use task network models that are run externally to the combat models to set performance parameters during a combat model run.

Let us go through each of these individually and discuss the approach as well as its expected advantages and disadvantages.

Develop and include task networks of human performance within combat models

This approach is simply to find the elements of human performance within the combat models and represent these elements through more detailed task networks. Figure 4 presents a graphic illustration of this approach where the large circle represents the overall combat model and the small circles within represent the "pockets" of the overall model in which human performance plays a central part. To use this approach, the combat model developer would need to perform the following general types of tasks:

1. Identify the pockets of human performance within the combat model for which a more detailed representation is required.
2. Separate human performance activities from other system component activities in these pockets.
3. Identify the variables which will link human performance to the rest of the combat models (e.g., times to perform a group of tasks, error probabilities).
4. Decompose human activity into a set of tasks and define the task network structure for those tasks.
5. Define the tasks' effects on the variables linking the combat model to the human performance model (e.g., through task beginning effects and ending effects).

Of course, these steps are very general and the specific approach would depend on many other factors related to the size and scope of the combat model. However, these steps should provide the reader with an idea of the concept.

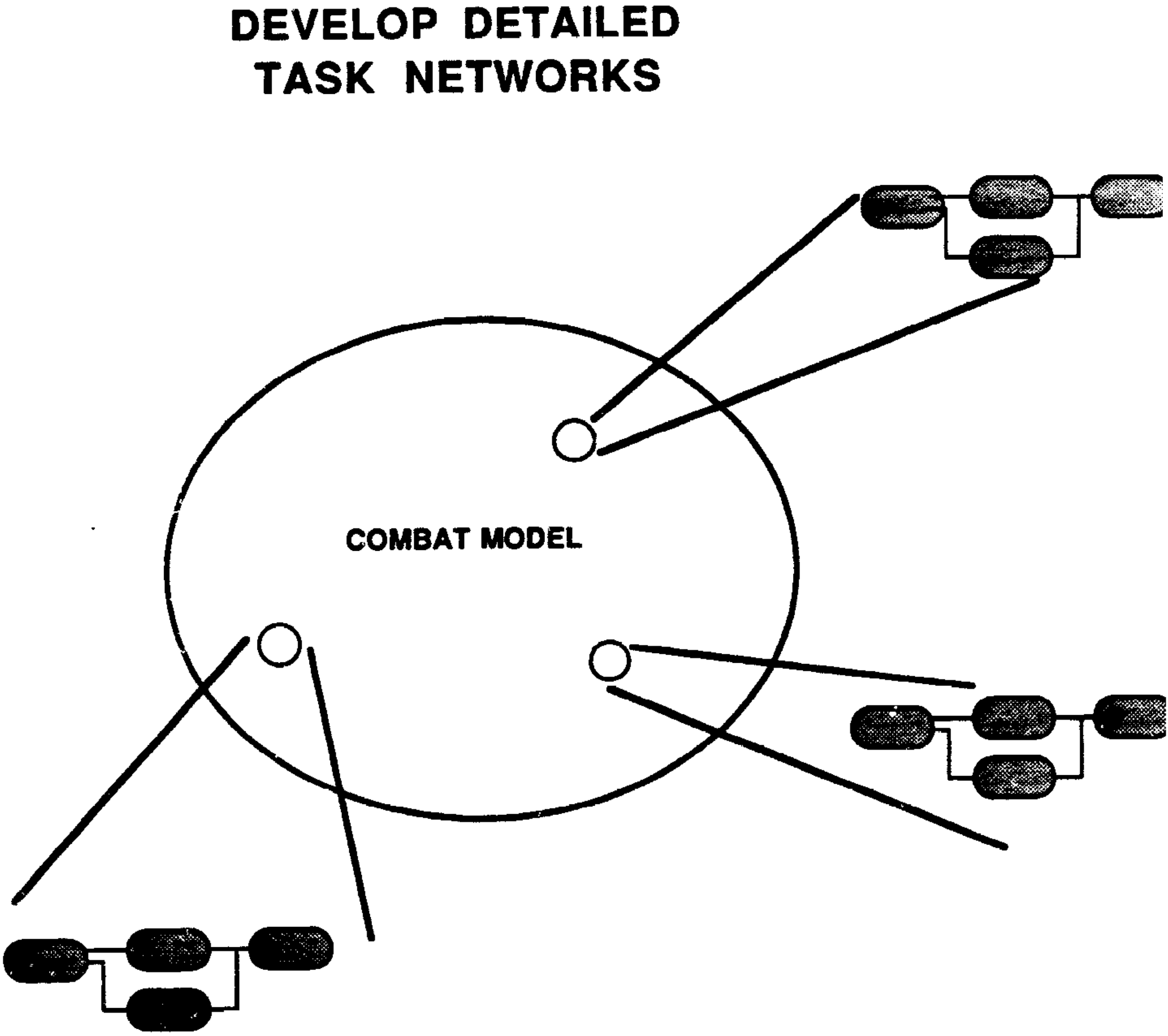
This approach offers several advantages:

The concept is straightforward - This approach is, simply one way of further detailing the human performance component of the combat model. It is easy to explain and offers a high degree of face validity.

It will allow the incorporation of the existing base of human performance models - There is a significant base of task network models of human performance currently available and under development. For example, over the next few years we are

Figure 4

GRAPHICAL REPRESENTATION OF EMBEDDING TASK NETWORK MODELS WITHIN COMBAT MODELS



developing libraries of task network models specifically related to human performance in combat. These can and will provide a ready data base for using this approach.

Existing simulation engines (e.g., Micro SAINT) exist so that models could be constructed externally to the combat model - The human performance network models could be constructed and tested with existing tools. Then, the execution parts of these tools could be embedded within the combat model code and used to run the human performance network models.

It would be relatively easy to interface task network models with existing small unit engagement models (e.g., platoon level) - As a practical matter, the task network modeling approach is very consistent with small unit engagement models.

Human decisions can be modeled - Using the branching logic constructs from task network modeling, complex tactical decisions can be rigorously modeled. There are many possible ways to do this from the very simple (e.g., probabilistic branches) to the very complex tactical decision models as illustrated in Laughery (1982).

Task network models are frequently compatible with hardware models - Many portions of the combat environment within a combat simulation are represented using a reductionist approach and often networks explicitly. This permits easy integration with a task network model.

Some of the disadvantages of this approach are:

Combat models are often frame driven rather than event driven which poses "interweaving" problems - Most combat models advance by advancing the simulation clock by a fixed amount and then updating the state of the system based upon what is projected to have happened during that interval. Task network models advance the clock based on the beginning or completion of a task. These two approaches are fundamentally different, though not opposing. The inclusion of both approaches within a combat model would require some consideration being given to how the two approaches must play together.

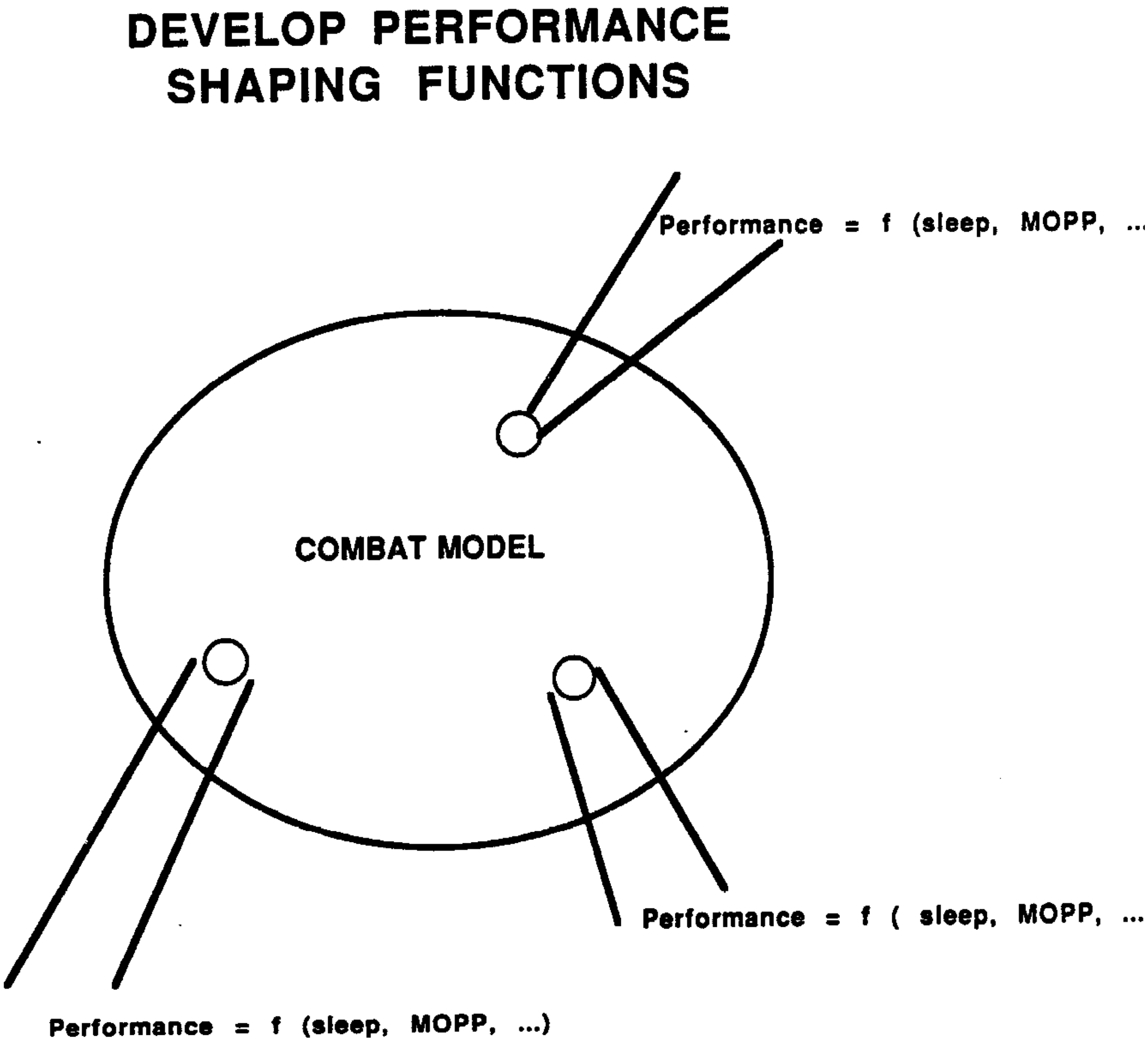
The level of detail is too extensive for large unit engagement models - Obviously, models at the division level and above would require a large number of humans to be represented. While there are some modeling "tricks" which would greatly reduce the number of task networks included in a division level model, I believe that the computational burden would be excessive on today's computers. However, as computational capacity grows, this problem will become less of a factor.

Develop performance shaping functions to dynamically manipulate human performance parameters during a combat model run

Performance shaping functions manipulating human task performance as a function of human stressors could be used to manipulate combat task performance. In doing this, human task time could increase and accuracy decrease as a function of increased stress attributable to sustained operations, NBC effects, or whatever combat stressors were of interest. These can be represented through mathematical and logical interrelationships. This approach is presented graphically in Figure 5. This approach has been applied in the AURA model extensively as described by Klopcic (1988) and a set of performance shaping

Figure 5

GRAPHICAL REPRESENTATION OF BUILDING PERFORMANCE SHAPING FUNCTIONS INTO COMBAT MODELS



functions for classes of human tasks are described in Laughery and Gawron (1983) and Laughery and Gawron (1984).

To use performance shaping functions in a combat model, the following general steps would have to be taken:

1. Identify the pockets of human performance within the combat model where relationships between performance and human stressors are required.
2. Separate human performance activities from other system component activities in these pockets.
3. Identify the variables which will link human performance to the rest of the combat models (e.g., times to perform a task, error probability).
4. Build into the combat model the variables representing the values of the human stressors (e.g., time since last slept, radiation exposure level).
5. Add the subroutines or functions representing the human performance shaping functions as well as the calls to these functions from the relevant pockets of human performance within the combat models.

The advantages of this approach are:

Once the performance shaping functions are developed, they are inexpensive to include - Using this approach does not require the development of detailed submodels of human activity. They require little computer space or time and the cost of selecting which moderator functions to call and when is a relatively small task.

Performance shaping functions easily allow the inclusion of the effects of stressors on combat performance - Task networks do not explicitly provide a mechanism for modeling stressors. In fact, performance shaping functions are the mechanisms used within task network models to define the relationships between task performance time and accuracy and stressors. I am proposing here that they can be used directly by the combat models.

The disadvantages of this approach are:

The data for creating the performance shaping functions are still relatively sparse - As stated earlier, the human performance research community has been remiss over the years in building quantitative relationships between stressors and human performance. There is data available in the archives, but the analyses conducted on them have been weak in this regard as is described in Laughery and Ditzian (1982). Reanalysis of existing data and the collection of new data are required if this approach is to become widely used.

To work, the combat model must separate human performance from machine performance - Human performance shaping functions cannot be overlaid on a time or error number which is heavily influenced by non-human elements of the system. For example, assume we have a performance shaping function which states that "human error probability will increase by 20% after 24 hours without sleep." We could not apply this function to a task which was 95% automated since only that 5% of the task which involved human performance would be affected and the projected increase in error rate would be erroneously large. Therefore, there must be a relatively distinct separation of human and non-human performance components.

Use task network models run externally to the combat models to set performance parameters during a combat model run

Separate task network simulations could be run to provide the combat models with time and/or accuracy parameters of human performance. In essence, the task network simulations could be used to determine the appropriate values to plug into the combat models. This approach is depicted graphically in Figure 6.

The steps required using this approach are:

1. Identify the pockets of human performance within the combat model where better human performance values are desired.
2. Identify the variables which will link human performance to the rest of the combat models (e.g., times to perform a task, error probability).
3. Build task network models to predict those variable values.
4. Set the values within the combat models to represent the predictions of the task network models.

The advantage of this approach is:

Existing combat models would not need to be modified - In essence, all human performance activity could be performed outside of the combat model. Appropriate values to include within the pockets of human performance would simply be plugged in.

The disadvantage of this approach is:

Closed-loop and system dependent aspects of human performance cannot be directly included in the combat model - Obviously, the parameter determined by the task network model cannot be changed dynamically based on a new prediction by the task network model while the simulation is running. What would be required to represent closed loop human performance is a constant readjusting of the parameter based on multiple runs of the combat model to see how the values of those variables affecting human performance change throughout the course of the simulation. An intermediate version of this would be to use the network models to create tables for a few key variables. For example, network models could be used to study the effects of fatigue on a specific task's accuracy level. Then a table relating fatigue to task performance could be created and embedded within the combat model.

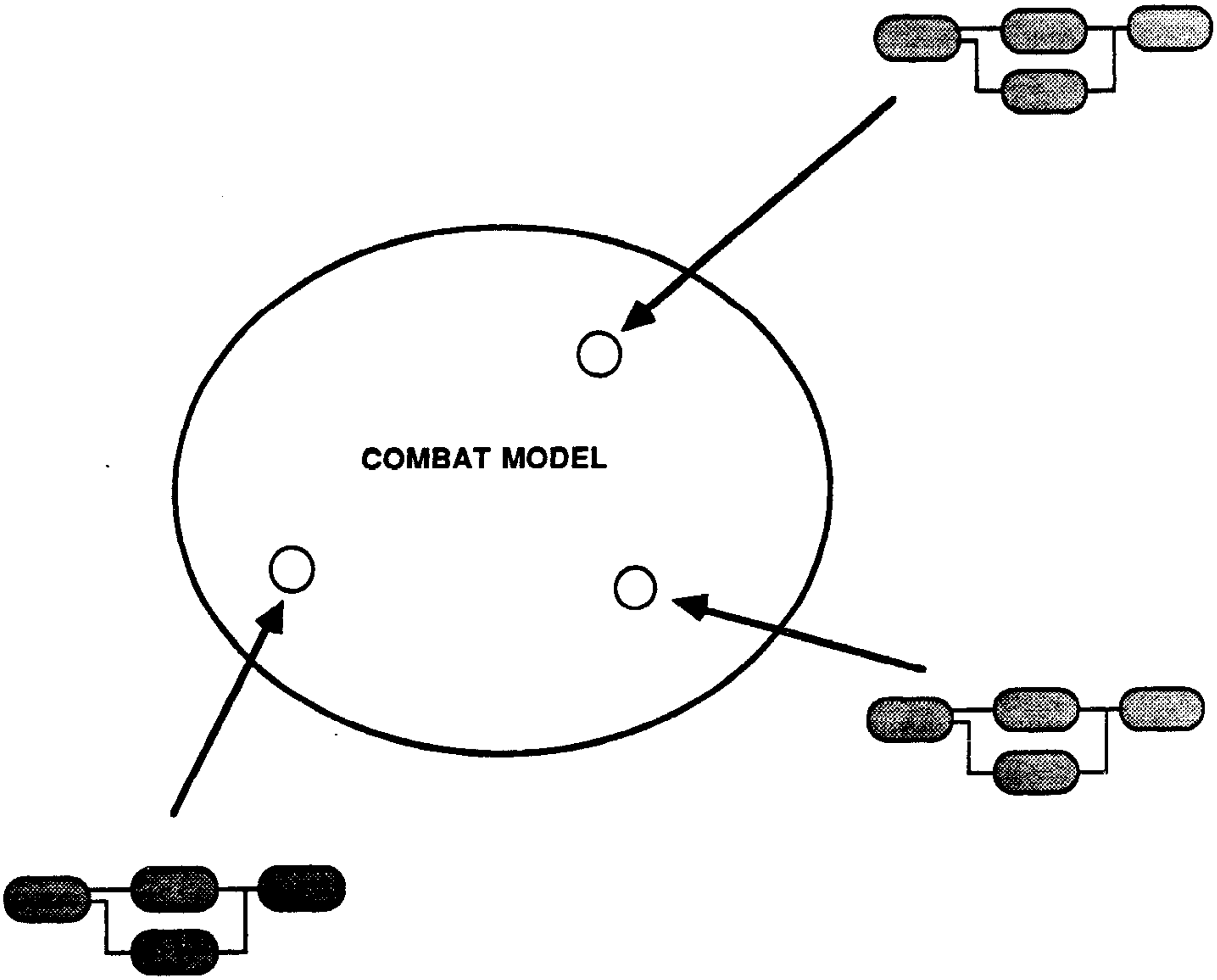
The three approaches are not mutually exclusive

The above three general approaches to include human performance in combat models are in no way mutually exclusive. For example, you could include performance shaping functions into task network models which are embedded in combat models. Or you could build sophisticated network models of human performance and study human behavior

Figure 6

GRAPHICAL REPRESENTATION OF THE USE OF TASK NETWORK MODELS EXTERNALLY TO COMBAT MODELS

CONDUCT SEPARATE HUMAN PERFORMANCE SIMULATIONS



outside of the combat model and then include a reduced version of the network within the combat model itself. In fact, they are three "dimensions" representing ways to include human performance within combat models as illustrated in Figure 7. For any pocket of human performance within a combat model, portions of any or all of the above approaches could be employed.

NEAR TERM RESEARCH AND DEVELOPMENT IDEAS

One of the rationales for using the above discussed approaches is that they are in no way pie-in-the-sky concepts. Task network modeling of human performance is becoming widely understood and accepted and it dovetails well with existing combat models. Therefore, I propose several near term projects which may yield a high and early payoff. Some of these are discussed below.

Embed task network models within some existing combat models - In other words, try it and see if it works. Undoubtedly, there are hidden technical problems that will need to be addressed, but existing models such as JANUS and existing task network simulation tools such as Micro SAINT can talk to one another without redesigning either system to any great extent. Additionally, soldier performance models exist for tank crews, artillery crews, helicopter crews, maintenance crews, and others. By trying it we will prove its feasibility as well as providing a research bed for future studies.

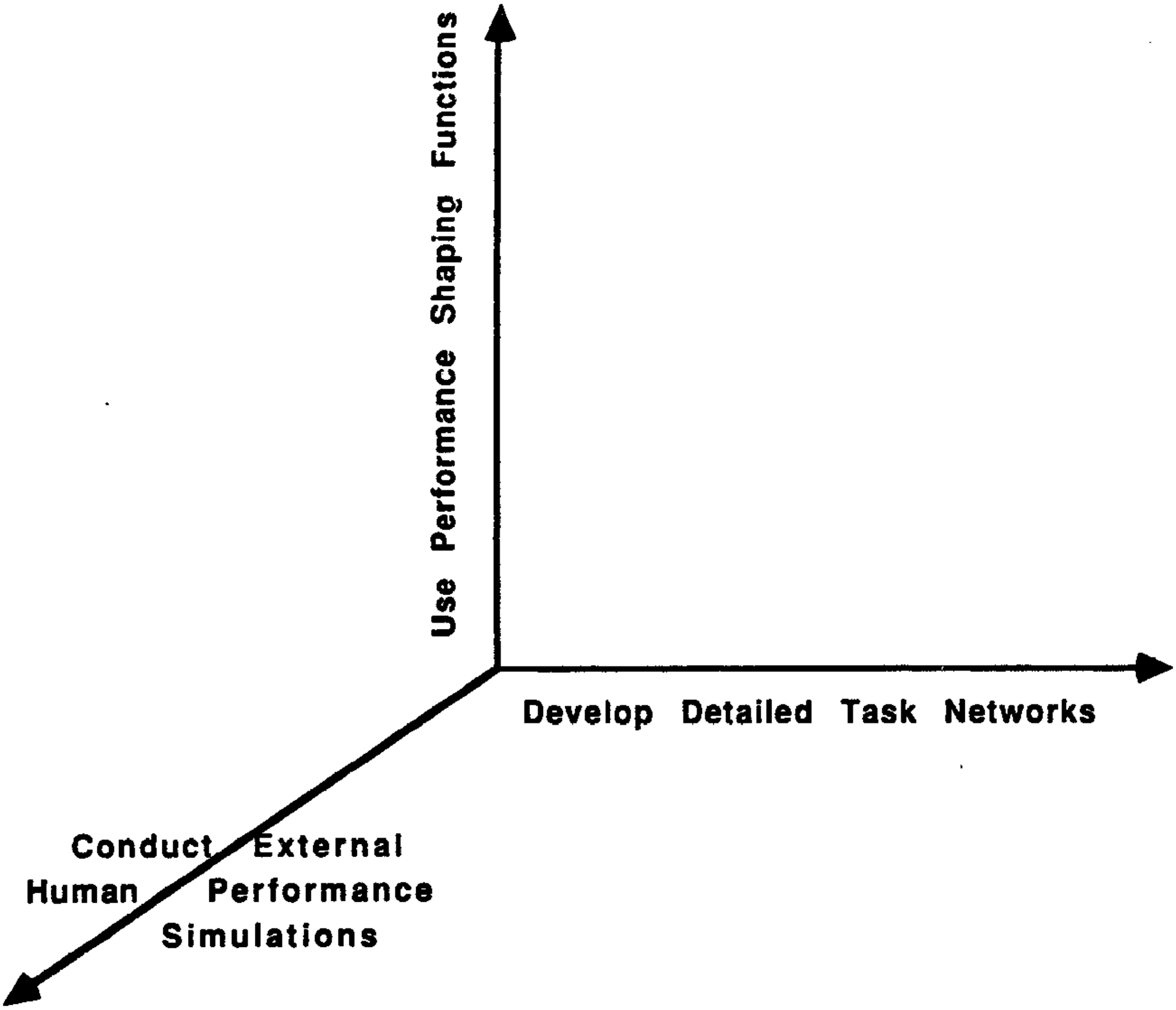
Begin developing human performance shaping functions from existing human performance data - As stated earlier, performance shaping functions can be developed from existing data bases through reanalysis. In previous work, I was able to build these functions from graphed data in journal articles, albeit with difficulty. Certainly, there are better and more raw data than I uncovered and analyzed. A key element of this research may be to build a matrix of "human task types by stressor variables" to see where the data is weakest. For example, there are extensive data relating visual detection task performance to atmospheric haze yet there are very few data relating logical reasoning task performance to sleep deprivation. Future research would be directed by the empty cells in this matrix (i.e., where little or no data exist).

Set up a "clearinghouse" for human performance shaping function data and models - There are already several activities and centers that are performing this kind of function including the Ballistics Research Laboratory, the Joint Working Group on Drug Dependent Degradation in Military Performance (JWGD3 MILPERF), and the Army Research Institute. However, these activities should be better coordinated. Furthermore, agreement on common human task taxonomies and the format of the performance shaping functions would be a significant step forward.

Begin the research to fill in the missing data - We by no means have all of the data relating changes in human performance to all of the stressors of interest. However, there does exist a set of test batteries and measurement instruments developed by the JWGD3 MILPERF which are currently available for conducting this research in the laboratory. These instruments are powerful, easy to use, and were specifically designed for the collection of these types of data. The use of these test batteries will result in cheap and reliable data. Rather than reinventing new research methods and data collection strategies, I believe that we should embrace these tools and put them to use on this problem.

Figure 7

THE CONCEPT OF THE INDEPENDENCE OF THE THREE APPROACHES



Think about a validation study at the small unit level - There are a host of potential data collection opportunities at places like the National Training Center and ARTEPs which could be tapped for a validation study of many of the concepts discussed in this paper. The Army Research Institute is already collecting data on small unit performance in some of these units.[1] It is well within the realm of possibilities that a validation study could be instituted to determine the extent to which the above approaches have predictive validity.

SUMMARY

I believe that we are closer than many may think to being able to provide useful input to combat models. Human performance modeling is no longer a black art or non-existent technology. Tools and data are available which can be used to provide a better representation of human performance within combat simulations and, ultimately, to make better predictions of our ability to successfully engage the enemy.

REFERENCES

Drews, C., Laughery, K.R., Kramme, K., and Archer R., LHX Cockpits: Micro SAINT Simulation Study and Results. Report prepared for Texas Instruments - Equipment Group, Dallas, Texas, June 1985.

Klopcic, T., Presentation at the Human Performance in Combat Models Workshop at Lawrence Livermore Laboratories, December 1988. Meeting details available from Mr. William Banks, Lawrence Livermore National Laboratory, Livermore, CA.

Laughery, K.R., Decision Making in MOPADS (Modeling Operator Performance in Air Defense Systems). Report prepared for Army Research Institute, Fort Bliss Field Unit, March 1982.

Laughery, K.R. and Ditzian, J.L., A Literature Review on Quantitative Human Performance Modeling . Report prepared for the Army Research Institute, Fort Bliss Field Unit, February 1982.

Laughery, K.R. and Gawron, V.J., MOPADS Skill Moderator Function Subroutines. Report 5.10 prepared for the Army Research Institute, Fort Bliss Field Unit, June 1983.

Laughery, K.R. and Gawron, V.J., Making SAINT Models More Representative of Human Behavior: The Theory and Application of the MOPADS Skill Moderator Function Subroutines. Report 5.6 prepared for the Army Research Institute, Fort Bliss Field Unit, March 1984.

---

[1] Detailsof these studies can be obtained from the Army Research Institute Field Unit at the Presidio of Monterey.

CREW DRILL MODELS FOR OPERATIONAL TESTING & EVALUATION

Stanley F. Bolin, Nigel R. Nicholson & Edwin R. Smootz
Army Research Institute

This paper describes a study using Micro SAINT (Laughery, 1985) to build crew drill models for use in a model-test-model approach to operational evaluation. The key tactic was to build crew drill models linking crew performance directly to system performance. Two crew drill models were developed for the Howitzer Improvement Program. Exercising these models over a range of critical task error rates produced plausible results illustrating how average performance time may degrade very slowly in spite of errors. We conclude that crew drill models written in Micro SAINT offer a promising way to inform operational testing and evaluation.

Model-Test-Model & MANPRINT Concepts

The U.S. Army Operational Test & Evaluation Agency (OTEA) is advancing a model-test-model approach. MG Hilmes, CG OTEA, recently spoke of this approach by breaking M-T-M down into two successive parts M-T and T-M. Each part defines different practical objectives in modeling. Models are needed first to know what to measure during operational testing and how to plan for system evaluation. Then, immediately after operational testing, models are needed to sift through, focus and interpret all the data available from system instrumentation, video cameras, observers and soldiers, themselves. M-T-M flows neatly. We use a model to design and conduct an operational test; we improve the model based on test results and use that improved model to evaluate system performance and availability. M-T-M seeks to improve both our model describing how the system ought to work under realistic conditions with soldiers in the loop and our evaluation of the system under test. Of course, the concept is simpler than the reality. Modeling and simulation present many technical problems (Hughes,1984; Murtaugh, 1987)and controversies (GAO, 1987; Gropman,1987).

The MANPRINT program (AR602-2 para 2-12) requires operational testing and evaluation (OT&E) efforts (1) to collect soldier time and error measures and (2) to assess the impact of soldier performance on system effectiveness and availability during simulated combat operations. Both MANPRINT and the M-T-M approach call for including soldier performance in combat models, particularly those developed as realistic operational test scenarios.

Crew drills were selected for modeling because individual soldiers do not fight, operate or maintain equipment systems in isolation. The squad -crew -team level of organization is the smallest productive element in the hierarchy. Individual training leads to squad - crew - team training as the building block of collective training in units. Crew drills practice specific move, shoot, communicate, maintain and resupply functions. Such drills are repeated and amplified at each successively higher level of unit training.

Although crew drills may be the same in terms of techniques and standards (SOP), the network of consequences and contingencies increases as collective training moves up through company-battery-troop to battalion levels and above. By building sets of drill models for different kinds of crews and linking them in a network of networks, unit performance may be simulated. By adding a combat scenario with command and control, combat simulation may become possible with crew-sized units in the loop. But first, we need a means to produce crew drill models.

Micro SAINT was chosen because it was literally designed for the job, building networks of networks. Self-Propelled Howitzer crews were selected to start because the Howitzer Improvement Program (HIP) is preparing for operational test and the HIP MANPRINT team had prepared the way.

Howitzer Improvement Program (HIP) Crew Drill Requirements

Operational Sequence Diagrams and timelines developed for the Howitzer Improvement Program (HIP) were used to build two crew drill models with Micro SAINT software. HIP gun crews will use automatic fire controls to enable them to respond quickly to fire mission orders (Department of Army, 1988). Gun crews will be trained to perform various crew drills to insure mission responsiveness. To these ends, the HIP prime contractor prepared Operational Sequence Diagrams (Geer, 1981, Meister,1985) to examine, study and document the details of performance necessary to accomplish specific missions. Operational Sequence Diagrams show networks of interactions among the crew members, the automatic fire control system (AFCS), and external input/output (I/O) for a single HIP vehicle. They diagram functional flows of information and action over time and events.

The fire-from-road-march drill requires the crew to pull off the road and fire a first round within 60 seconds after receiving a mission order. This march fire mission diagram begins with mission alert alarm and ends with firing a first round. It shows 110 nodes or points of interaction among crew members and their weapon system.

The 12-round-volley drill requires the crew to sustain fires at three rounds per minute after firing the first round. The volley fire diagram begins with recoil from the first round and shows the next round in a volley of fires. It shows 58 nodes or points of interaction for each additional round to be fired.

The HIP prime contractor used the Operational Sequence Diagrams to develop sets of schematics, timelines and assumptions necessary for specific crew drills. The schematics spread crew and system functions over time in seconds. The timelines list key events and their approximate times in approximate accumulative time order. The assumptions describe initial conditions for the weapon system and the crew.

The HIP crew consists of four soldiers, a chief of section (COS), a gunner (GNR), a cannoneer (CAN), and a driver (DRV). The HIP weapon system (SYS) is assumed to be well maintained, supplied and moving along in a state of readiness for a mission order. The crew is assumed to be well selected, trained, rested, and alert for a mission order. The time estimates, informed by expert judgment and experience, have been set to meet minimum time requirements.

These two crew drills, march fire and volley fire, were selected for model development because the HIP is moving toward initial operational test and evaluation. There is a practical need to estimate task time and error tolerances in crew performance. Such estimates might prove useful in formulating training objectives and in evaluating crew performance. Good computer models might even be used to show crews in training why they need to drill toward the kinds of time and error tolerances built into the models. Good computer models might, finally, help to reduce and evaluate field test data. But, first, there have to be good models.

This report is based on work done by the HIP prime contractor MANPRINT team. Our dependence on that work is too great for a mere footnote. Our research required direct cooperation and assistance from the HIP Program Manager and the prime contractor. They did exceptionally thorough work which made our work possible. If these models fail to be good ones, the fault is entirely ours. If they prove to be good ones, substantial credit rests with the HIP Project Manager and the HIP prime contractor MANPRINT team.

Technical Objective: Simple Start-Up Models

Although we hope these models are considered good ones, this report cannot properly address the quality or value of the models. Such evaluation depends on what use is made of models as well as judgments about the assumptions built into the models. From our point of view as researchers and developers, questions of validity and verification are premature. These models represent a new start. Our interests and objectives are in the technology for building and understanding crew performance models.

However, we are not purists, interested in models for their own sake. We want to build models that can be easily understood and rapidly adjusted for use in the MANPRINT program (Department of Army, 1987). In particular, we believe that crew drill models may help to include the soldier in combat models (Van Nostrand, 1988), especially operational test models for new weapon systems in their early stages of development.

Therefore, these models have been built very simply. They make few assumptions and the assumptions are simple. They have been built using an interactive computer system that is easy to learn and use. This system is called Micro SAINT (copyright 1985 Micro Analysis and Design, Boulder, CO). It can be run on commonly available personal computers; the term Micro stands for micro-computer. The acronym SAINT stands for systems analysis of integrated networks of tasks.

Micro SAINT software (Laughery, 1985) is designed to build networks of networks, but our models are simple networks. The software provides a choice of sampling distributions for random events (normal, gamma, exponential, and rectangular). We elected rectangular because it is simplest and it can be argued that well trained crews perform with machine-like precision. Although Micro SAINT supports complex contingency tables and dynamic modeling, we elected to use direct error-correction loops. If it is wrong, do it again until it is right. No changes were made in the time limits for repetitions. This kind of simulation of the "green machine" is very mechanical. We figure we can complicate the simulation later. We want the start-up models simple so they can be easily understood (Chubb, Laughery & Pritsker, 1987). We must understand each drill network before we link the drills in scenario networks.

Errors may occur anywhere in a human network, of course, and no crew position is completely free of error, but in these initial models, errors are limited to two critical tasks. Only the simulated gunner and cannoneer make errors, and the simulated errors are made in tasks that are checked by the chief of section. These tasks are the retrieval of projectiles by the cannoneer and the retrieval of charges by the gunner. The simulated checking process is flawless. Each error found in checking leads to a repetition of the retrieval and check tasks. This model design provides opportunity to see how crew error rates might impact on crew drill completion times.

Ten error rates were examined. They varied from zero up to 64% per individual duty position and task. For simplicity, we gave the two individual tasks the same error rate when we varied the rate. We sought in the start-up models to keep the relation between individual and crew as straightforward as possible. With just two independent tasks, the aggregate crew error rate is equal to the sum of the individual probabilities minus their product; so the individual task error rates of 64% yield an aggregate crew rate of 87%.

Although simple in computation, 87% may seem to be an unreasonably high rate of error, but that is a positive feature of modeling. We can examine variables beyond reasonable limits and do so with other variables under perfect simulated control. For example, in our two drill models, the chief of section always catches every error and makes sure there are no errors before firing. There are no simulated short or long rounds because of crew error. Only the time to completion is free to vary in computer trials of these two drills and the only cause is crew error in two critical tasks. Without error, the drill models meet required time standards because they were built to do so.

The models start out delivering exactly the kind of results we want. The interesting questions and observations to be made from running such models are those suggesting how, how often and how badly performance may degrade. Putting the matter positively, how sturdy and dependable is excellent model behavior?

Fire-from-Road-March Drill Model

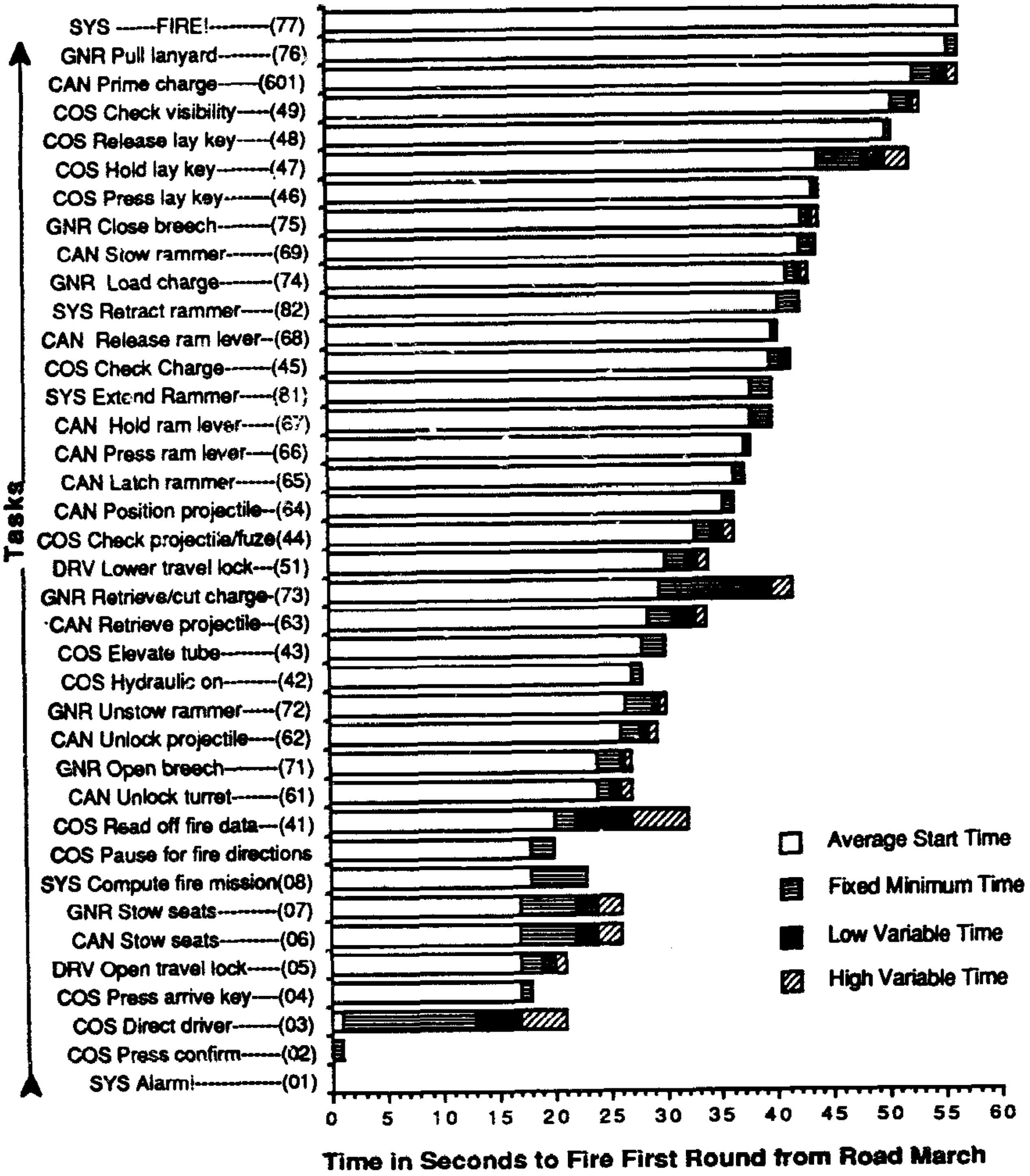
March fire timelines are shown in Figure 1. This figure shows each task in the model network in sequence from "SYS Alarm!" to "SYS Fire!" The average start time is time spent waiting on a preceding task. Note that "COS Direct driver" begins immediately after "COS Press confirm" which is a fixed time task.

Note further that "COS Press arrive key" begins at the mid-point of "COS Direct driver" between low and high variable time. That mid-point is the average of a rectangular distribution of equally probable performance times shown as low and high variable time. The model takes 16 seconds, on the average, give or take four seconds to simulate finding a place to stop and stopping. That eight second spread is filled randomly on every computer trial or run of the model.

Since the sequence shown is based on average performance times for each task, the mid-point of variable time is regularly the starting moment for some later tasks. Task numbers shown in parentheses were used to index the model. They may be used conveniently here to point to such instances. For example, tasks 4 through 6 start immediately after task 3. Other contingencies may be inferred from end-to-start alignments, but they may be separated by intervening tasks on the chart. For example, task 8 follows upon task 4.

The critical path through this particular network sequence happens to flow through the COS, GNR and SYS tasks including CAN only at priming the charge (601). DRV is on the path only in being directed by COS to an emplacement. The critical path is not shown on the chart because it changes depending on task times and error conditions. This particular sequence is based on fixed conditions of average time for each task and no errors. When errors and corrective actions are introduced by the model, the critical path changes back and forth between GNR and CAN tasks.

Figure 1
March Fire Timelines



Errors were simulated at "CAN Retrieve projectile" (63), "GNR Retrieve/cut charge " (73), "COS Check projectile/fuze" (44) and "COS Check charge" (45). If-Then loops at each check task generated repetitions of associated retrieval-check task sequences.

We explored ten individual task error rates. They were zero, 1%, 2%, 4%, 8%, 16%, 24%, 32%, 48% and 64%. We stopped at 64% because the aggregate crew error rate from either one or both of the round preparation tasks approaches 90%. When chance produces at least one time-consuming error in nine out of ten drills and no errors are allowed to get by, it can take a long time to complete a drill. When large numbers of error-laden drills are simulated on a personal computer, run times can increase from a few minutes to hours. As a practical matter, 64% seems high enough as an upper limit.

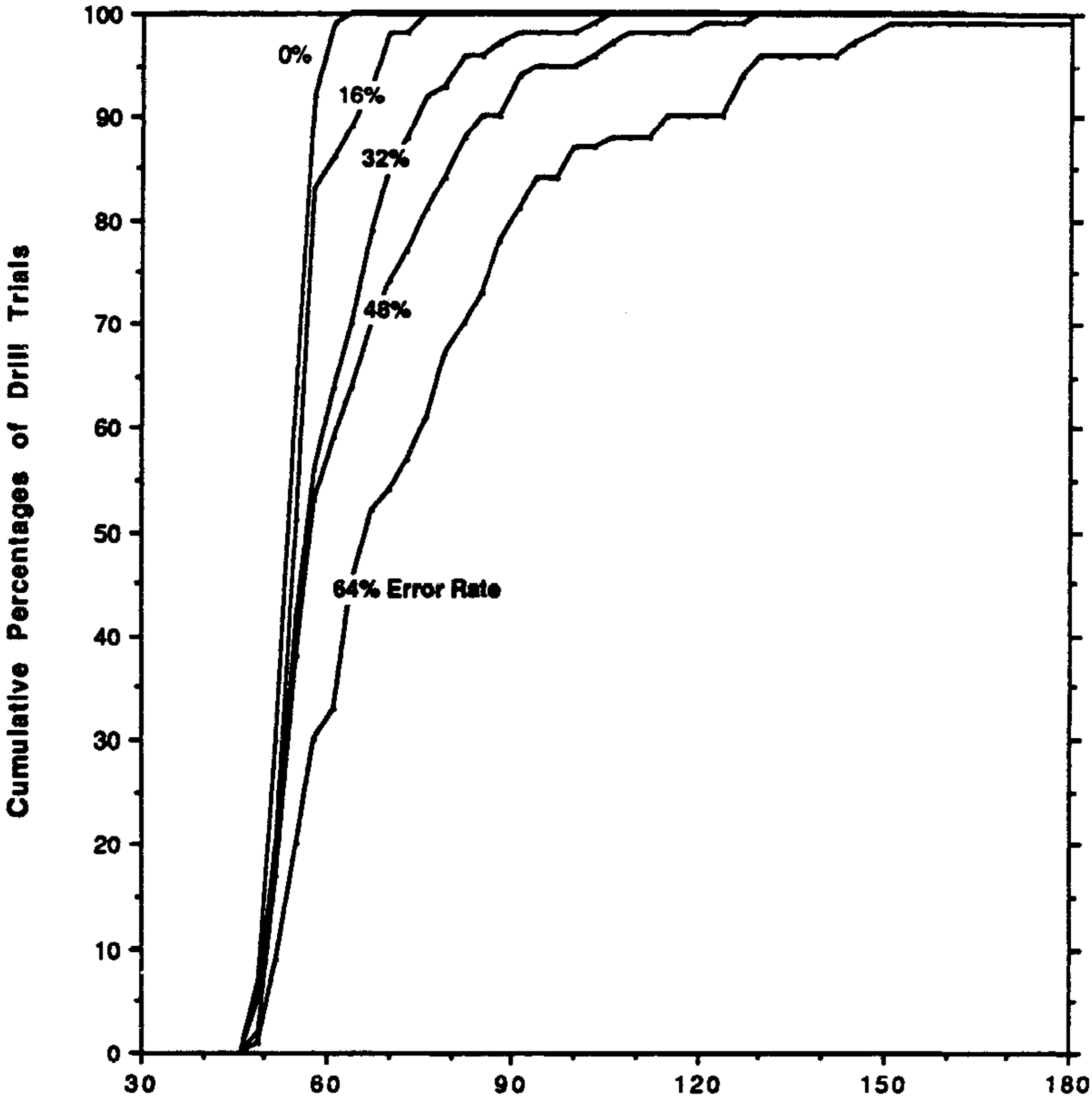
We began our exploration with a doubling rate scale which can be seen up to 16%, but we quickly found that low rates made little differences in means but larger differences in variability. We then introduced half steps at 24% and 48% to look more closely at this unexpected robustness in means relative to variability. Figure 2 is a picture of the results for five error rates.

The 0% curve shows that all error-free drills took less than 60 seconds. Minimum time was 46 seconds for 0% and every other rate. The common point of origin for these curves is not a fact of nature nor is it a computational necessity. It results from our use of a network model and the number of runs or trials per error rate. It is the critical path time for the lowest built-in task times, which is a network model feature. It happened because 100 trials was enough to insure that the least-time path occurred at least once for each error rate. It is a reminder that we are looking at a simulation of crews trained to a set of tight time standards and, by model design, effectively maintaining those standards.

A counter-intuitive feature of this picture is the tight clustering of the stems for 0% through 48% error rates up to the 50th percentile of the cumulative distribution. The 50th percentile is the median. Median completion time is well within the 60 second requirement in spite of individual task error rates running up toward one in two drills. The aggregate crew error rate equivalent of 48% is 73% or seven in ten drills. Even the highest error rate yields median time less than 70 seconds. These medians show the robustness of central tendencies in this model.

Increasing variability can be seen in the progressive growth in the right-hand tails of the distributions. Yet that growth is constrained for individual task error rates at or below 16%. All of the small error rate curves are contained within that thin region shown between 0% and 16%. To see constrained variability, read across the 85th percentile to intersect the 16% curve and then to the time scale. Even the 85th percentile appears to be

Figure 2
March Fire Drill Times by Error Rate



Time in Seconds to Fire First Round in 100 Computer Trials per Error Rate

sturdy in the face of considerable error. Among crews as good as the simulated crews, the 60 second requirement might be met in spite of critical task error rates up to one in six drills or aggregate crew error rates up to three in ten drills.

At high task error rates, however, the frequencies of long completion times increase rapidly. More than one-third of drill trials fail to meet the 60 second requirement at 32 and 48 % task error rates. More than two-thirds fail to meet it at the 64% error rate. These results illustrate potential zones of intolerable failure.

12-Round-Volley Drill Model

Volley Fire Timelines are presented in Figure 3. This drill model begins where the march fire drill ends, but it is not meant to be run or practiced in tandem with the march fire drill. This drill assumes that the weapon has been prepared for volley fire. Projectiles have been fuzed and charges have been cut in readiness for a volley, fire-for-effect. Time standards differ from march fire. The two critical tasks take less time. The chief of section checks projectiles and charges; the AFCS has already been set. The figure shows just one round after the first in a series of thirteen. Interpretation follows that for Figure 1. The difference is that required 12-round time is not shown; it is 240 seconds for 3 rounds per minute.

Figure 4 presents results for volley fire in a series of error rate curves. The curves are separated more than they were for march fire. The separation results from repetition and the accumulation of time losses from round to round. Yet the interpretation is much the same as it was for march fire.

In particular, the drill time requirement was met 85 to 100% of the time with individual task error rates as high as 16%. But note that the curves do not originate at a common point. Repetitive models show their built-in differences. No simulated crew operating at 64% task error meets the required time. Yet lower error rates make the requirement some of the time.

In this model, median time is below the required minimum time at task error rates as high as 24%. That rate is two notches lower than the highest tolerable rate in march fire, but it still demonstrates dependable performance inspite of considerable error. Sustained fire is more sensitive to error, however, and the potential zones of intolerable failure are lower than they are for march fire.

If-Then Implications

To summarize and directly compare overall results for our two models, ratios were computed to see the increases for each error

Figure 3
Volley Fire Timelines

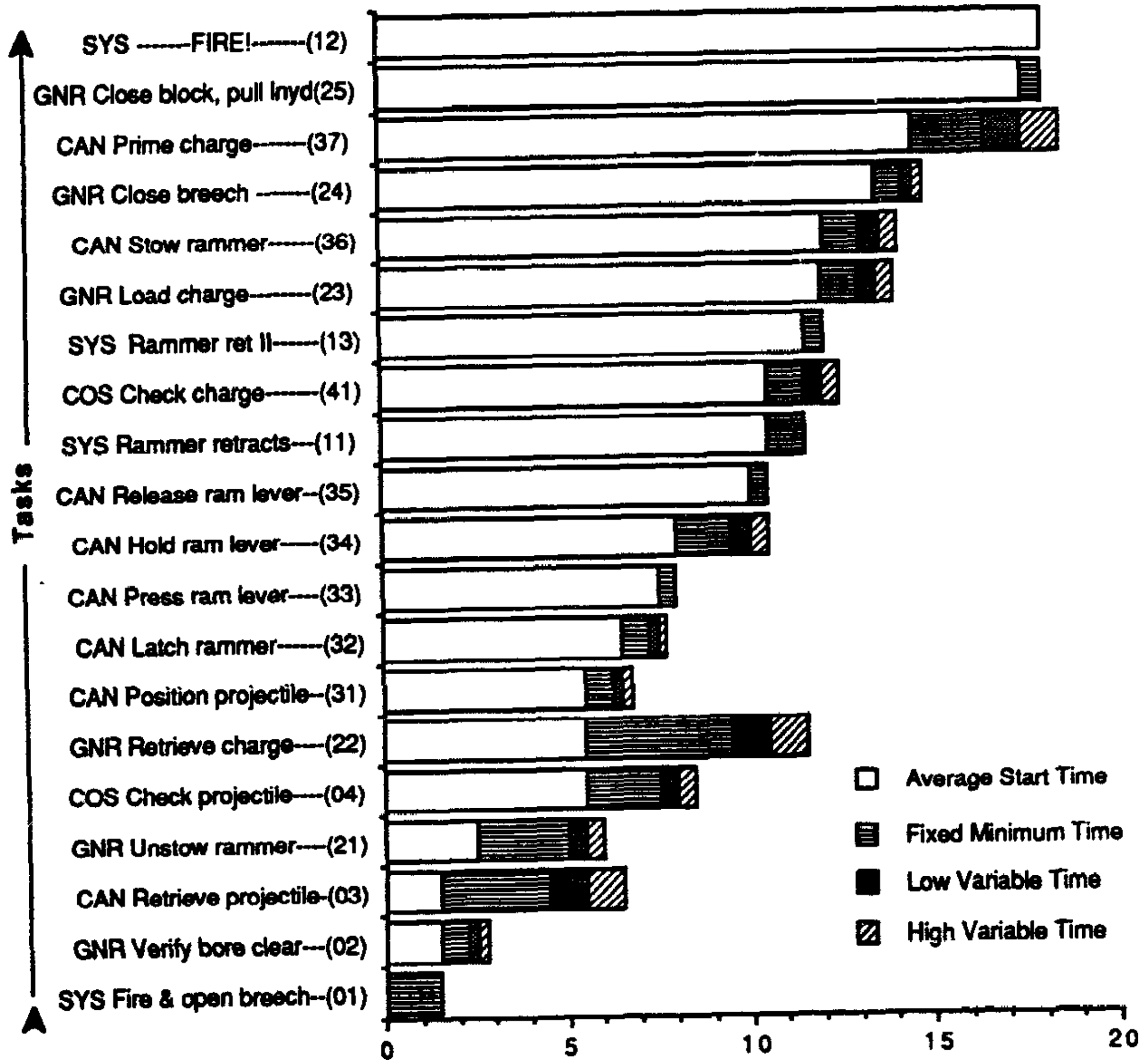
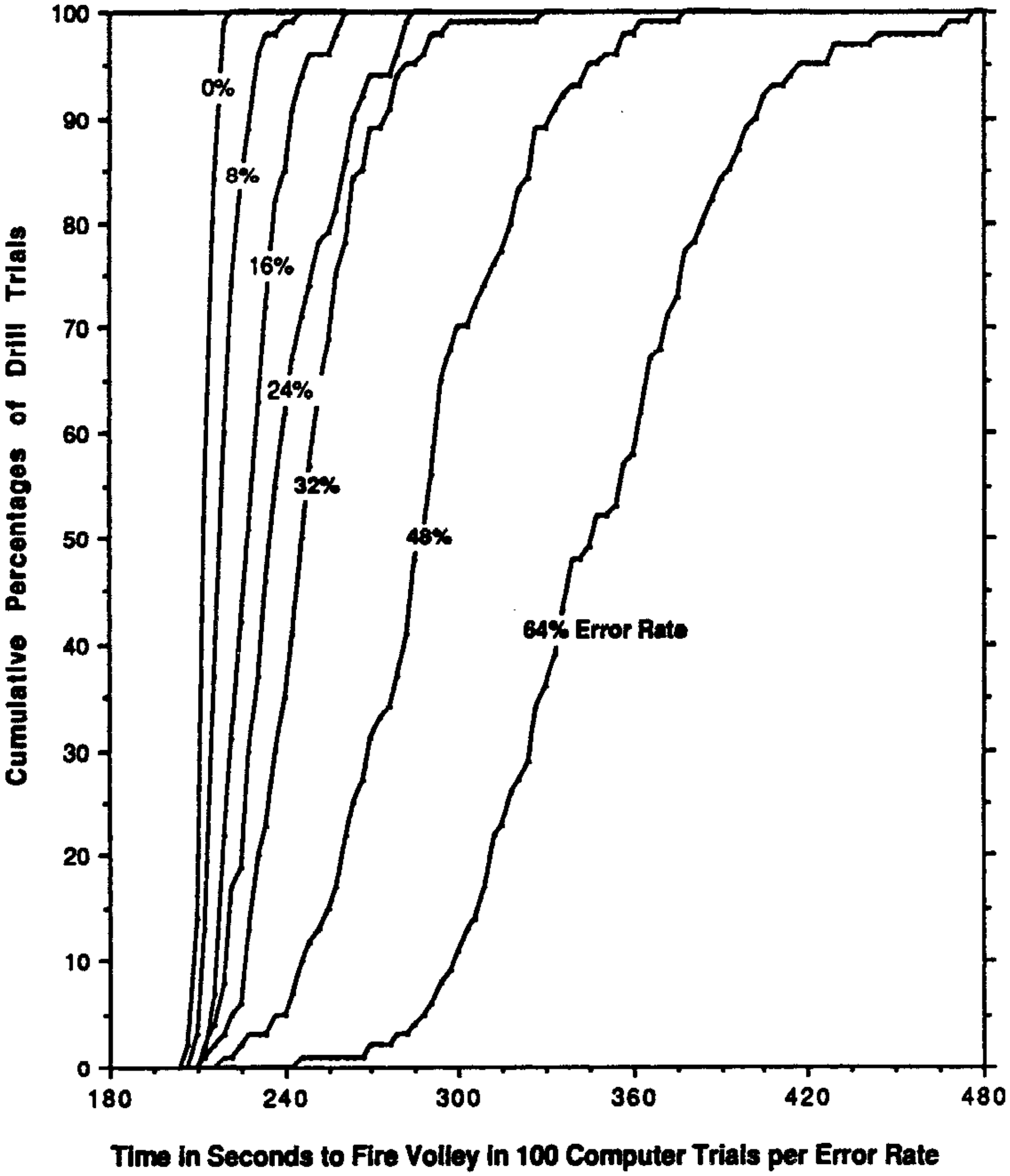


Figure 4
12 Round Volley Drill Times by Error Rate



rate relative to zero error. Such ratios were computed for average times and for the standard deviations. The relative increases for both statistics derived from the two crew drill models are shown in Figure 5.

The robustness of average time over error conditions is clearly shown for the two crew drill models. Even with high error rates, the relative increases are less than two-fold. The standard deviations, however, can increase ten-fold and fourteen-fold.

We believe the increases in variability against the apparent robustness of the averages are important observations. Another expression of the same phenomenon from these computer runs is seen in the apparent ability of "well trained, well rested, well equipped" simulated crews to meet drill time standards 85% of the time with critical task error rates as high as 16% .

These start-up model observations may have some useful implications. These If-Then implications are oriented toward crew training and modeling sustained or continuous operations.

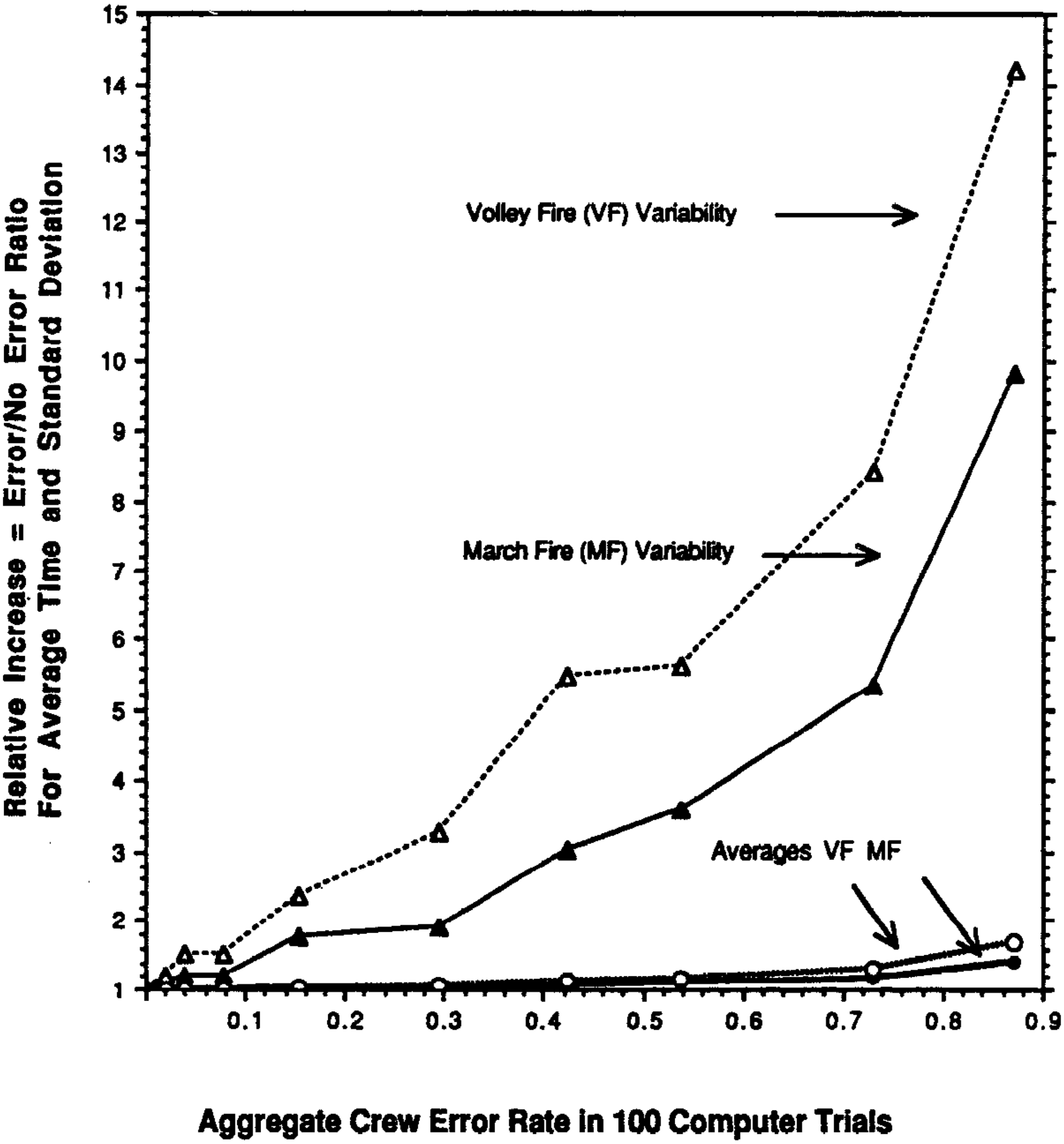
If real crews can be trained to sustain the kinds of task time standards used in these models, they may learn to satisfy demanding drill time requirements. Whatever else, the drill models and their results illustrate the value of high and tight time standards in training.

Another potential training application is in the control of errors. Specifically, the models might be used to show crews how important it is to create slack time to absorb occasional errors. Model-based drill training might be used to encourage a new performance method called "error management " (Frese and von Rosenstiel, 1988). This approach aims to reduce or eliminate the negative consequences of errors rather than pretending that errors never happen.

If error rates increase as a function of continuous or sustained operations, however, then even highly trained crews may be subject to excessive performance variability. The drill model results illustrate how intolerable increases in variability may occur with literally no change in task time standards. Slow and small degradations in average performance can be accompanied by large increases in poor performance.

There is something paradoxical about consistently high standards and slowly changing averages being associated with rapid increases in intolerable failures. Our simple start-up models show it can happen, but the real world process is not simple. We suspect that some kind of progressive concatenation of errors may be involved when there are several interacting team members and their individual performances gradually decay under continued performance stress.

Figure 5
Impact of Error
On Average Time and Variability



Although we have not modeled "stress" or fatigue decay, we have modeled the presumed consequences, error and increasing error. It is possible to consider the changes generated by increasing error rates as expected performances in continued or sustained operations. This line of thinking suggests that modeling to understand the concatenation or cascading of error in crew drills might provide leverage against performance decay under stress.

Note that in Figure 5, we have used the aggregate crew error rate, rather than the individual task error rate, on the x-axis . We made this change to focus attention on the simulated crews in considering these relative increases. The increases in variability shown here result from right-hand tail growth. That lop-sided growth is based on accumulated losses in time generated by a simulation that will not fire a badly prepared round. Less dramatically, quality control is rigid and time, rather than quality, is lost.

The Boolean algebra and Venn diagrams associated with the computation of agggregate crew error rates from the individual task error rates lead us to deal with three elements. The computation is as simple as a Venn diagram. The aggregate crew rate is the sum of the individual task rates minus the product of the two rates. It is straight forward and it is right for an overall rate of any single error or combination of errors.

However, it does not weight the elements according to their time penalties as these occur in our models. Note carefully that we are not saying a probability estimate should do so. Instead, we observe that the three elements represent different time penalties which contribute differently to the tail growth observed in Figure 5.

Each individual task being repeated has its expected time penalty. When both tasks must be repeated, the time penalty may be as much as the sum of penalties or it may be no greater than the largest penalty, depending on the sequence and parallelism in the task network. We believe that the two critical tasks in our two models were in sequence with little overlap in time in most runs. If so, then high error rates with high likelihoods of joint errors would generate large summed penalties. In effect, the tail grows much faster when both critical tasks must be repeated.

The growth may further increase if the chief of section becomes a bottle neck and corrections must wait for inspection. In any case, it appears that growth in variability accelerates more rapidly than the average time because of compounded time penalties. To find out if this is so, we need to set up the software to count error compounding and delays within crew drill trials. These simple models are not as simple as they appear.

The next steps in development toward a simulated crew operations technology (SCOT) are technical ones. We need to satisfy ourselves that we understand how these models work. Crew drills may be decomposed or simplified still further. Assumptions and choices of sampling distributions need to be explored. While such technical work is being done, the timelines and geography for an operational test scenario may be developed for modeling.

A major advantage of modeling in Micro SAINT is that interested parties may readily check results and test different assumptions. Micro SAINT software models may be run on commonly available personal computers. The practical result is that complex networks of networks can be assembled from component parts. A distributed development effort can be organized using the common language of Micro SAINT.

Conclusion and Recommendation

Two crew drill models have demonstrated potential for estimating the consequences of critical task errors. Model results have shown how average performance time may degrade very slowly in spite of errors if crews can maintain high and tight task time standards. Whether or not real crews can do so is an empirical question beyond inference from models, but the models do illustrate the value of training to such task standards. They also suggest that operational performance requirements may be satisfied by training to such task standards. Therefore, we conclude that crew drill models written in Micro SAINT offer a promising way to study time and error measures before training and testing real crews.

Crew drill models may also provide ways to evaluate soldier performance in operational testing under simulated field conditions. In particular, error rates may be systematically introduced to simulate likely consequences of stress and fatigue in continuous or sustained military operations. The present models were limited by design to simple procedural tasks and just two critical tasks. The simulated chief of section never made an error of judgement in checking rounds or charges. There were no other errors of judgement or process. There was no context of events , no scenario of move, shoot, communicate, resupply, maintain and do it all over again, hour-by-hour, day-after-day. There were no breakdowns or interruptions, no sudden changes characteristic of battlefields. Such events may be built into an operational test scenario simulation. If a variety of crew drill models were run in such a simulated environment, we would have the beginnings of simulated crew operations technology (SCOT). We recommend that SCOT be developed to include a number of simple crew drills and to combine them with command and control actions during operational test scenarios. We intend to conduct further research toward this end.

References

Chubb, G.P., Laughery, K.R., and Pritsker, A.A.B. (1987). Simulating Manned Systems. G. Salvendy, (Ed.), Handbook of Human Factors ,(pp.1298-1327). New York: John Wiley & Sons.

Department of the Army (1987). Manpower and Personnel Integration (MANPRINT) in Materiel Acquisition Process (AR 602-2). Washington, D.C.

Department of the Army (1988). United States Army Weapon Systems 1988 ASA(RDA) Handbook. Washington, D.C.

Frese, M. and von Rosenstiel, L. (1988). The Organizational Psychology and Psychology of Economics Unit, Department of Psychology, Ludwig-Maximilians University of Munich, Federal Republic of Germany. The Industrial Psychologist, 25/3, 39-43.

GAO, United States General Accounting Office (1987).Assessing DOD Simulations for Credibility. GAO/PEMD-88-3.

Geer, C.W. (1981). Human Engineering Procedures Guide (Report AFAMRL TR-81-35, P. 114). Wright-Patterson AFB, OH: Aerospace Medical Division.

Gropman, A.L. (1987). On Pursuit of the Holy Grail. Phalanx, 20/4, 1,16.

Hughes, W. (1984). Military Modeling. Alexandria, VA: Military Operations Research Society.

Laughery, K.R. (1985). Modeling Human Operators on a Microcomputer: A Micro Version of SAINT. Simulation, 44, 10-16.

Laughery, K.R. (1985). Micro SAINT User's Guide (Including computer program software). Boulder, CO: Micro Analysis & Design.

Meister, D. (1985). Behavioral Analysis and Measurement Methods (pp. 66-69). New York: John Wiley & Sons.

Murtaugh, S.A. (1987). Of Human Behavior and Actions as an Essential Ingredient for Realism in Combat Modeling. Phalanx, 20/4, 22-24.

Van Nostrand, S.J. (1988). Including the Soldier in Combat Models (Executive Research Project Report S73). Fort McNair, D.C.: Industrial College of the Armed Forces.

Micro SAINT MODELING OF
THE CLOSE-IN WEAPON SYSTEM (CIWS) LOADING OPERATION:
INTERNAL VALIDATION AND SENSITIVITY ANALYSIS

Louis Tijerina and Delia Treaster
Battelle Columbus Division
Columbus, Ohio

INTRODUCTION

This is a period of rapid growth in quantitative and computational approaches to human factors engineering (HFE) and human-system interaction. There is a need for HFE inputs to system design which are compatible with other engineering and science data. There is also a need to quantitatively model human-system performance for purposes of description, prediction, and control. A variety of methods have been developed to help accomplish quantitative modeling of humans interacting with systems. One class of such quantitative methods is simulation.

The advent of relatively inexpensive, powerful, easy-to-use computers and a professional community educated in computer use has prompted a renewed and expanded interest in modeling and simulation of human performance. In particular, much attention has focussed on network modeling with a PC-compatible software package called Micro SAINT (Drews, 1986; Archer, Drews, Laughery, Dahl, and Hegge, 1986). To build a simulation, this methodology requires the analyst to draw a task network (i.e., the sequential and simultaneous arrangement of activities) and provide statistical descriptions for each activity. Unlike other human performance models, Micro SAINT is relatively empty of inherent behavioral assumptions; because it is really a language rather than a model per se, the analyst is free to include whatever details and theoretical constructs are deemed appropriate.

This paper describes the development of Micro SAINT simulations of the Close-In Weapon System (CIWS) loading operation with a 3-man crew. The CIWS is a critical topside weapon system aboard U.S. Navy ships which serves as the last line of defense against enemy missile or aircraft attacks. The speed and accuracy with which it can be loaded and fired may thus have life-and-death consequences; this is why it has been selected for analysis. The objective of the study was to model the loading operation in terms of:

(1) Discrete subtasks performed by each crew member
(2) Average time to complete each subtask
(3) Distribution of completion times
(4) Probability of completing a subtask successfully
(5) Analysis of the recovery procedures to be followed in the event that a subtask cannot be completed successfully.

This paper presents a case study on the use of Micro SAINT to model a multi-man shipboard tactical operation. It will provide an example of how one might go about building and validating a sequential network model of a biomechanical tactical operation; the problems encountered in this effort will also be discussed. The uses of such a simulation for combat modeling will be considered. Finally, some of the major issues we uncovered which surround modeling human behavior and performance will be listed. It is hoped that this paper will contribute to a more informed use of sequential network models of human performance in the future.

METHOD

Subjects. Three crew members on two naval frigates (hereafter identified as Ship #1 and Ship #2) participated in the data collection effort. Two members of Ship #1's team were CIWS-qualified and had worked together for at least six months; the third member was unfamiliar with the CIWS and served as a "helper." All three members of Ship #2's team were CIWS-qualified and had worked together for at least six months.

Apparatus. The loading operation was videotaped with one tripod-mounted video camera equipped with a character generator (for superimposing titles and annotations on the tape), zoom lens and an internal clock. The clock provided the times of the event recorded on the tape. The videotapes were analyzed using a 5-head VCR with slow motion, frame-by-frame advance, and freeze frame capabilities.

Materiel. The crews were observed aboard their respective vessels, which were equipped with the General Electric Phalanx Close-In Weapon System (CIWS). Because of safety considerations, the loading operation was conducted using dummy rounds which resemble live rounds in shape, size and weight. However, the dummy rounds were not sheathed in the protective mesh which blocks high-frequency signals that could activate a live round. The dummy rounds were packed in 100-round, 76 lb. ammunition boxes. Because live rounds are primed with depleted uranium, crew members, as appropriate, wore protective gloves used to prevent heavy-metal poisoning when handling the rounds. Crew members wore standard uniforms; no arctic apparel or individual protective equipment were worn.

Data Collection Procedure. The environmental conditions present were suitable for the development of a baseline model. The ships were docked and ship motion was negligible. Weather conditions were dry, with mild temperatures, and the sky varied from cloudy to partly sunny. The videotapes were made in daylight, under natural lighting conditions.

The crew members were instructed to perform the loading operations in their usual manner. Differences observed between the equipment and procedures used by the two crews are summarized in Table 1.

Performance of Tasks 2, 3, and 4 was contiguous, and Task 1 was performed separately. This was done in order to avoid repositioning the camera during the transition from Task 1 (which began at the ammunition lockers) to the other Tasks (which were performed at the CIWS). Figure 1 illustrates the sequence and repetitions of the Tasks.

TABLE 1. DIFFERENCES BETWEEN CIWS TEAMS AND MATERIEL

Ship #1	Ship #2
Worked on starboard side of CIWS	Worked on port side of CIWS
Used maintenance platform	Did not use maintenance platform
2 CIWS qualified crewmembers	3 CIWS qualified crewmembers
Hydraulics activated on opposite side of principal work area	Hydraulics activated on same side as principal work area
Spanner wrench used on same side	Spanner wrench used on opposite side
Rounds laid out on deck	Rounds loaded directly from ammo box
Uploaded dummy rounds	Did not upload dummy rounds
Defective pins	Nondefective pins
Older CIWS	Newer CIWS
Simulated unlocking ammo lockers (one of three locks nonfunctional)	Did not simulate locking ammo lockers
Removed bar lock from ammo locker (two additional locks plus stowage)	Did not use bar lock
Third deck person sometimes helped by activating hydraulics	
Fixed positions for crewmembers	Rotating positions for crewmembers
Initial belt length to insert into exist unit was longer than that used by Ship #2.	

Note: Significant differences were observed in crew organization and training.

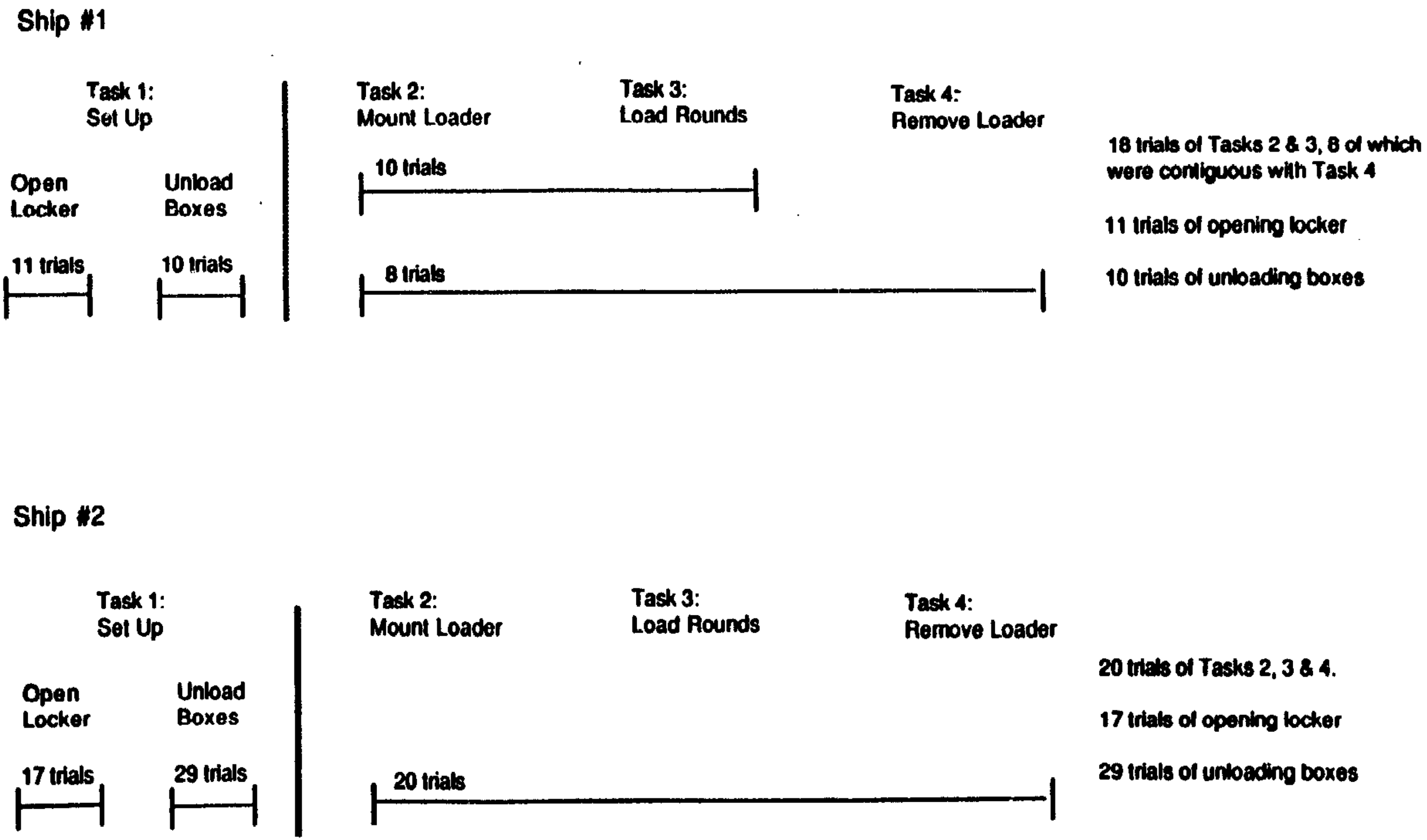


FIGURE 1. SEQUENCE AND REPETITION OF VIDEOTAPED TASKS

ANALYSIS

Data reduction, as defined here, encompassed three phases:

(1) Operational definitions of each activity element were developed and refined, culminating in an activity element description

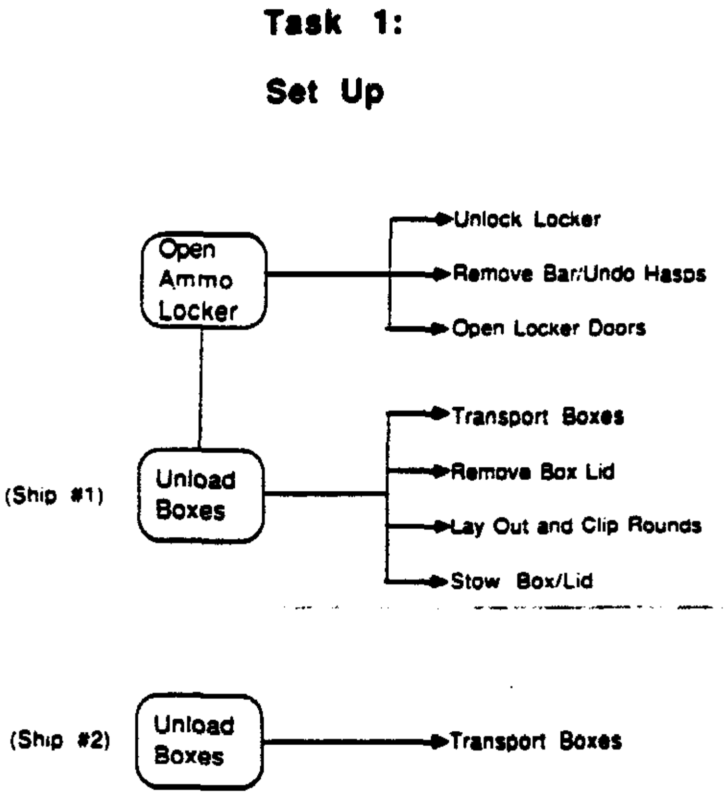
(2) These activity elements descriptions were used to extract timing data from the video tapes and then to calculate completion times for all the activity elements of each task.

(3) The sequential and simultaneous nature of each activity element within the tasks was determined.

Phase I: Activity Element Description. The first phase of the data reduction effort was oriented toward development of operational definitions for the activity elements (tasks, subtasks, and task elements) which comprise the CIWS loading operation. The importance of this effort is underscored by the fact that these descriptions would later be used to exhaustively describe the loading operation in the network models; those activities which were not included would not be represent in the models. Hence, the listing of activity elements was considered to constitute both a necessary and sufficient description of the loading operation. Part of the descriptive effort required the establishment of pre- and post-conditions, which represent a required state of the system before or after the activity element which must exist in order for that activity element to be completed. In those cases where no pre- or post-conditions are given, these are implicitly taken to be the end state of the preceding activity and the start state of the next activity, respectively.

Using this framework, activity element descriptions were generated for the time and motion study in an ad hoc fashion. The researchers reviewed the videotapes, discussed the actions which transpired, and arrived at a consensus of the appropriate activity descriptions. These descriptions underwent numerous revisions and modifications before being finalized. The final task hierarchy thus derived is illustrated in Figures 2 through 5 along with the activity descriptions.

The activity descriptions possessed several features. First, they had to be visible on the videotapes, although perhaps not on every repetition. This was essential to allow the completion times to be collected from the videotapes. Secondly, the descriptions were of discrete events with definite start and end states. There were several instances in which the activity elements flowed continuously from one to the other; despite the fact that they were logically separable, clearly definable start and end states were not possible. In these cases, the logically separate but continuous activity elements were subsumed under a single higher level category for which start and end states could be defined. A third attribute was that the activity elements be of reasonable duration (> 3 seconds) since an activity element must be long enough to be accurately timed. Finally, the level of description had to be appropriate for the purposes of the study. In this regard, no cognitive functions were modeled, although these may affect task performance.



TASK 1 ACTIVITY ELEMENT DESCRIPTION

Task 1: Set Up

Subtask 1.1: Open Ammo Locker

Task Element 1.1.1: Unlock Locker
Start State:
Operator touches (first) lock.
End State:
Operator releases (last) lock.

Task Element 1.1.2: Remove Bar/Undo Hasps
Start State:
Operator touches bar or hasp.
End State:
(Bar set-aside/all hasps up) Operator releases hand from bar/hasps.

Task Element 1.1.3: Open Locker Doors
Start State:
Operator touches door.
End State:
(All doors open) Operator release hand from last door prop.

Subtask 1.2: Unload Boxes (Ship #1)

Task Element 1.2.1: Transport Boxes
Start State:
Operator touches box in locker.
End State:
(Box set on deck) Operator releases box and begins to rise.

Task Element 1.2.2: Remove Box Lid
Start State:
Operator puts hand(s) on lid handles.
End State:
Operator releases set-aside lid.

Task Element 1.2.3: Lay Out and Clip Rounds
Start State:
Operator touches belt end.
End State:
Operator releases laid out rounds (begins to rise).

Task Element: 1.2.4: Stow Box/Lid.
Start State:
Operator touches box/lid.
End State:
(Box/lid transported and stowed) Operator releases set-asid box/lid.

Subtask 1.2: Unload Boxes (Ship #2)

Task Element 1.2.1: Transport Boxes
Start State:
Operator touches box in locker.
End State:
(Previous box transported to deck) Operator touches next in locker.

FIGURE 2. TASK 1 HIERARCHY AND ACTIVITY DESCRIPTION

Task 2:

Install Loader

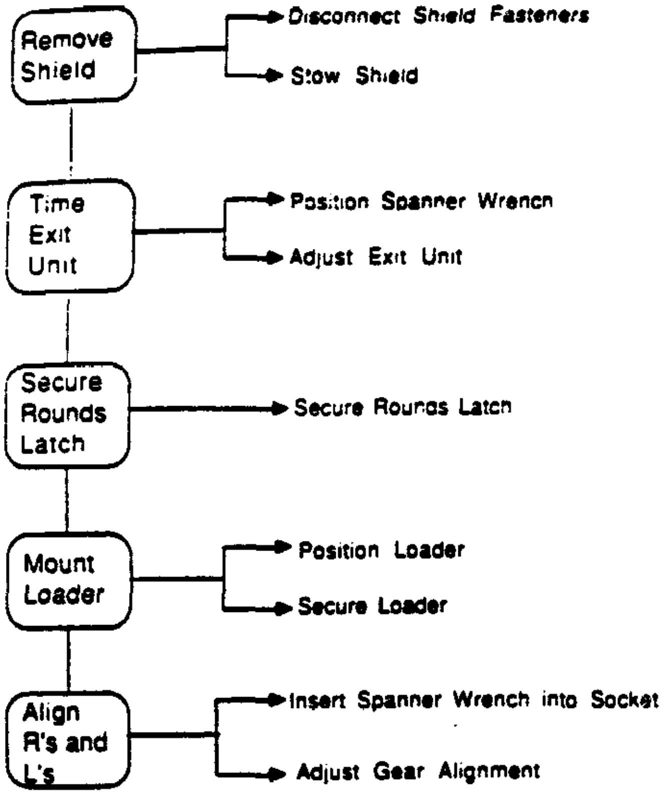


FIGURE 3. TASK 2 HIERARCHY AND ACTIVITY DESCRIPTION

TASK 2 ACTIVITY ELEMENT DESCRIPTION

Task 2: Install Loader

Subtask 2.1: Remove Shield

Task Element 2.1.1: Disconnect Shield Fasteners
Start State:
Operator lays hand(s) on fasteners/first fastener pulled
End State:
(All fasteners removed) Operator lays hand(s) on shield
lift off and set aside.

Task Element 2.1.2: Stow Shield
Start State:
Operator lays hand(s) on shield to lift off and set asid
End State:
(Operator has set shield aside) Operator releases hand
from set-aside shield/arm movement indicates shield rele

Subtask 2.2 Time Exit Unit (Ship #1)

Task Element 2.2.1: Position Spanner Wrench
Start State:
Operator grabs wrench with either hand.
End State:
Wrench pushed into wrench socket.

Task Element 2.2.2: Adjust Exit Unit
Start State:
Wrench in exit unit.
End State:
(Timing button depressed and wrench pulled out of socket
Operator releases hand(s) from set-aside wrench.

Subtask 2.2 Time Drum (Ship #2)

Task Element 2.2: Time Drum
Start State:
Operator places hand on timing button.
End State:
Operator removes hand from timing button.

Subtask 2.3 Secure Rounds Latch

Task Element 2.3.1: Secure Rounds Latch
Start State:
Operator touches latch assembly/pins
End State:
Operator releases pin with latch in downward position.

Subtask 2.4 Mount Loader

Task Element 2.4.1: Position Loader
Start State:
Option 1: Operator releases pin with latch in downward
tion.
Option 2: Operator begins to move loader to mount.
End State:
(Loader set in place) Operator places hand on any faste

Task Element 2.4.2: Secure Loader
Start State:
Operator places hand on any fastener.
End State:
Operator releases last fastener.

Subtask 2.5 Align R's and L's (Ship #1)

Task Element 2.5.2: Insert Spanner Wrench into Socket
Start State:
Operator grasps or moves spanner wrench.
End State:
Wrench placed in socket.

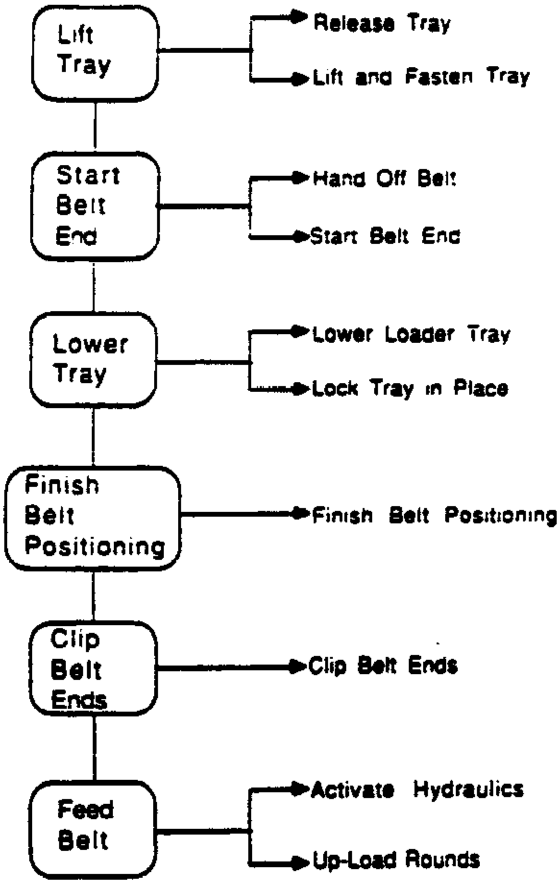
Task Element 2.5.3: Adjust Gear Alignment
Start State:
Wrench placed in socket.
End State:
(R's and L's aligned) Operator releases set-aside w

Subtask 2.5 Align R's and L's (Ship #2)

Task Element 2.5.1: Adjust Gear Alignment
Start State:
Operator bends down and looks into peep-hole to chec
status.
End State:
Operator begins to get up.

Task 3 (Ship #1):

Feed Belt/Load Rounds



TASK 3 ACTIVITY ELEMENT DESCRIPTION

Task 3: Feed Belt/Load Rounds (Ship #1)

Subtask 3.1: Lift Loader Tray

Task Element 3.1.1: Release Tray
Start State:
Operator lays hand on side tray clip.
End State:
(Loader tray unclipped) Operator releases side tray clip.

Task Element 3.1.2: Lift and Fasten Tray
Start State:
Operator begins to lift up tray.
End State:
(Tray in up position and fastened in place) Operator releases
tray.

Subtask 3.2: Start Belt End

Task Element 3.2.1: Hand Off Belt
Start State:
Belt moves into view (Deck person hands to Crewmember #2).
End State:
Belt end comes into contact with feeder.

Task Element 3.2.2: Start Belt End
Start State:
Belt end comes into contact with feeder (visual).
End State:
Operator places hand on stowed tray or tray hook.

Subtask 3.3: Lower Loader Tray

Task Element 3.3.1: Lower Tray
Start State:
Operator places hand on stowed tray or tray hook.
End State:
(Tray is in down position) Operator releases lowered tray.

Task Element 3.3.2: Lock Tray in Place
Start State:
Operator places hand on side latch.
End State:
(Tray is latched) Operator releases tray latch.

Subtask 3.4: Finish Belt Positioning

Task Element 3.4.1: Finish Belt Positioning
Start State:
(Tray is in down position) Operator places hand on wrench/wre
is turned.
End State:
Operator releases set-aside wrench.

Subtask 3.5: Clip Belt Ends

Task Element 3.5.1: Clip Belt Ends
Start State:
Operator brings ends of belts together.
End State:
Operator moves hands away from connected belts.

Subtask 3.6: Feed Belt

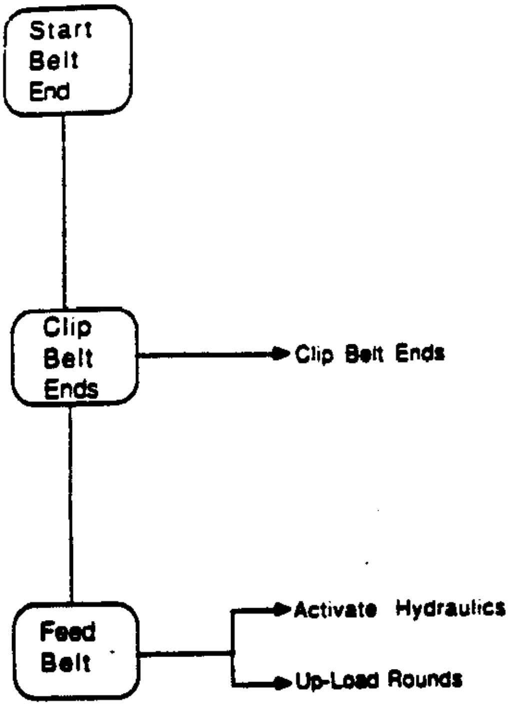
Task Element 3.6.1A: Activate Hydraulics
Start State:
Operator moves to hydraulics (walks).
End State:
Operator places hand on manual feed-rate control lever.

Task Element 3.6.2: Up-Load Rounds
Start State:
Operator places hand on manual feed-rate control lever.
End State:
End of belt exits (falls out) of loader.

FIGURE 4. TASK 3 HIERARCHY AND ACTIVITY DESCRIPTION

Task 3 (Ship #2):

Feed Belt/Load Rounds



Task 3: Feed Belt/Load Rounds (Ship #2)

Subtask 3.1: Start Belt End

Start State:
Option #1: Lift Tray (followed by Position Belt End).
Option #2: Position belt end (followed by Lift Tray).
End State:
(Tray lifted out of way, belt end started) Operator brings belt ends together.

Subtask 3.5: Clip Belt Ends

Task Element 3.5: Clip Belt Ends
Start State:
Operator brings ends of belts together.
End State:
Operator moves hands away from connected belts.

Subtask 3.6: Feed Belt

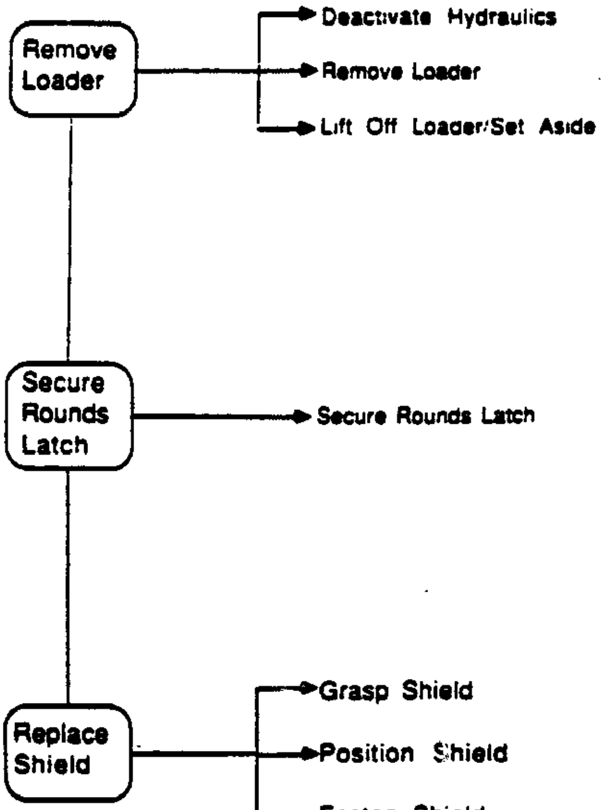
Task Element 3.6.1: Activate Hydraulics
Start State:
Operator moves to hydraulics (reaches)
End State:
Operator places hand on manual feed-rate control lever.

Task Element 3.6.2: Up-Load Rounds
Start State:
Operator places hand on manual feed-rate control lever.
End State:
End of belt exits (falls out) of loader.

FIGURE 4. (Continued)

Task 4:

Stow Loader



TASK 4 ACTIVITY ELEMENT DESCRIPTION

Task 4: Stow Loader

Subtask 4.1 Remove Loader

Task Element 4.1.1: Deactivate Hydraulics
Start State:
Belt falls free of loader tray.
End State:
Ship #1: (Operator walks around to hydraulics) Operator
from hydraulics switch.
Ship #2: (Hydraulics deactivated) Operator removes hand fro
hydraulics switch.

Task Element 4.1.2: Remove Loader
Start State:
Operator removes hand from hydraulics switch.
End State:
(All fasteners loosened) Operator begins to lift off loader

Task Element 4.1.3: Lift Off Loader/Set Aside
Start State:
Operator begins to lift off loader.
End State:
(Loader removed and set aside) Operator releases grip on loade

Subtask 4.2 Secure Rounds Latch

Task Element 4.2.1: Secure Rounds Latch
Start State:
Operator lays hand(s) on latch assembly.
End State:
Operator releases latch assembly/pins.

Subtask 4.3 Replace Shield

Task Element 4.3.1: Grasp Shield
Start State:
Operator releases latch assembly/pins.
End State:
Operator lays hand(s) on shield.

Task Element 4.3.2: Position Shield
Start State:
Operator begins to lift shield into position.
End State:
(Shield situated in place) Operator reaches for first fasten

Task Element 4.3.3: Fasten Shield
Start State:
Operator reaches for first fastener.
End State:
(All fasteners secure) Operator removes hand(s) from fastene

FIGURE 5. TASK 4 HIERARCHY AND ACTIVITY DESCRIPTION

Rather, the elements which were included were manual tasks performed by individual crew members.

Phase II: Completion Times. The next step in the data reduction effort was the calculation of completion times for the various activity elements. Once a set of activity elements had been derived and operationally defined, the analysts who performed the data reduction were instructed on how to identify those elements on the videotapes. For each observer, this instruction included an orientation to the project, a review of the CIWS loading operation's task hierarchy, an explanation of the activity descriptions (including nomenclature), and a preview of the videotapes with commentaries by the principal researchers to develop the perceptual skills needed to identify the activity elements.

In addition to calculating the activity completion times, the analysts also categorized all observed activities into one of three mutually exclusive categories: cyclic, noncyclic, and foreign. Cyclic elements were those present in every loading. For example, the task REMOVE SHIELD was part of every loading operation, and thus was included in the data reduction and modeling efforts. Noncyclic elements were those elements which were legitimately a part of the loading operation but which occurred infrequently or probabilistically. Error states were the primary noncyclic elements. These were captured as they occurred as well as the recovery procedures needed to return to a normal (cyclic) state. Foreign elements were defined as those which were not a legitimate part of the loading operation, such as lighting a cigarette. These foreign elements were not recorded as part of any other task activity but were the principal components of "dead time," i.e., the time between defined activity elements. In the case of error states, it was usually impossible to separate the error state from its recovery procedure; the two often flowed into each other. Therefore, analysts recorded the entire duration from start to end state of the error condition. A sample data collection form is shown in Figure 6. Pertinent information about a given trial was annotated on the backside of the data collection form. Start and end times were rounded to the nearest second, and task durations times were calculated as the difference between end and start times which were supplied by the clock times superimposed on the videotape.

Phase III: Sequencing. After the completion times were collected, the sequences in which these activities occurred was determined. Also identified were the person/crew positions who participated in the activity. Timelines were generated for each repetition of Tasks 2, 3, and 4. Because these timelines were referenced with respect to absolute time, they provide a graphic display of the sequential and simultaneous nature of the activity elements being modeled. Analysts indicated on these timelines who (crew position/crew identity) completed each activity element. An sample timeline is shown in Figure 7.

Timelines were generated for each repetition or trial of the loading operation. From these, flow diagrams were created, representing the sequence, allocation of person/position to the activity, relative frequency of occurrence, etc. The timelines were also used to establish sequential or

OBSERVER:

TASK:
SUBTASK:
TASK ELEMENT:

START STATE:
END STATE:

TRIAL #	SHIP	TAPE	COUNTER #	START TIME mm ss	END TIME mm ss
1				:	:
2				:	:
3				:	:
4				:	:
5				:	:
6				:	:
7				:	:
8				:	:
9				:	:
10				:	:
11				:	:
12				:	:
13				:	:
14				:	:
15				:	:
16				:	:
17				:	:
18				:	:

FIGURE 6. SAMPLE DATA COLLECTION FORM

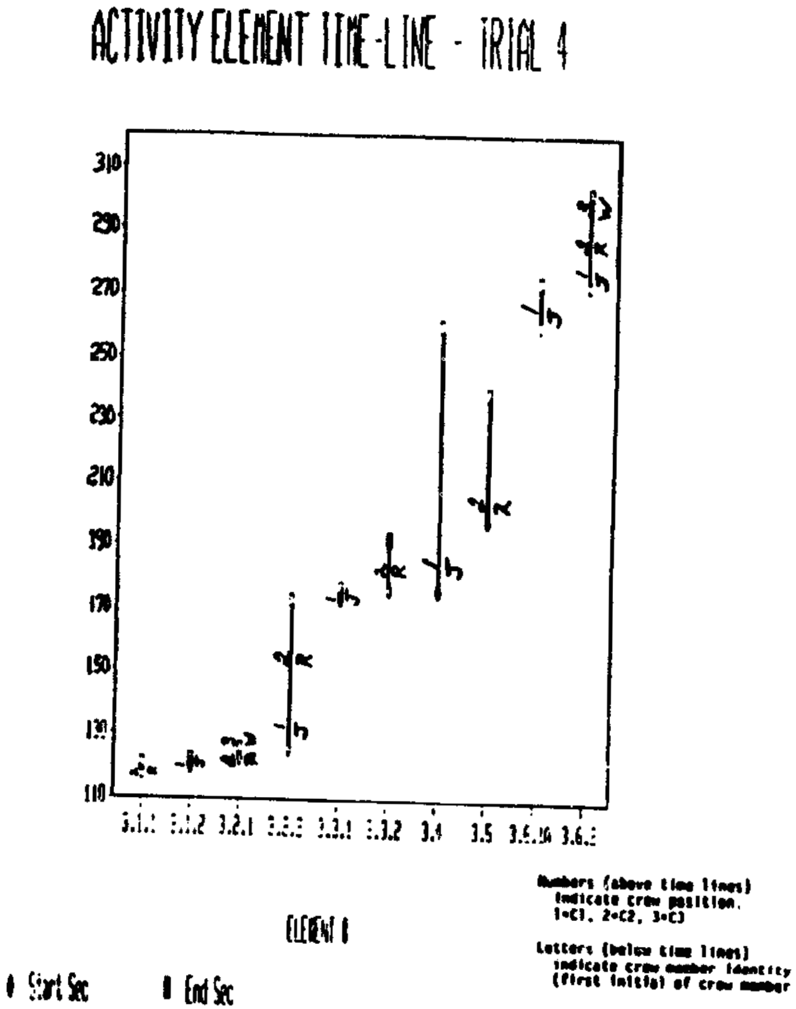


FIGURE 7. SAMPLE ACTIVITY ELEMENT TIMELINE

procedural variations. Network models were created from the flow diagrams with branches representing the alternative procedures or crew positions. In order to capture the uniqueness in crew structure and crew methodology, the two crews were modeled separately. These network models were annotated with mean time and standard deviations for each activity element. A sample flow diagram is shown in Figure 8.

Micro SAINT Model Construction. Micro SAINT is a PC-based software package which enables the user to perform network simulation of discrete systems. It uses a methodology known as sequential network modeling, in which activities are represented in a flow diagram as blocks with arrows between blocks showing of those activities. Each activity in the network is represented in Micro SAINT by three parameters: a mean, a variance, and a distribution (normal, exponential, gamma and rectangular distributions are available.) A hierarchical network structure allows the use of the concept of top down development. In building a task network, the modeler begins with major tasks followed by decomposition of these tasks into networks of smaller tasks until the desired level of detail is obtained. A network is an arrangement of tasks which represent an operation.

The menu-driven interface facilitates ease of use and obviates the need to learn a programming language. Statistics on various system parameters (i.e., completion time, crew member busy/idle time, etc.) can be collected. Since Micro SAINT is not a compiled language, the execution of models is interactive. Micro SAINT is able to evaluate algebraic expressions which gives it the computing power of programming languages such as FORTRAN. The random number generator seed is used to generate a vector of random numbers that are used to calculate the actual times for task execution and the probabilistic paths taken through the model. Using the same random number seed generates the same set of random numbers that drive model execution, a useful feature for debugging.

Every attempt was made to represent the real world data in the simulation as closely as possible. The flow diagrams were modified as needed to reflect sequencing alterations (the order in which activity elements were performed) and to reflect the crew position which accomplished any given activity element. For example, activity elements 2.1.1, 2.1.2, 2.2 and 2.3 were observed to occur in the following sequences, and each was represented as a sub-network in the model:

Alternative 1: 2.1.1 2.1.2 2.3 2.2
Alternative 2: 2.2 2.1.1 2.1.2 2.3
Alternative 3: 2.1.1 2.1.2 2.2 2.3.

The probability of taking one of these alternatives was determined by its observed frequency in the videotape data. For example, the observed frequency for Alternative 1 was 10/19, for Alternative 2 was 1/19 and for Alternative 3 was 8/19. The inputs to Micro SAINT for these activity elements were the empirical mean and standard deviation of completion time. The gamma distribution was tentatively chosen as the input distribution for all activity elements. For Alternative 1, the mean and standard deviation were derived from 10 trials, for Alternative 2, from 1 trial, and for Alternative 3 from 8

Task 2 (Ship #2)

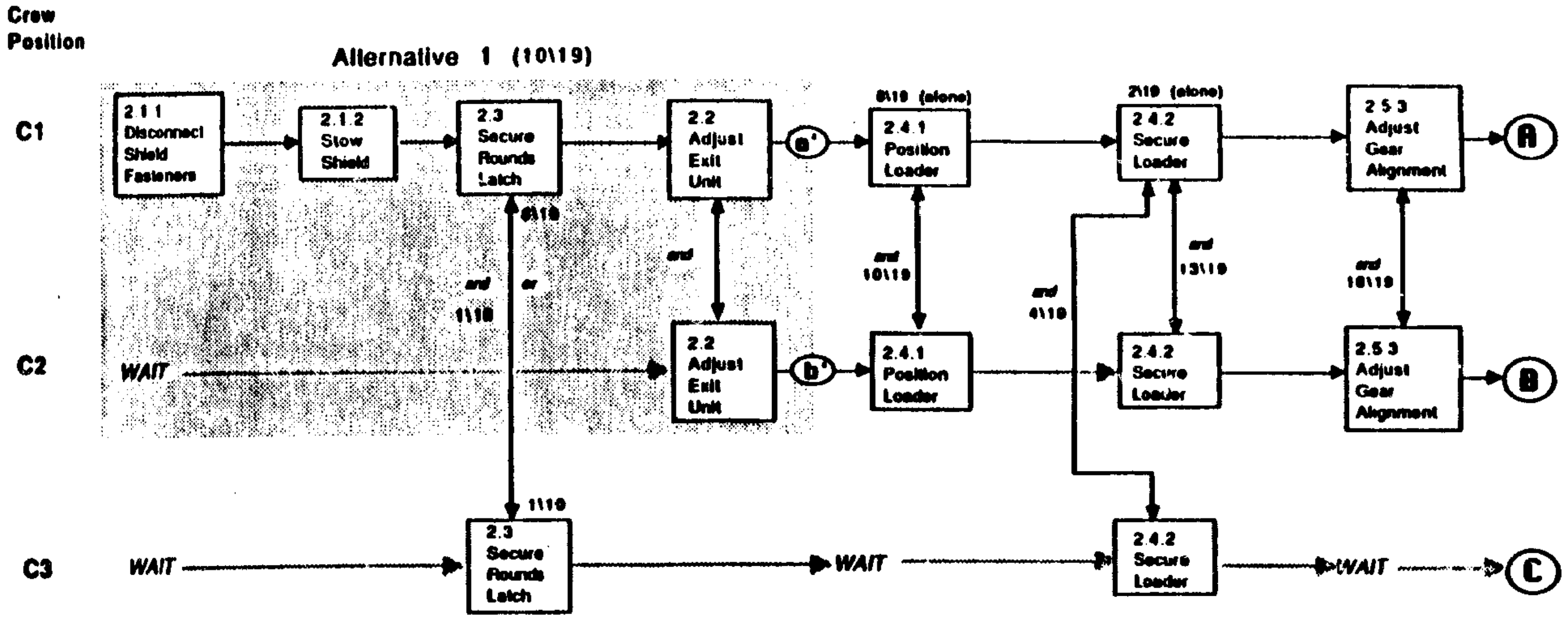


FIGURE 8. SAMPLE FLOW DIAGRAM

trials. A similar methodology was employed to account for differences in which crew position performed the activity. In this manner, the observed order and position of performance of the loading operation was preserved as closely as possible. However this fidelity to reality produced statistical artifacts: an alternative procedure may have been observed only once; in such cases the average completion time was the single observed time, and the standard deviation was zero.

The rationale for delineating alternative procedures and crew positions in the model was that these represented qualitatively different operating procedures. Changes in the sequence or manning allocation may be reflected in the overall completion times.

Validation Approach. One of the project tasks was to check the Micro SAINT models against the videotape data. There are two stages involved in checking a model. Verification is the process of checking that the simulation program operates as expected, i.e., debugging. Validation is the process of checking that the simulation model is a sufficiently close approximation to reality for the intended purpose. The credibility of any model depends upon a demonstration that the model represents at least the sample of reality from which it was derived.

A preliminary validation of the CIWS models was performed during the data collection on board the frigates. The CIWS teams provided face validation of the models as they reviewed the task inventory for completeness and correctness. They also commented upon the sequence of tasks and subtasks within the loading operation.

The Micro SAINT models were also validated by means of internal data validation. Each model was run 30 times to create a simulation data set; from this, the average complete time with a 95% confidence interval was generated. Then a check was made to determine whether or not the average completion time for the real world data was contained within the simulation's confidence interval.

Missing values were common in the sample data. Whenever an activity element was contaminated by a foreign element or error, it was dropped from the data set used for model building. A single trial or loading evolution, however, required that all activity elements making up the loading operation be present in order to develop completion times. Therefore, any missing values were estimated by the mean of the remaining values for their associated activity elements. Thus, the reader should be aware that the real world data used for the validation are derived from the actual videotapes, but do not necessarily correspond to a single trial as it was actually performed in its entirety.

The use of the confidence interval about the simulation mean was chosen for two reasons. First, such a test emphasizes uncertainty in the computer model (where the uncertainty properly belongs, from a model validation standpoint). Secondly, the probabilistic nature of the simulation provided independent runs or trials as data points for the sample of completion times, making classical statistics applicable only to the model outputs.

One other limitation with the procedure described is the fact that the trials of observed performance were not independent. "Warm-up" effects were observed in that the crew members worked faster as they gained additional practice with each repetition. Improvements in crew coordination also resulted in shorter completion times. Classical statistical techniques (with the exception of time series techniques) assume independence between observations in a sample. Making use of the simulation confidence interval bypasses this problem to some extent since model output, by definition, is independent across runs (the computer equivalent of trials.) Furthermore, the danger in violating the independence assumption is minimized, to a large extent, by careful consideration of the scope of inference for this study. As an internal validation, the scope of inference is limited only to performance of the crews upon which the models were built and the conditions under which they performed. This makes full use of all the available data more defensible. It must be remembered, however, that internal validation says nothing about how well a model describes or predicts the performance of other CIWS crews. These latter validations are separate issues in modeling and simulation.

Sensitivity Analysis Approach. The sensitivity analysis was performed by systematically varying the average task completion time for selected activity elements one at a time. These changes were made by multiplying the average time by a factor, M, which varied from 0.0 to 2.0 in 0.10 increments. The standard deviation was left unchanged. After 100 repetitions at one value of M, the next value of M was specified and 100 more repetitions of the model were run without terminating the simulation. This was repeated for 10 increments of M with 100 runs per increment; thus one simulation session consisted of 1,000 runs of the model. All sessions started with the same random number seed of 1; this was thought to ensure that observed changes in the model results were attributable to the intentional changes, and not to the random changes from different random number seeds.

The activity elements selected for modification were those considered to have the greatest impact on the total completion times: the activities in Tasks 2 and 3. These were highly critical elements; failure to perform any of these correctly would result in failure to complete the CIWS loading operation. On the other hand, neither Task 1 nor 4 was deemed critical in the loading operation. Task 1 consisted of unloading ammunition boxes and removing the rounds; Task 4 consisted of dismounting the loader and replacing the cover. The sensitivity analysis was conducted on activity elements one at a time. Simultaneous changes in more than one element can be used to determine synergistic effects on the modeled process. However, it is not feasible or cost effective to exhaustively evaluate all possible combinations. The only sensible way to manipulate multiple parameters simultaneously is to look at meaningful combinations. Because in the present work there is not clear guidance available on what such meaningful combinations might be, sensitivity analysis was restricted to individual activity elements taken one at a time.

RESULTS

Validation

Results of the validation tests for Ship #1 are present in Table 2; for Ship #2 in Table 3. These tables include the confidence intervals generated about the mean completion time for each task, and for the total loading operation. Data derived from the videotapes of the crews (denoted as real world data) and the simulation models are presented for comparison. The results are discussed below.

The originally proposed validation of the models was to build a 95% confidence interval about the simulation mean and determine whether the observed (real world) mean fell within that interval. This approach assumes there is no variability about the observed mean, an obviously false assumption. In actuality, there is always variability about a sample mean. The modified approach therefore, was to build the confidence interval about the observed mean and compare it for overlap to the confidence interval about the simulation mean. Overlap indicates that, at the 95% confidence level, the simulation is valid. No overlap indicates the opposite: at the 95% confidence level, the simulation is not validated.

The validation results for Ship #1 show validation for all tasks and marginal validation for the total loading operation. If the confidence intervals are widened to the 99% level, then the total loading operation is also validated. With respect to Ship #2, the overall operation and all tasks are valid at the 95% level with the exception of Task 1. The reasons for the discrepancy are not known.

Overall, the two simulation models appear to be well validated by the methods used; this is especially encouraging in light of the exploratory nature of the research which these models represent. However, several important caveats must be kept in mind. First, the real world sample data were not independent because the same crew performed repeated loading evolutions. Classical confidence intervals assume independence and therefore, the results reported here are approximate. Second, the real world data sets were censored to remove, within a single trial or observation, activity elements that exhibited either error states or foreign elements. In order to develop a completion time for a trial in which an activity element was dropped, it was necessary to fill that activity element with a placeholder, in this case, the means of the remaining activity elements. In summary, the results should be treated as approximate.

Sensitivity Analysis

Examination of the Micro SAINT simulation models reveals that they are largely linear in nature. All branches or paths stemming from the same node comprise the same activity elements, differing only in the sequence of those elements, or the crew position. Furthermore, all branches flow forward, i.e., there is no backward looping of activities. This is due to the fact there were, in general, no shortcuts in loading the CIWS as it was observed. By definition, the activity elements must be both comprehensive and exhaustive in

TABLE 2. VALIDATION RESULTS - SHIP #1

SHIP #1

Task	N_{rw}	X_{rw}	X_m[a]	CI_{rw}[b]	CI_m
1	11	213.23 [45.65][c]	188.98 [30.15]	(182.56, 243.89)	(177.24, 200.22)
2	16	65.61 [11.99]	64.08 [13.52]	(59.22, 72.00)	(59.03, 69.13)
3	16	123.45 [26.68]	111.45 [19.16]	(109.24, 137.66)	(104.30, 118.60)
4	8	57.42 [9.66]	54.16 [10.38]	(49.34, 65.50)	(50.28, 58.03)
ALL	20	459.70 [46.46]	418.67 [38.92]	(435.48, 483.92)	(404.14, 433.20)[d]

Notes:

a. All model statistics are based on N=30 runs.
b. Confidence intervals are calculated as:

$$CI: \bar{X} \pm t_{\alpha/2,\ n-1} (S / \sqrt{N}) \text{ where } \alpha = .05$$

c. Values in brackets are standard deviations corresponding to the means above them.

d. Lack of overall match at 95% confidence level.

TABLE 3. VALIDATION RESULTS - SHIP #2

SHIP #2

Task	N_{rw}	X_{rw}	X_m	CI_{rw}	CI_m
1	17	72.99 [5.88]	93.20 [3.87]	(69.97, 76.01)	(91.76, 94.64)[d]
2	19	97.04 [25.06]	95.12 [26.27]	(84.96, 109.12)	(85.31, 104.93)
3	20	99.38 [44.33]	88.70 [44.15]	(78.63, 120.13)	(72.22, 105.18)
4	20	43.27 [12.37]	40.93 [6.04]	(37.48, 49.06)	(38.67, 43.18)
ALL	20	312.68 [71.49]	317.94 [52.47]	(279.22, 346,14)	(298.35, 337.53)

Notes:

a. All model statistics are based on N=30 runs.

b. Confidence intervals are calculated as:

$$CI: \bar{X} \pm t_{\alpha/2,\ n-1} (S / \sqrt{N}) \text{ where } \alpha = .05$$

c. Values in brackets are standard deviations corresponding to the means above them.

d. Lack of Task 1 match for Ship #2 at 95% confidence level.

order to create a useful model. Furthermore, most of the elements were sequentially ordered, with few elements occurring simultaneously. In this regard, the uncertainty was in the completion times and the probability of taking a particular branch. Since the models excluded aberrant activities such as errors and foreign elements, they essentially capture perfect performance of the loading operation, which is appropriate for baseline data.

Tables 4 and 5 show the range of values for the activity elements for selected tasks. The elements marked with asterisks have the greatest impact on the task completion time. Changes in the task and total completion times were related to the magnitude of the permutation in the individual activity in question. Small changes in the activity completion time resulted in small changes at the task and total completion time levels; likewise, large changes at the activity element level resulted in large changes at the task and total operation levels. The changes were also proportional to the duration of the permitted activity: an element with a long completion time had a greater effect on the task and total times vis-a-vis an element with a shorter completion time. As expected, the impact on total time was smaller than on task time, since there are more component elements in total time which "dampen" the effect of one element. From the tables, it can be seen that in Ship #1, INSERT BELT END, with a mean time of 34.58 seconds, had the greatest effect on Task 3 (71% - 135%) as well as on total time (92% - 111%). Likewise, in Ship #2, CLIP BELT ENDS (mean time = 62.33) changed Task 3 by 47% - 169%, and total time by 85% - 119%. These results give some indications as to where changes could be made which would have the greatest impact on overall completion time.

DISCUSSION

Developing the Micro SAINT models of the CIWS loading operation was both challenging and enlightening. A number of important issues in manned systems simulation and modeling were encountered which required difficult decisions and were handled in the manner described (see also Tijerina and Treaster, 1987 and Treaster and Tijerina, 1988). Simulation and modeling of human behavior and performance is gaining popularity and attention. It is intrinsic to such DoD programs as MANPRINT and is a point of emphasis in many military research agendas. Yet, sequential network simulation of human performance is relatively new and many unknowns surround its application to military and other real world systems. In this section, we will briefly review some of the major "lessons learned" from our endeavors. We hope that the issues discussed here will foster more informed applications of sequential network simulation for integrating human performance and behavior in combat models.

TABLE 4. SENSITIVITY ANALYSIS RESULTS
(Ship #1)

		Task Completion Time		Total Completion Time	
Task 2	Range	% Change (seconds)	Range	% Change (seconds)	
Disconnect Shield Fasteners	2.1.1 ($\bar{x}$=1.81)	60.36-66.08	0.94-1.03	412.92-430.05	0.99-1.03
Stow Shield	2.1.2 ($\bar{x}$=2.44)	59.44-67.24	0.92-1.04	408.18-427.14	0.98-1.03
Position Wrench	2.2.1 ($\bar{x}$=1.81)	60.41-66.20	0.94-1.03	411.61-426.14	0.99-1.02
Adjust Exit Unit*	2.2.2 ($\bar{x}$=11.07)	52.32-74.51	0.81-1.16	399.81-435.25	0.96-1.05
Secure Rounds Latch	2.3 ($\bar{x}$=6.94)	56.54-70.74	0.88-1.10	411.85-431.49	0.99-1.04
Position Loader	2.4.1 ($\bar{x}$=8.94)	55.05-72.89	0.85-1.13	410.92-430.39	0.96-1.04
Secure Loader*	2.4.2 ($\bar{x}$=17.57)	47.66-81.84	0.74-1.27	402.47-439.03	0.97-1.06
Insert Wrench	2.5.2 ($\bar{x}$=1.43)	60.42-66.47	0.94-1.03	413.16-425.01	0.99-1.02
Adjust Gear Alignment*	2.5.3 ($\bar{x}$=13.54)	50.32-75.69	0.78-1.18	397.12-431.78	0.96-1.04

* Denotes activity elements which had the largest impact in Task 2.

TABLE 4. SENSITIVITY ANALYSIS RESULTS
(Ship #1)
(Continued)

Task 3	Range	Task Completion Time		Total Completion Time	
		% Change (seconds)	Range	% Change (seconds)	
Release Tray	3.1.1 ($\bar{x}$=1.71)	111.03-116.08	0.98-1.04	410.31-425.28	0.99-1.02
Lift & Fasten Tray	3.1.2 ($\bar{x}$=6.35)	104.26-118.79	0.94-1.07	408.56-428.33	0.98-1.03
Transport Belt	3.2.1 ($\bar{x}$=3.36)	107.18-119.27	0.96-1.07	411.08-428.99	0.99-1.03
Insert Belt End*	3.2.2 ($\bar{x}$=34.85)	79.04-149.80	0.71-1.35	382.30-460.44	0.92-1.11
Lower Tray	3.3.1 ($\bar{x}$=2.61)	109.85-117.78	0.99-1.06	411.30-427.03	0.99-1.03
Lock Tray in Place	3.3.2 ($\bar{x}$=6.71)	111.08-116.60	1.00-1.05	414.50-423.70	1.00-1.02
Finish Belt Positioning	3.4 ($\bar{x}$=16.62)	102.82-131.33	0.92-1.18	406.20-435.55	0.98-1.05
Clip Belt Ends	3.5 ($\bar{x}$=11.94)	101.27-126.26	0.91-1.13	405.29-432.38	0.97-1.04
Activate Hydraulics	3.6.1 ($\bar{x}$=14.20)	98.91-127.47	0.89-1.15	403.41-435.15	0.97-1.05
Upload Rounds*	3.6.2 ($\bar{x}$=24.93)	85.68-138.54	0.77-1.24	392.61-443.72	0.94-1.07

* Denotes activity elements which had the largest impact on Task 3.

TABLE 5. SENSITIVITY ANALYSIS RESULTS
(Ship #2)

Task 2, 3	Range	Task Completion Time % Change (seconds)	Range	Total Completion Time % Change (seconds)	
Disconnect Shield Fastener	2.1.1 ($\bar{x}$=7.8)	83.05-102.96	0.88-1.10	311.17-332.93	0.97-1.03
Remove Shield	2.1.2 ($\bar{x}$=3.94)	84.55-98.87	0.90-1.05	308.18-326.72	0.96-1.01
Time Exit Unit*	2.2 ($\bar{x}$=24.05)	69.38-115.12	0.74-1.23	290.02-345.03	0.90-1.07
Secure Rounds Latch	2.3 ($\bar{x}$=9.56)	81.96-101.38	0.87-1.08	306.21-335.73	0.95-1.04
Position Loader	2.4.1 ($\bar{x}$=13.39)	76.74-104.78	0.82-1.12	302.91-337.96	0.94-1.05
Secure Loader	2.4.2 ($\bar{x}$=19.15)	72.61-111.91	0.77-1.19	300.33-346.33	0.93-1.08
Adjust Gear Alignment	2.5.3 ($\bar{x}$=19.72)	76.09-112.48	0.81-1.20	299.05-343.23	0.93-1.07
Start Belt Ends*	3.1 ($\bar{x}$=62.33)	42.94-155.38	0.47-1.69	272.71-381.83	0.85-1.19
Clip Belt Ends*	3.5 ($\bar{x}$=23.1)	67.30-111.57	0.73-1.21	294.88-342.30	0.92-1.06
Activate Hydraulics	3.6.1 ($\bar{x}$=4.5)	80.17-98.93	0.87-1.07	311.28-329.44	0.97-1.02
Upload Rounds	3.6.2 ($\bar{x}$=9.45)	78.39-105.39	0.85-1.14	306.07-334.26	0.95-1.04

* Denotes activity elements which had the largest impact in Task 2 or 3.

A Model of Modeling

Figure 9 shows an overview of the process used to develop our models. It was prepared *post hoc* to conceptually capture the stages we went through in modeling the CIWS loading operation and to organize the critical assumptions made along the way. We believe Figure 9 depicts stages common to all modeling efforts. Unfortunately, books and coursework on modeling and simulation spend much more time explaining the syntax (techniques) involved rather than the semantics (meaningfulness) of the finished product. Therefore, a review of the "model of modeling" is worthwhile if it alerts the reader to at least some of the sources of bias which might make a simulation less than realistic.

In building our sequential network models of crew-system interaction, we had to collect data on the process and then build from that. Before collecting data from which to construct our models, we identified these preliminary assumptions to justify getting started:

1) There is structure in the world to be modeled. In our case, this meant we assumed there was a standardized procedure. Without this standardization we faced the possibility of building N models for N crews doing it in different ways.

2) The operation to be modeled is in a stable condition. For instance, we assumed that there would be no major change in equipment, manning, procedures, etc. which would invalidate our models.

3) The design, equipment, materials, etc. used during data collection are representative of the operation as it will exist for some reasonable period of time. For example, if the Navy decided suddenly to fully automate CIWS loading, a model of manual loading would be irrelevant.

4) Operators who serve as subjects in the data collection will know the proper method, have sufficient skill, ability, and practice for the job, yet are not exceptionally skilled. We didn't want hot shots or green recruits, since each would represent an extreme of ability.

Modeling is motivated by a desire/need to understand/predict the behavior of some aspect of reality. Ideally, this need is articulated at the outset with respect to the model's purpose (i.e., the specific questions to be answered by the model, the required precision of those answers) and the range of conditions under which the model should apply. These two points, purpose and range of conditions, deserve consideration because they constrain the level of detail, measures of merit, and the approach which will be suitable.

In the case of the CIWS loading operation, the purpose was to demonstrate that a sequential network model of crew-system interaction could be built from empirical data. The level of detail was specified to be the greatest detail

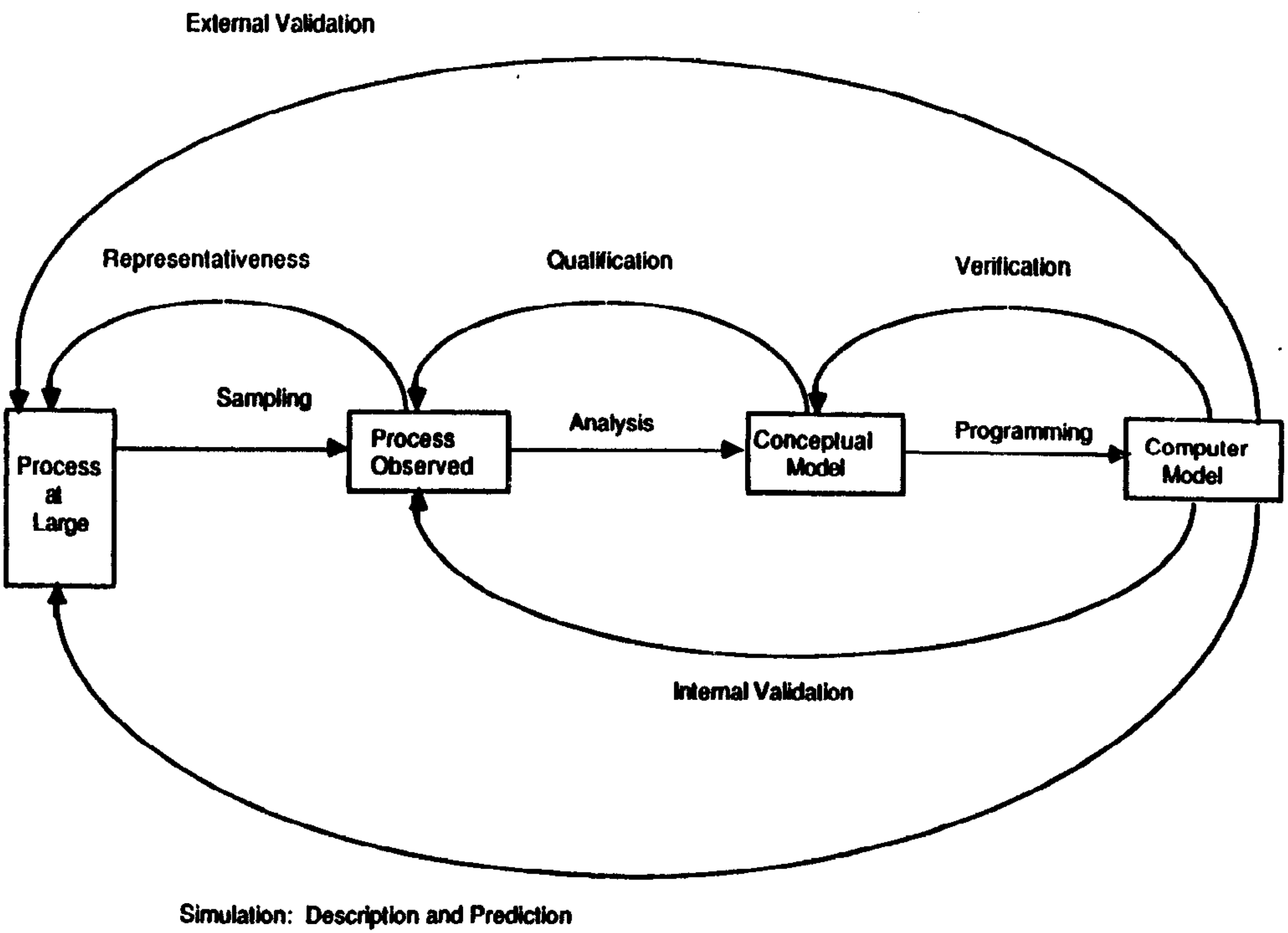


FIGURE 9. A MODEL OF MODELING

possible within the constraints of available resources and data resolution. Our technical monitor was interested in ship motion-induced biomechanical interference of topside manual handling tasks as well as procedural and manning impacts on those tasks. This influenced us to describe CIWS loading in terms of manual tasks the crew members perform, each characterized by time and accuracy. If interest had focussed instead on, say, electro-myographic activity of the back muscles or forces acting on a crew member's spinal column, our models might have been quite different.

Our models were constructed from data collected under conditions of calm seas, daytime lighting, temperate weather, and no winter apparel or individual protective equipment (IPE) worn by the CIWS crews. These are the conditions for which the models are most applicable. It was thought that activity elements could later be isolated under laboratory conditions and task time increases could be determined for conditions of MOPP, ship motion, or other performance shaping factors. This, incidentally, involves two additional assumptions. One is that it is legitimate to pull a node (or setof nodes) from the network and manipulate it in isolation, then plug it back into the model and see the impact on the operation overall. Second, it is assumed that the organization of the operation remains the same under different conditions; only the durations (and perhaps accuracies) change. This ignores the possibility that some conditions will alter the structure of the operation substantially, i.e., induce a "work around". An alteration of the sequence of tasks would make a baseline model invalid and require a new model (or parts of a model) to be build for the alternative procedure.

With purpose, scope, and feasibility in mind, data gathering may begin. A sample of reality is selected and observations are collected to provide the data for modeling. Sometimes this sample is a subset of real-world examples (e.g., a few CIWS crews from specific ships chosen from all possible crews on all possible ships). In other instances the sample is taken from a unique real-world system (e.g., simulation of the Mir space station) or from a database whose original empirical sources are no longer apparent (e.g., handbooks, databooks, predetermined time systems such as MTM, human performance databases, etc.). In any case, the sample data always consists of observations made at particular points in time under specific performance shaping factors (i.e., personnel, equipment, environmental constraints, scenarios, test instructions, etc.). The major issue at this point is the representativeness of the sample data with respect to reality. It is implicitly assumed that the sample used for modeling purposes is representative of the real world. The actual representativeness of the sample data determines the boundary conditions within which the model will be most applicable. Extrapolating beyond these conditions is a risky venture.

The sample observations are analyzed to derive a conceptual model of the real system. In the case of the CIWS loading operation, for example, analysis involved a task inventory and videotape data reduction. These efforts resulted in a conceptual model of the simultaneous and sequential nature of the loading operation, allocation of activities across crew members, and statistics (i.e., means, standard deviations, and tentative input distributions) for each of the activity elements. An iterative process of qualification is pursued to determine the adequacy of the conceptual model to provide an acceptable level of agreement for the domain of intended application (Schlesinger, et. al., 1979). For the CIWS project, this qualification process revolved around refinement of the modelers' understanding of the operation, limitations in the data reduction procedures, and directives provided by the sponsor; these points are discussed further in Tijerina and Treaster (1987).

Once the conceptual model is adequately developed, it is programmed. The simulation programming environment will impose certain ways of representing data in which the conceptual model's graphic, verbal, and quantitative aspects must be expressed. Because things can go wrong in this translation process, iterative verification (debugging) cycle is needed to ensure the conceptual and computerized versions of the model match.

A debugged computer model of reality is a suitable candidate for validation. Model validation is defined by the SCS Committee on Model Credibility as "substantiation that a computerized model, within its domain of applicability, possesses a satisfactory range of accuracy consistent with the intended application of the model" (Schlesinger, 1987, p. 104). The present authors have thought of validation as being of two types: internal to the simulation development effort and external to the simulation development effort. The first, which we call internal validation, is really an extension of the verification effort: it determines how well the computer model accurately mimics or describes the behavior of the sample from which it was built. This is a critical step because the decomposition of an operation into its component elements for modeling can introduce distortions. Internal validation provides some measure of proof that the computer model is an adequate representation of at least the sample. The scope of inference in internal validation is, of course, limited to the sample which was used to build the simulation.

The second type of validation, external validation, assesses how well the computer model describes or predicts reality which was not a part of the original sample used for model building. The scope of inference in this case is beyond the sample and into the population of interest. The tie that binds the two types of validation together is sample representativeness. The computer model should first be an adequate representation of the sample and the internal validation checks this. As the computer model checks on internal validity, then the external validation, in effect, checks on the representativeness of the sample on which the model is based. The present report contains the results of internal validation on the CIWS models. External validation assesses the degree to which other CIWS crews perform like the models. Additional predictive validation is also possible to determine the match between model and real world performance when a procedural, manning, personnel, or engineering intervention is simulated.

Some Possible Uses of Models

We would next like to consider some uses for sequential network models of crew-system interaction. Originally, the CIWS models were thought to be useful for evaluating engineering, manning, and procedural interventions to the CIWS. Bottlenecks might suggest equipment design improvements. The relationship between loading crew size and total completion time might suggest an optimal manning allocation. Changes in who does what and when could be investigated with the models in an effort to, via a procedural change, shorten the loading times required. And, as mentioned earlier, the impacts of various performance shaping factors like ship motion or MOPP gear could be assessed in principle. These are the most immediate uses of a sequential network model of an single shipboard tactical operation.

Can models like those discussed here also introduce more operational realism in modeling of combat? Depending on the uses to which modeling of combat will be put, we think the answer is 'yes'. Three specific applications are discussed below.

Combat, like other real-world operations, is carried out with finite resources. One of those resources is time. Time windows of opportunity open and then close. Tactical activities, from individual crew member tasks to battle force operations, take time to complete. Certain performance shaping factors such as individual protective equipment, high sea states, night operations, and sustained/continuous operations manifest their effects in the form of time lags and errors which require additional time for correction. Prevailing under such circumstances requires new strategies and gambits which take the time lags and loss of finesse into account. Therefore, it seems worthwhile to try and emulate the pace and rhythm of combat simulations by integrating realistic assessments of how long it takes humans to complete various tactical operations.

The combat effectiveness of a military unit, such as a ship, is a function of all the tactical operations which take place in that unit. Models of individual shipboard operations, like the CIWS models, therefore suggest a bottom-up approach. For example, one might model shipboard operations such as launch and recovery of aircraft, underway replenishment, shipboard communications, rigging shoring, and weapons loading. The completion time distributions of such models might then be rolled up, by means of a more global model, to provide more realistic estimates of the 'responsiveness' of the ship as a whole to commands from a higher authority or to the exigencies of combat. In turn, individual ship completion times could then be rolled up to yet higher levels of simulation (e.g., the battle group). This notion of micro-models (like the CIWS model) nested within macro-models in a unified simulation system like Micro SAINT should be considered further.

Another use of sequential network models of individual tactical operations might be in the form of tactical decision aids for commanders. For instance, the CIWS models might be manipulated to generate distributions of completion times as manning changes from 1 to 6 crew members. Such data could help a ship CO determine the extent to which he can trade off extra hands for time. Similarly, one might represent the thermal load of such topside operations by representing, say, the physical exertion associated with various activity elements. A sequential network model so configured could provide statistics on cumulative heat stress which would aid the officer in determining reasonable watch lengths and crew rotations. Decision aiding possibilities like this also should be explored.

A third potential application of sequential network models of individual tasks might be to provide inputs to other models. Consider the area of CBR defense. Data from field exercises suggests that wearing MOPP gear slows down performance in some combat operations, has no effects on others, and makes yet others virtually impossible to complete. Efforts have been made to develop task time multiplier models (TTMs) which will indicate the time difference (relative to completion times in a shirt-sleeve environment) to complete a task in MOPP gear. One such model arrives at these multipliers by characterizing the human performance requirements of the task in terms of a behavioral taxonomy (Ramirez, Rayle, DaPolito, and Shew, 1987). In this regard, the CIWS model and others like it could be used to provide the baseline times to which TTMs are applied. The models could also be used to determine the locus of the task degradation effect (if any), which might then suggest procedural, engineering, or training enhancements.

To our knowledge, such uses of Micro SAINT models as suggested above have yet to be tried out or evaluated; other applications to combat modeling are undoubtedly also possible. Documentation is still not available which can tell us how well such models work or for what applications they work best. Hopefully, the next few years will give us answers to some of these questions.

Conclusions

In the present study, data were collected and models were developed of the Close In Weapon System (CIWS) loading operation; these models were programmed in the Micro SAINT simulation language. An internal validation was conducted by comparison testing between the simulation results and the empirical data from which the model was built. Sensitivity analyses were also performed to determine the magnitude of parameter deviations which significantly impact total task completion performance and to assess the rank order of activity elements in terms of their relative impact on individual tasks and the loading operation as a whole. Additionally, critical tasks/paths were identified which have implications for affecting total system performance.

In general, the internal validation was successful; the model for Ship #1 was fully validated and only Task 1 was not validated for Ship #2. With regard to the sensitivity analysis, the results indicated which activity elements are most likely to yield a relatively high payoff for any enhancements to the loading operation. Conversely, these activity elements are also most likely to degrade the loading operation as a whole in the face of performance shaping factors which have negative consequences (e.g., environmental stressors such as ship motion).

A number of interesting conclusions can be offered with respect to the methodology used. Many issues germane to simulation of manned systems surfaced during the project and these were dealt with in the best ways known to the investigators. In general, however, the simulation approach proved to be both data intensive and analytically demanding. The proliferation of easy-to-use simulation modeling tools is a substantial accomplishment worthy of praise. However, easy-to-use simulation software should not be taken to imply that manned systems simulation is easy to perform. In the present case, the investigators could only proceed by making many assumptions; it is indeed true that we buy data with assumptions. The degree to which these assumptions will support the conclusions drawn from the models remains to be seen from attempts to use them.

References

Archer, R., Drews, C., Laughery, R., Dahl, S., & Hegge, F. (1986). Data on the usability of Micro SAINT. *IEEE 1986 National Aerospace and Electronics Conference NAECON 86*, *3*, 855-858.

Drews, C. (1986). *Anybody can simulate! A discussion of Micro SAINT*. Paper presented at the 1986 Eastern Simulation Conference, Norfolk, VA, March 10, 1986.

Ramirez, T., Rayle, M., DaPolito, F. & Shew, R. (1987). *Improvements to the Method for Determining Task Time Increase Caused by the Individual Protective Ensemble*. Paper presented at the Human Factors Society 31st Annual Meeting.

Schlesinger, S. et al. (1979). Terminology for model credibility. *Simulation*, *32*(3), 103-104.

Tijerina, L., & Treaster, D. (1987, April). *Close-In Weapon System (CIWS) Part I: Micro-SAINT model development* (Contract No. DAAH001-84-D-0001, D.O. 0062). Columbus, OH: Battelle Columbus Division.

Treaster, D., & Tijerina, L. (1988, April). *Micro SAINT Modeling of the Close-In Weapon System Loading Operation: Internal Validation and Sensitivity Analysis*. (Contract No. DAAL03-86-D-0001, D.O. 0494). Columbus, OH: Battelle Columbus Division.

DISCUSSION OF "TASK NETWORK MODELING CONSTRUCTS"
by R. Laughery
"CREW DRILL MODELS FOR OPERATIONAL T&E"
by S. Bolin, N. Nicholson and E. Smootz
and
"MICRO SAINT MODELING OF THE CLOSE-IN WEAPON SYSTEM LOADING OPERATION"
by L. Tijerina and D. Treaster

DISCUSSANT: Susan Dahl, Micro Analysis and Design

The point of Dr. Laughery's paper was to present three potential methods through which the Human Factors community could help the combat modelers better incorporate human variables in their combat models.

Dr. Laughery offers quantitative human sub-models one of the potential solutions to this long-standing and complex problem. While it is true that the concept is straightforward and intuitively pleasing, it may take a lot of effort to implement. In order to implement the method we have to be able to pull the human tasks out of the existing combat model. To some degree that division of tasks is dynamic because it is a function of the level of automation in changing weapon system designs. In other words, it seems that the function allocation to humans or systems will heavily impact the entry points for the new human performance submodels.

That isn't meant to say that this particular concept doesn't have value. As we build and modify combat models, it will be of great help if any modifications can be made with "building in this separability" in mind.

The reasons I just mentioned lead me to believe that the most promising approach right now, is Dr. Laughery's idea of using human performance models as "preprocessor predictors" for the combat models. As Dr. Laughery mentions in his paper, this solves many problems with integrating the human performance into combat models. Unfortunately, this approach also has a cost, and that is a loss of fidelity when compared to the "perfect solution".

An obvious recommendation that would come from accepting the viability of these concepts is that our human performance modeling community should try to apply such a method in an effort to build a strawman which could be evaluated to determine whether it does yield an improved combat model as a result of the additional human performance input. This exercise would results in "lessons learned" which would aid our ability to expand these concepts in later applications.

Dr. Bolin's paper provides a very nice follow-up to the concepts presented by Dr. Laughery, by providing a specific example of how task network modeling could be added to a combat scenario and useful results could be realized. In Dr. Bolin's example, crew drill models become the building blocks for Dr. Laughery's concept.

One of the most interesting methods used by Dr. Bolin's team were the operational sequence diagrams. These were new to me, but they provided a very effective method of illustrating the performance being modeled and inter-task dependencies. I encourage any of you not familiar with these to make a point of noticing them when you read Dr. Bolin's paper.

One comment on the models themselves, Dr. Bolin and his team used rectangular time distributions to model performance on each task. Also, error rates were modeled as simple repetitions. I believe that these two time-based relationships may actually be biasing the mission performance time by artificially increasing the variance. In other words, it would be interesting to see whether when some weight came off the tails of the task performance distribution (and we could accomplish this by using normals or gammas instead of the rectangular) we would increase the central tendency and therefore decrease the variance. I believe that this is on Dr. Bolin's "to do" list, and the results will be interesting, although I don't believe it will affect the general trends that we currently see in his output.

The implications of this study are discussed quite thoroughly in the paper.

All in all, this is an extremely well written paper. It is easy to read and provides a very nice example of how we should proceed towards accomplishing the goals put forth in Dr. Laughery's paper.

The third and final paper I will discuss comes from the folks at Battelle. This paper also fits very nicely with the session because it provides a thorough discussion of not only a specific modeling application but also the trials and tribulations that can affect a model validation effort.

Dr. Tijerina notes that validation is difficult because it is hard to decide when a task really begins and ends by simply watching the task or viewing a videotape. This difficulty is exacerbated by the inconsistencies in the CIWS in the way that different crews performed the same tasks. These are both "real world" difficulties. However, in order to represent the real world more correctly, the models themselves would need to be much larger and more complex. After viewing Dr. Tijerina's results, we have to ask ourselves whether the cost of increased fidelity will buy us all that much in improved results.

Also, if it were true that the existing model didn't account for error states or foreign elements, then it seems reasonable to expect the model performance time should be consistently less than the measured time, because the model should reflect "perfect performance". However, this is not what happened in Dr. Tijerina's model. Because of this, it seems likely that the time to make and recover from an error is somehow already included in the performance times for each "error-prone" task. For this reason, I suspect that the cost-benefit tradeoff of increasing fidelity would have to be evaluated for each separate instance of modeling, rather than deciding that increased fidelity is always better and therefore should be pursued.

Finally, the general comments made near the end of Dr. Tijerina's paper regarding the modeling process itself reiterate a very important point. The objective of the modeling study drives the model development process. This seems obvious on the surface, however my experience with modeling studies has shown that it is not all that unusual for people to build a model, run it, and then find out that it doesn't answer the questions they are interested in. This strengthens the case for up-front analysis.

I think that we can all appreciate the amount of effort that goes into a validation effort like the one completed by Battelle. It is encouraging to see that so much thought and energy is devoted to the least glamourous element of a modeling study.

In conclusion, the three papers complement each other nicely and demonstrate that we have some very intelligent and dedicated people working on the problems of modeling human performance in military tasks. The three papers also show that computer-based human performance simulation has finally emerged from the realm of just computer scientists and programmers. The authors agree that tools such as Micro SAINT are indeed appropriate for studies of this type and that the extension of the Micro SAINT models to the larger problem of studying how human performance impacts combat models is promising. They are to be commended for their work.

SAINT PERFORMANCE ASSESSMENT MODEL OF A SURFACE TO AIR MISSILE (SAM) SYSTEM

Gerald P. Chubb
Constance M. Hoyland

SofTech, Incorporated

INTRODUCTION

Model development requires use of modeling tools even if that means nothing more than paper and pencil along with one's training in a science or engineering discipline. The use of any tool depends on skill in application. The product therefore depends on at least two factors, the quality of the tools and the skill of their user(s). When multiple users are needed to put the tools used to work producing a final model, there is a further need to organize the model development and implementation process to facilitate communication, assure proper coordination of effort, and provide a basis for controlling and documenting what is done. This is especially important for large, multidisciplinary efforts.

This paper is an extension to an earlier treatment by Chubb and Hoyland (1988). It illustrates an approach to model development, implementation, and validation that was proven effective in a series of studies. Since then it has been adapted by SofTech as the basis for other modeling and simulation efforts. It therefore appears to be a generalizable approach. The concept itself has been termed IDEAL (Integrated Design Evaluation and Analysis Language) by its sponsors, the Armstrong Aerospace Medical Research Laboratory. IDEAL is based on the combined use of two previously developed tools. The first is SofTech's Structured Analysis and Design Technique (SADT™), also known as IDEF$_0$ (ICAM Definition language). IDEF$_0$ is a non-proprietary version of SADT provided to the Air Force as part of the Integrated Computer Aided Manufacturing (ICAM) program. SADT and IDEF$_0$ are therefore virtual equivalents (though not strictly identical). The second tool used in IDEAL is the Air Force developed Systems Analysis of Integrated Networks of Tasks (SAINT). IDEAL provides a systematic method for developing a static description of system functions before building a dynamic model of system behavior. This approach is an integral part of what Wallace, Stockenberg, and Charette (1987) refer to as Unified System Development Methodology.

This introduction will review the historical evolution of the IDEAL concept. It will also provide additional references to each of the component parts (SADT and SAINT). The application of the concept will be discussed in the context of an Air Defense missile system, although the entirety of that model will not be presented because of its scope and detail. The progressive

---

™SADT is a Trademark of SofTech, Inc.

implementation, validation, and use of that model will be briefly reviewed, and its subsequent extension to treat the effects of exposure to chemical weapons will be described.

SADT or IDEF : Functional Decomposition

SADT began as a software requirements engineering tool (Marca and McGowan, 1988). It was subsequently extended to other applications (Ross, 1985) where a top-down, hierarchical decomposition was needed in order to describe a system, particularly when the intent was to examine and describe how that system does or should function. The principal goal of SADT or IDEF$_0$ is to provide a structured approach for breaking a complex system down into more elemental components that are simpler to deal with. This is the basis for contemporary system engineering practices where a large, multidisciplinary team must be employed to design, develop, and produce some product.

The original Greek root for our word analysis is λύω. The meaning of that Greek word is enlightening. While it means to loose or free, it also may mean to break or destroy. The Greeks apparently recognized that breaking something down can destroy the nature of the whole. This is well recognized in biology where studies of living organisms done in vivo may provide different results than laboratory studies of specific organs done in vitro. A similar problem was noted by the Gestalt psychologists studying perception when they recognized that the whole is something more than the aggregate of its parts. During partitioning one tries to keep account of both functional structure and certain relationships. SADT is a disciplined analysis method that keeps track of functional structure and data relationships. SADT has proven especially useful when a team of analysts must be used to get a complete, accurate description of a large, complex system. It has been used to describe hardware, software, and human functions in the context of weapon system operation, manufacturing process definition, and other applications. It is a general-purpose systems engineering tool.

SAINT: Dynamic Behavioral Modeling

A static description of the system's functional structure does not convey the evolution of performance over time. That takes a model capable of representing the dynamic aspects of real-time system behavior. Those behavioral dynamics capture the sequential dependencies among groups of tasks, the uncertainties of concurrent activity sequences, variations in activity duration, conflicts in resources demands, situation specific rule implementation, and a variety of other interactions between operators, tasks, the equipment, and the environment. SAINT was developed to provide a tool for analyzing the behavior of large, complex systems (Seifert and Chubb, 1978). The general character of SAINT and some simple models were presented in Chubb (1981). Early applications of SAINT were reviewed by Seifert and Chubb (1978), and more recent applications were subsequently reviewed by Chubb (1986). Some additional examples will be briefly discussed later. Historically, SAINT was the chronological precursor to the Simulation Language for Alternative Modeling (SLAM), a proprietary product (Pritsker, 1987).

SAINT is non-proprietary and similar in many ways to SLAM. SAINT provides a general-purpose, FORTRAN-based, simulation language within a network-oriented framework. It can be used to exercise Petri-net representations quite easily, but is not limited to them. It has often been used to represent and assess the implications of a task analysis, but not every use of SAINT has to begin with task analysis. Nor does every use have to begin with SADT. Chubb, Stodolski, Fleming, and Hassoun (1987) recently used SAINT to perform sensitivity analyses of a closed-form analytic model of pilot workload.

This flexibility in the use of SAINT is both a strength and a weakness. As a strength, SAINT's flexibility makes it generally useful to a very broad class of systems modeling problems. The wide variety of SAINT applications to date aptly attests to this utility. As a weakness, SAINT's flexibility does not guide the modeler into any systematic definition of the problem or its representation. Often, the biggest difficulty one faces in solving system design problems is to define the nature of the problem itself and to decide how that problem can be resolved by studying a suitably constructed representation of the system's dynamic behavior. Typically, this requires an identification of the performance requirements the system must meet. Then an analysis of the behavioral model is performed to discern whether a particular design can be expected to meet or exceed those requirements. If not, the design may need to be changed, the requirements relaxed, or both.

IDEAL: Combining SADT and SAINT

Backert, Evers, and Santucci (1981) first described the need for a front-end analysis tool as an aid for SAINT model development. They identified the close correspondence between concepts used in SADT and similar constructs in SAINT. They also noted where there were gaps in the transition from SADT to SAINT. These gaps must be bridged by building a performance data base that captures information needed for implementing and executing a SAINT model. The performance data base contents are based upon the functions and data relationships identified in a static, structural SADT model of the system, but these data are then augmented with descriptions of how activity duration will be specified, how activity sequencing will be controlled, and what may affect the values assigned to various attributes incorporated into the model. These attributes typically describe characteristics of the system, resources, tasks, and information. Activity completion often affects such variables, changing their value.

Bachert, Evers, and Santucci (1981) also describe the modeling and simulation process in an SADT diagram. Figure 1 is a slightly modified version of that description. The activities are described by the verbs in the boxes. The results of an activity are shown as labels on the lines leaving a box on its right side. The input data required to perform that activity are shown by lines entering the box from the left. Arrows at the top of a box signify control data that will influence how an activity is executed. Controls may describe conditional dependencies that affect the implementation of a particular function. Finally, arrows entering a box from below are termed mechanisms. They are the means by which the function or activity is performed. Mechanisms are usually synonymous with resources. Resources may be hardware, software, people, or anatomical (or cognitive) components one wants to treat as resources. In some cases (for example, workload analyses), it may be desirable to represent anatomical features of the individual human operator, such as eyes, right hand, left hand, etc., or mental resources.

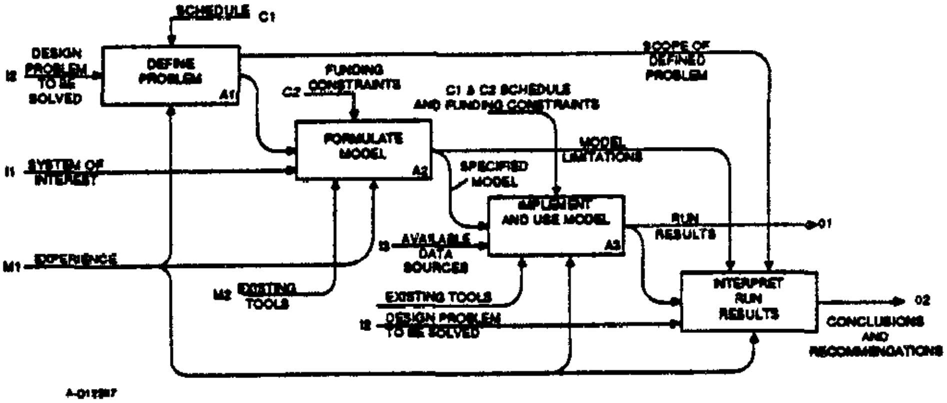


Figure 1. The Simulation Process

As the diagram implies, there are important aspects of modeling that must be addressed besides describing the system and representing its dynamic behavior. One must also identify the objective of the studies to be done with the model. This will define what outputs the model should produce, what functions require more or less detailed treatment, and the quality and quantity of input data that will be needed to support model development and validation. Figure 1 also implies that the quality of the behavioral model is critically dependent on the quality of the system description available to the modeler. The utility of SADT for SAINT model development is that it forces a careful examination and integration of the available system technical information. IDEAL further supports the SADT-to-SAINT transition by suggesting a standardized Performance Data Base (PDB) format for organizing source document information that will be needed during model implementation.

Under the Cockpit Automation Technology (CAT) Program, SofTech has also recently developed a customized version of SAINT called C-SAINT. C-SAINT is upward compatible with SAINT. Everything done in SAINT can also be done in C-SAINT. However, several limitations of SAINT have been relaxed and some new modeling features have been added. About half of the original SAINT source code has also been rewritten in FORTRAN 77 using structured programming techniques. C-SAINT features briefly reviewed in Chubb, Hoyland, and Ganote (1988) are more thoroughly treated in Hoyland, Chubb, and Evers (1988). The C-SAINT User's Manual supplements the existing SAINT User's Manual (Wortman, et al., 1978) and provides: a) a general introduction to modeling and simulation (with references to pertinent literature), and b) a hypothetical model of an avionics architecture/advanced cockpit to illustrate possible uses of the new modeling features incorporated into C-SAINT.

For its own use, SofTech has developed several versions of SAINT for use on IBM/PC and Macintosh computers (512, Plus, and II). While these are not sold as commercial products, they are made available to use, with some restricted rights. The VAX version of C-SAINT that Boeing is delivering to the Air Force is wholly non-proprietary.

IDEAL DEVELOPMENT OF A SURFACE TO AIR MISSILE MODEL

The IDEAL methodology was first applied to the description of a surface to air missile (SAM) battery (Bachert, Evers, Boyland, and Rolek, 1982). This particular paper was not approved for public release before the conference but is available from the authors now. In this particular case, the SAM battery was actually a generic simulator rather than a specific system. However, it was instrumented for studies of human operator performance in the context of air defense. By suitable parameterization of simulator dynamics, specific systems can be emulated. The SAM Simulator therefore provided an excellent test bed for applying the IDEAL concept and validating the resultant SAINT model. The simulator was designed to use three trained operators to acquire, track, and engage simulated penetrating aircraft. Simulating missile launch, the impact of operator tracking errors and strategy shifts then could be evaluated in terms of miss distance. By using an Air Force approved missile flyout model, it was also possible to estimate aircraft probability of hit, and probability of kill. Consequently both proximal and distal criteria of operator performance were identified and measured.

Figure 2 presents a schematic diagram of this system. Three sensor subsystems were included: 1) an acquisition radar, 2) a tracking radar, and 3) a tracking television. Three responsibilities were allocated among the

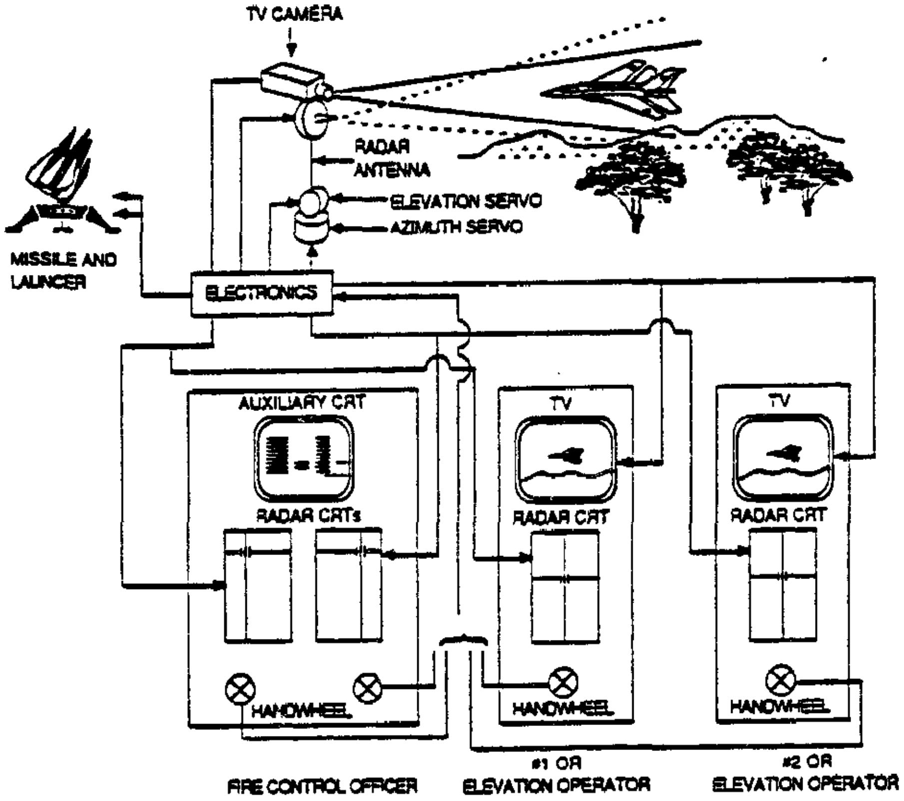


Figure 2. Schematic of the Generic SAM Simulator

operators: a) elevation tracking, b) azimuth tracking, and c) launch control. The last of these three responsibilities was given to an operator called the Fire Control Operator (FCO). The FCO was also responsible for aircraft acquisition and overall system control, as well as missile launch decisions. Two tracking operators shared the responsibility for keeping track of where the aircraft was. This was to be done in the azimuth-elevation plane. Thus, the two-axis problem was reduced to a single-axis tracking problem for each of two operators. The FCO monitored the coordinated activity of the two tracking operators and decided when their joint error was small enough to permit successful launch, once the aircraft was within range.

As a detected aircraft flies inbound toward the SAM site, the crew must evaluate any Command, Control, and Communication (C^3) data available from an external area defense surveillance team. This forms a perceptual set influencing the acquisition phase of their mission. The next major decision after acquisition is determining when to enter the tracking mode and selecting whether tracking will be done by radar or by television. The terminal decision is whether to launch a missile or not. Figure 3 identifies the overall decision sequence the crew must execute.

Static Model Development

Table 1 presents a ten-step model development process that has proven useful for guiding the evolution, test, and use of a model. What one chooses to do in modeling a system depends on a number of factors. These include the training and experience of the modelers, the people they are working for, and whoever might review the results or conclusions drawn from the model. These are not irrelevant to problem formulation, but they are issues independent of IDEAL.

The stages in table 1 are listed sequentially, but in practice there may be a reordering or even repetition and looping among these various stages. This is an idealized description of a much more dynamic process, especially when it involves more than one person.

A seven person team was involved in SADT model construction for the SAINT Performance Assessment Model of a SAM System (SPAMSS). This team included subject matter experts (familiar with the SAM system), human factors experts, and project managers. The SADT model was developed in a set of IDEF$_0$ diagrams to produce a static model of the system's functional structure. An IDEF$_0$ diagram results in a top-down hierarchical decomposition of system functions. The process of developing IDEF$_0$ or SADT diagrams is more completely described in Marca and McGowan (1988). Details of the system's architecture and operation are progressively refined in greater detail as the decomposition continues to break apart a parent function into its children. The relationships among these children (siblings) are also described at each level in the decomposition. Emphasis is placed on identifying logical dependencies and data relationships among the various functions at any particular level. At this stage of modeling, the sequencing and duration of those functions is intentionally ignored. Sequencing and duration of activities is later treated as the performance data base is constructed to implement the dynamic behavioral model of system performance. This allows hiding certain details at

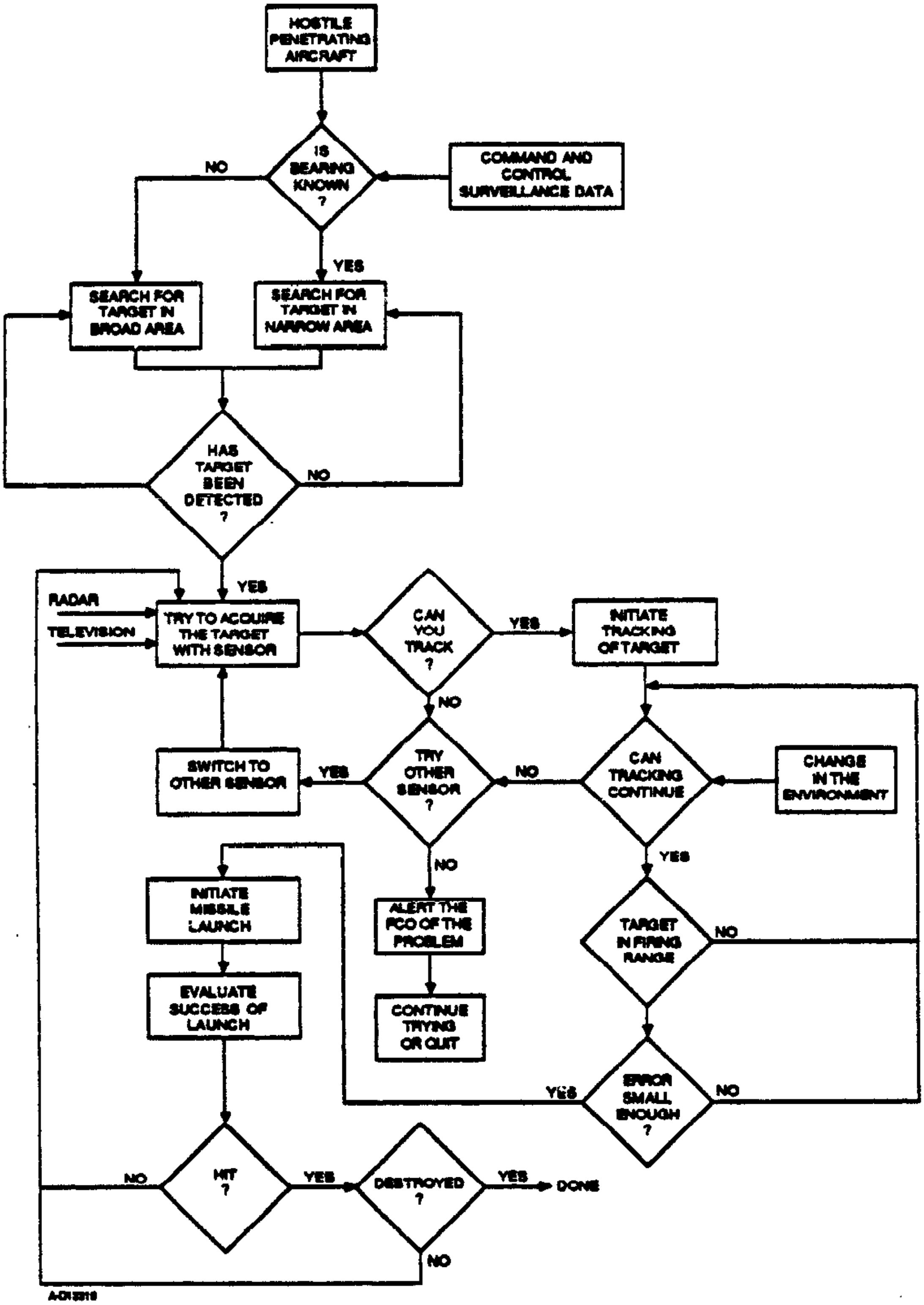


Figure 3. Overall Decision Sequence

TABLE 1. STAGES OF MODEL DEVELOPMENT

(1) Problem Formulation

The definition of the problem-solving objective. What design uncertainties need to be resolved? How will system performance be evaluated so alternatives can be compared? What model outputs are needed as dependent variables?

(2) Model Building

The abstraction of the human/machine system into mathematical/logical relationships in accordance with the problem formulation. What is critical? What can be ignored, left out, or treated later? This is where IDEAL is applied, specifically SADT.

(3) Data Acquisition

The identification, specification, and collection of data. For example, estimates of subtask durations, the frequency/probability of taking optional pathways, etc. This is where the Performance Data Base (PDB) is generated.

(4) Model Translation

Preparing the model for computer processing as prescribed by the selected simulation language (e.g., applying the contents of the SAINT User's manual, as recently updated by the C-SAINT User's Manual).

(5) Verification

The process of ensuring that the computer executes as intended. In this context, this step is done when no error codes occur and the run results appear reasonable.

(6) Validation

The process of establishing that the desired accuracy or correspondence exists between the simulation model behavior (reflected in run results) and what is known about, has been observed, or has been measured with respect to the real system's behavior. This requires subsequent study using results from empirical testing (e.g., using results from a prior experiment's data, from other studies, or results from new experiments.)

(7) Strategic and Tactical Planning

The process of establishing the experimental conditions for using the model, including statistical experimental design. Identify the independent variables, range of variation, and number of levels; specify what will be held constant; and establish what factors will be randomized across treatment conditions.

(8) Experimentation

The execution of the simulation model to obtain output values that achieve the desired precision in estimating dependent variable statistics.

(9) Analysis of Results

The process of analyzing the simulation outputs to draw inferences, fit regression equations or response surfaces to run results, and make recommendations for problem resolution.

(10) Utilization

The process of implementing decisions resulting from the analysis of simulation results. This step is crucial, but not really technical in nature. It is important at the beginning to remember that this is the payoff. Consider this step in all of the foregoing steps, especially problem formulation.

one level of development, revealing or dealing with them later when the modeling problem is better understood.

The Author of the IDEF$_0$ diagrams is responsible for identifying the purpose and perspective for the static description. The first step is to define the context within which decomposition will occur. This consists of identifying system outputs, inputs, mechanisms, and controls at a global level. The Readers of the diagram have the responsibility of making explicit any objections or questions they may have about the stated purpose, perspective, or context as presented by the Author. All disagreements over terminology are resolved before proceeding to the next level of decomposition. This assures that all participants agree on the objectives, definitions, and descriptions as they are initially stated, later modified (if necessary), and ultimately validated as a correct representation of known facts about the system. An IDEF$_0$ diagram is not unique. If the purpose or perspective is changed, the nature of the decomposition of system functions may change as well. Each team member may have access to different information about various aspects of the system's operation, based on their training, experience, and understanding of available documentation. Development of the IDEF$_0$ model as a precursor to developing the SAINT model of system behavior serves an important role: it makes sure the known facts about the system are gathered together, interpreted in a well-integrated description, and certified as acceptably correct by all members of the project. In that role, it serves as a knowledge engineering tool.

This process incorporates information such as the system's specifications, various operating procedures, and mission performance parameters. It focuses attention on where source data are available to define various aspects of system function, and where there are data voids that need to be filled by analysis, measurement, or speculation. Thus the development of this static model can eliminate building a non-supportable model behavioral dynamics.

This methodical, systematic evolution may seem frustrating to those who want quick results, but in large-scale modeling, this process precludes disaster. It is worth the investment because it reduces the risk of developing a good model for the wrong purpose. It also tends to assure you get a good model and one that others will accept. Premature entry into behavioral modeling may inadvertently leave important details undiscovered until late in development when it becomes more difficult to implement changes. Moreover, the IDEF$_0$ description provides a foundation for building the road map that will guide the behavioral model development. It forces the team to agree on what level of detail is sufficient, how to quantify various aspects of the problem, and what factors will be intentionally ignored or suppressed. Bachert, et al., claim the use of SADT produces a 60% savings compared to building SAINT models less systematically (Bachert, Evers, Hoyland, and Rolek, 1983).

Performance Data Base Construction

The Performance Data Base (PDB) is the bridge between SADT and SAINT. It serves to map the IDEF$_0$ model into a corresponding SAINT model. This transition is essentially a two-step process. First, the modeler examines the

IDEF$_0$ diagram and fills in the slots in a PDB frame. Second, these data are then used to construct the SAINT representation of behavioral dynamics that corresponds with the IDEF$_0$ decomposition of system functions. Table 2 summarizes the linkage between IDEF$_0$ constructs and SAINT modeling concepts. Table 3 briefly describes the SAINT concepts identified as column headings in table 2. This is an abridged list. It only shows those SAINT concepts directly linked with IDEF$_0$ constructs. The letters in the cells of the mapping matrix (table 2) refer to the following major categories of information:

a. Global System Characteristics - variables and their parameters that relate to all functions and system constraints.

b. Scenario Specific State Conditions - variables and conditions relating functions and information flow through the network for various operational scenarios.

c. Resource Attributes - a list of machine or operator attributes, the values of which describe such things as physical characteristics, stress levels, skill levels, etc.

d. Function or Task Characteristics - details about an activity, may include label, time statistics, priority level, probability of successful completion, precedence relationships, resource requirements, etc.

e. Environmental Factors - variables which influence an activity, e.g., lighting conditions, noise, etc., should also be noted.

TABLE 2. IDEF$_0$ SAINT MAPPING MATRIX

SAINT / IDEF$_0$	LABL	TIME	PREDECESSOR	RESOURCE	PRIORITY	INCM	DMOD	UTCH	SWIT	MODRF	BRANCHING	STATE VAR	MONITOR	CLEARING	NETWORK
INPUT			d			b, d	b				b		b		
FUNCTION	d	d						e		d		b		d	
CONTROL					d	b, d	d		d		d				a, b
MECHANISM		c		c				e		c, e	c			c, e	
OUTPUT											a, b, c d			d	
NETWORK		b							b	b, c, e	a, b, c d	b	b,d	b,c	a, b

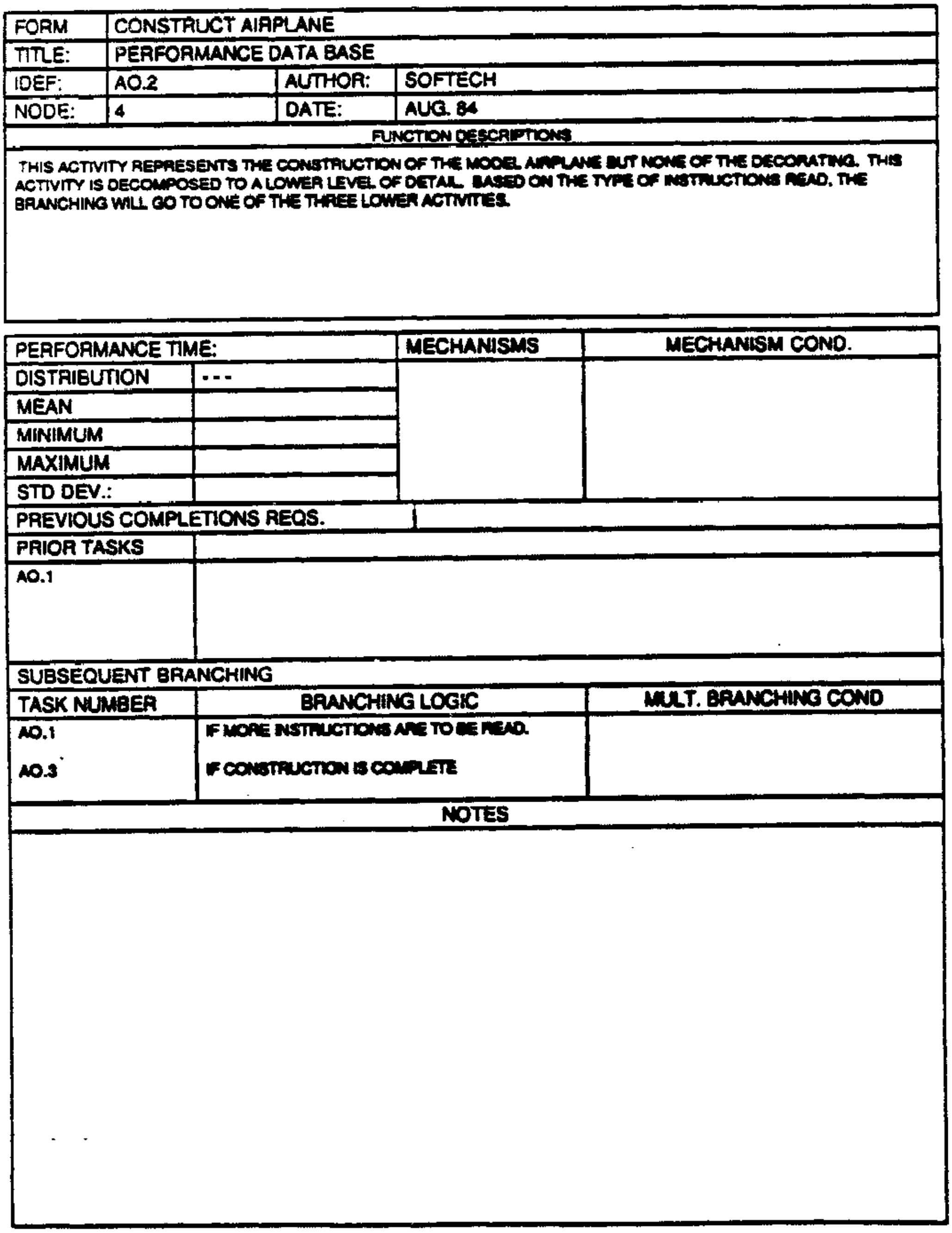
A-D13341

TABLE 3. SAINT MODELING CONCEPTS (ABRIDGED)

LABL:	An eight character mnemonic used as a task or activity name (label).
TIME:	The duration of a task or activity. Task duration may be defined by a Monte Carlo sampling from a specified distribution, as a referenced moderator function (see MODRF below) or as a fixed, constant value, or as a computed value from a referenced moderator function (see MODRF below)
PREDECESSOR:	A prior event which must have occurred before the present activity is permitted to begin.
RESOURCE:	If required (and some activities may not require resources), the activity is not to begin until one (or all) of the specified resources is available.
PRIORITY:	Controls which activity will be started first if more than one could begin at the same time.
INCM:	Information choice mechanism for determining which arriving SAINT information packet will be saved and examined (the first arriver, last, or some other option); information packets flow through a SAINT network and packet contents may be used in many different ways for modeling system dynamics.
DMOD:	Distribution modification allows TIME samples to be drawn from an alternate distribution when some particular event occurs.
UTCH:	User task characteristics are scalar values the modeler may use to describe the attributes of a task (e.g., level of difficulty, task complexity, etc.).
SWIT:	Switches or binary state indicators that may be used to represent status changes (mode change, indicator light, etc.).
MODRF:	Moderator function defining activity duration, by an equation. User code in a FORTRAN subroutine so SAINT can call it; in C-SAINT, a set of predefined functions are available for defining such equations.
BRANCHING:	Control of node exits; determines what comes next after the current activity is completed.
STATE VARIABLE:	A continuously changing value (like aircraft azimuth and elevation as it flies by the SAM site).
MONITOR:	Linear functions that look at state variables to determine when they reach certain values of interest to the modeler.
CLEARING:	The process of pre-emptively interrupting either on-going tasks or presently busy resources.
NETWORK:	The interconnection of IDEF$_0$ functions or SAINT nodes.

Table 4 presents an example of a representative PDB frame. The upper portion is a header for record keeping purposes. This is useful when several efforts may be ongoing or when a project may be of a size or duration where it becomes important to keep track of who filled in the information and when it was last reviewed.

TABLE 4. A REPRESENTATIVE PDB FRAME



FORM	CONSTRUCT AIRPLANE		
TITLE:	PERFORMANCE DATA BASE		
IDEF:	AO.2	AUTHOR:	SOFTECH
NODE:	4	DATE:	AUG. 84

FUNCTION DESCRIPTIONS

THIS ACTIVITY REPRESENTS THE CONSTRUCTION OF THE MODEL AIRPLANE BUT NONE OF THE DECORATING. THIS ACTIVITY IS DECOMPOSED TO A LOWER LEVEL OF DETAIL. BASED ON THE TYPE OF INSTRUCTIONS READ, THE BRANCHING WILL GO TO ONE OF THE THREE LOWER ACTIVITIES.

PERFORMANCE TIME:		MECHANISMS	MECHANISM COND.
DISTRIBUTION	---		
MEAN			
MINIMUM			
MAXIMUM			
STD DEV.:			

PREVIOUS COMPLETIONS REQS.	
PRIOR TASKS	
AO.1	

SUBSEQUENT BRANCHING

TASK NUMBER	BRANCHING LOGIC	MULT. BRANCHING COND
AO.1	IF MORE INSTRUCTIONS ARE TO BE READ.	
AO.3	IF CONSTRUCTION IS COMPLETE	

NOTES

A-D12530

There will be one PDB frame for each pair of an IDEF$_0$ function with a corresponding SAINT node. While this frame is the key data collection record, it is not the only record a SAINT modeler should be keeping. Several other tabular lists need to be generated in conjunction with the PDBs. However, these secondary lists simply keep track of what has already been done on prior PDBs. The modeler will therefore want to keep the following lists updated as new PDB frames are filled in:

- Distribution Sets: number and parameter value
- Moderator Functions: number and definition
- System Attributes: number and meaning
- Information Attributes: number and use
- Resources: number, nature, and how many attributes each has
- Resource Attributes: resource number, nature of the resource attribute, and the assigned resource attribute number
- State Variables: number and name
- Monitors: number and definition
- Task Characteristics: number and meaning
- User Functions: number and definition

Using the PDB as a template, the modeler has a guide for collecting the information that will be needed to represent dynamic task execution. One of the first questions addressed is what governs activity duration. A particular activity might be of fixed or variable duration. If it is variable, it may be determined as a function of some set of factors or may vary randomly according to some particular stochastic distribution. If randomly varying, then the modeler will specify the distribution type and its parameter values. If these are the same as for some other activity, then the modeler can simply refer to the existing distribution set number that was used before and annotate his PDB accordingly. If this task duration is a unique distribution type or uses a unique set of parameter values, then a new distribution set needs to be defined and added to the list of distribution sets that are being separately recorded.

Alternately, task timing may be a function of task characteristics, the present value of system attributes, the value of one or more information attributes, and/or the present value of one or more attributes of the resource that is required for task execution. If task duration is to be specified as being a function of such factors, then the timing is described differently. The modeler needs to specify the mathematical formula to be used to calculate the time value and the argument list (set of independent variables for that equation). This will then become a statement (or set of statements) in MODRF. The modeler will then assign a number to this particular function and add it to his list, or he will simply reference some function on the list if a suitable function was already defined for some previous PDB frame.

In C-SAINT, the modeler may establish a function library and reference those equations like distribution sets. This eliminates the need to program the equation in a FORTRAN subroutine (which was necessary in SAINT). However, if the C-SAINT function-set options are not adequate for the desired calculations, C-SAINT users can supply their own formulas, using the MODRF approach as currently done in SAINT.

Before or after timing considerations have been defined, it is also necessary to ask what governs when the activity will start. At least two factors control initiation. One is precedents and the other is resources. Precedents are simply activities which logically must have been completed before the present one is allowed to occur. A distinction may also be needed between the first time something is done versus all subsequent repetitions of that activity. For example, to start your automobile engine, you must first insert the key then turn it, but if the engine fails to start, you will need to turn the key again, but it doesn't need to be reinserted because that precedent was satisfied on the first attempt to start the engine. The second factor governing task initiation was availability of the requisite resources. Two situations routinely arise. The first is where several different resources are all required. The second case is where any of several resources may be suitable substitutes for one another. The PDB will specify both precedence and resource requirements for this particular node in the network, and if some new resource is implicated, it needs to be added to the list being kept.

It may happen that an existing resource is to be used, but in this case, the nature of that resource influences how long the task takes. Perhaps the modeler did not note this situation previously. Now the modeler may want to add a new attribute to the list of previously defined resource attributes and also go back and modify the definition of activity timing. In this fashion, the PDB allows for continuing review and updating as the modeling process evolves.

As a part of task accomplishment, various attributes may take on new values, reflecting the impact of starting or completing this particular activity. For example, a switch may be set to a new position, an error may occur, or some state variable may be regulated (changing its value). Or alternately, the performance of that activity may change a resource attribute (e.g., skill level) or some task characteristic (e.g., perceived difficulty). These changes are also noted on the PDB along with annotations about how the new attribute values will be assigned. They can be fixed or variable just as time values were. Attribute values that are to be changed according to a functional relationship require that the modeler specify the nature of the equation to be used and the argument list for that equation.

Finally, attention must be given to what happens after activity completion. Branching on exit from a node is along one or more paths to successor nodes and can be controlled three basic ways. So the PDB has slots for identifying what the successor (next) nodes will be and which set of rules will control exit path selection (deterministic, probabilistic, or conditional). If branching is conditional, the conditions must be stated and the form of evaluation must be identified. SAINT recognizes two forms of conditional branching evaluation: 1) take the first branch for which the stated condition is met, or 2) take all branches for which the conditions are met.

While these are the basic elements of a PDB frame, there are a variety of special cases that can be represented in a SAINT network. The PDB has a slot for notes to permit annotations that call attention to the need to exercise any of these specialized modeling features (e.g., task or resource clearing).

In the transition from IDEF$_0$ to SAINT, the dynamic behavioral modeling can be done at more than one level of detail. The IDEF$_0$ decomposition of static functions will provide a hierarchical partitioning. The SAINT model could easily be implemented at any particular level or strata in the hierarchy. However, it is also possible to build a model which operates at one level of detail for some functions and at another level of detail for other functions. Thus a single activity at one (more molar) level of detail may be replaced by a set of functions representing that activity at a more detailed (molecular) level.

As a standard rule, when a function at one level of detail is to be replaced by an activity network representing a more detailed level of breakdown, two dummy tasks are added to initiate and terminate the detailed SAINT network. These consume no time but control the branching and tunneling of information when going from one level of IDEF$_0$ to another. This procedure also simplifies control of looping and interactions among various modules. This convention of using dummy nodes assures a higher degree of modularity, simplifying the number of changes one must make if the model is changed later. The dummy tasks tend to insulate and isolate changes to localized areas while maintaining a tight coupling between the static IDEF$_0$ model of functions and the dynamic SAINT model of behavior.

Construction of the SAINT Performance Assessment Model of a SAM System (SPAMSS)

The next stage of model development re-examines the problem being addressed and lays out an architecture for various submodels that need to be constructed. From the IDEF$_0$ decomposition, it became apparent that there were interfaces where inputs or controls existed in the static IDEF$_0$ model that needed to be provided but were not incorporated in the PDB describing operator function execution (Bachert, Evers, Hoyland, and Rolek, 1983). For example, the threat aircraft state needs to be represented since it drives all other activities in the model. This was achieved by writing appropriate equations and incorporating them in subroutine STATE, which is linked with the SAINT code after it is successfully compiled. Correspondingly, there is a need for a tracking model and for a scoring model, the latter being incorporated in USERF which is also linked to SAINT after compilation.

Figure 4 shows the overall SPAMSS architecture as presented by Bachert, et al. (1983). Figure 5 portrays a sample threat aircraft flight path. Different flight paths can be represented by altering the module that contains the A/C Model SAINT STATE Equations. This can be done without affecting the contents of the other modules in the model. Four primary paths were incorporated with two to five variants of each. Path representation was

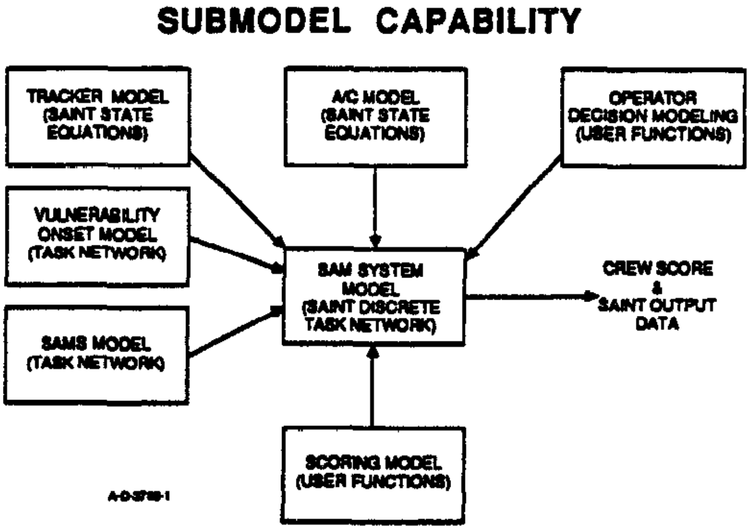


Figure 4. Overall SPAMSS Submodel Architecture

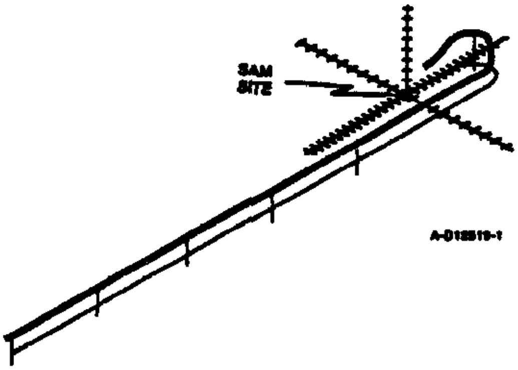


Figure 5. Aircraft Flight Path (Typical)

accomplished by using least squares regression to fit a polynomial to actual flight path measurements taken from the generic SAM simulator. The three factor polynomial equation was then implemented in SAINT subroutine STATE. The simulated flight paths were then correlated with the simulator by comparing the velocity and acceleration data with respect to the SAM site. This assured that the polynomial was a valid representation of the aircraft flight path.

The operator tracking model considered multi-task time sharing, as the tracking operators continually rotate the firing pedestal and also perform other associated duties. The pedestal was modeled as moving at a constant rate, using rate equations implemented in subroutine STATE. The operators regulated that rate (up or down), and this regulation task was treated as a discrete activity in the SAINT task network. When operators had no competing responsibilities, they could dedicate themselves to evaluating tracking error and adjusting pedestal rate accordingly, which produces a smaller tracking error. By contrast, as diversions increase, the rates are not changed as often and must be of greater magnitude, so tracking errors increase.

The scoring model provides a record of how well the operators performed under the various conditions studied. It keeps track of the elevation and azimuth tracking errors and the range from the SAM site to the aircraft. These scores are then used to determine when the aircraft is vulnerable to a missile strike. The Fire Control Operator can then be scored on how long it takes to make a firing decision relative to the first possible choice, the optimal (maximum hit probability), and last feasible firing solution. If a decision was made to launch, the outputs from the scoring model are used as inputs to a missile fly-out model. That model will then evaluate the estimated miss distance, probability of hit, and probability of kill for this particular engagement. Consequently, both behavioral and system measures are available.

VALIDATION AND UTILIZATION

Because the SAINT model was based on an instrumented simulator, there was a considerable body of data available from prior experiments that could be used to estimate parameter values for the SAINT model. Moreover, the simulator was available to validate model predictions subsequent to model implementation.

It was possible then to validate the SAINT predictions in terms of several measures and for more than one set of operating conditions. Figure 6 shows some sample results from SPAMSS (Hoyland, Evers, and Snyder, 1985). Along the top of this graph are indications when particular mode changes occurred. SAINT predictions were not significantly different from those found in the empirical studies. The continuous tracking error is also plotted over time, and the windows of vulnerability shown are where the errors were small enough to permit successful missile launch. Again, SAINT provided valid predictions of error and start/stop times on the windows of vulnerability.

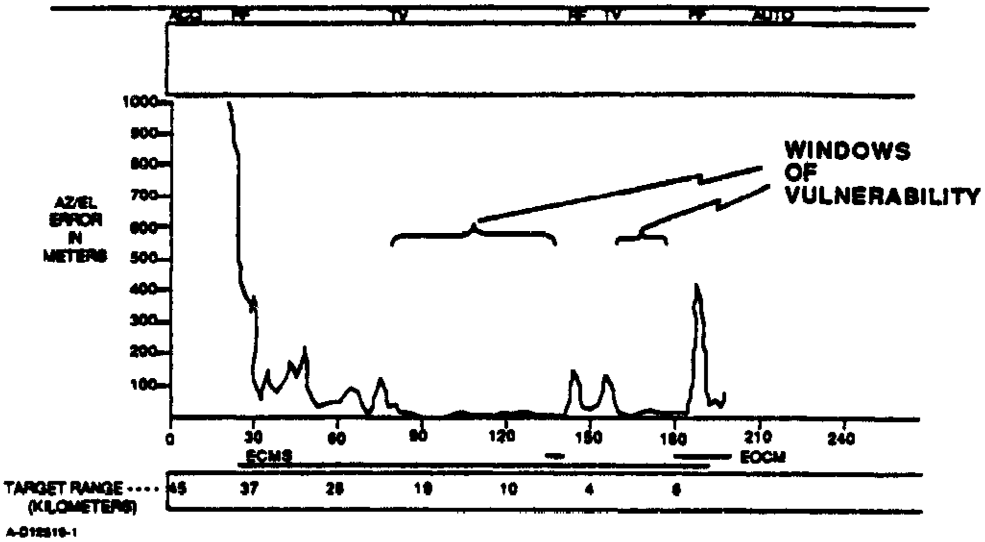


Figure 6. Model Results

Preliminary Utilization

Subsequent tests looked at changes in the use of countermeasures that impact the use of automatic tracking and the use of various sensors for manual tracking. In this case, the SAINT model was first used to estimate what should be observed when the empirical tests were run. Following the empirical studies, actual operator and system performance was compared with the SAINT predictions. Again, no statistically significant differences were found.

Having developed a valid representation of the existing simulator based on data from prior studies, it was then possible to use the model two ways. First, one could try out new studies before running human subjects. Second, one could examine conditions that could not be tested in the simulator. Both uses were pursued.

The first set of studies is especially useful in systems analysis and empirical research planning. One can examine the hypothetical implications of changes in system design or in operating procedures. Alternatives can be quickly examined to see which are worth validating by empirical test. Some concepts may perform so poorly there is little reason to spend money on real-time, human-operator studies. In other cases, the model may predict results that appear more promising and clearly seem to warrant empirical test to validate those predictions.

The second set of studies is more controversial. Here we speculate on conditions that cannot (or at least will not) be empirically tested. For example, it is of some interest to estimate how well operators might perform under actual combat conditions where they may be exposed to various weapons effects. Because those weapons effects cannot be safely administered at large exposure levels, modeling provides a means of estimating what could occur under conditions that would be unsafe for empirical study. A general methodology was developed for treating this problem, one that could be used for a variety of environmental effects (Evers, Hoyland, and Hann, 1986).

Speculative Modeling

Since no one can ethically validate by empirical test a model that tries to predict the performance impacts of exposure to debilitating or lethal weapons effects, some prefer to treat such predictions as speculations rather than as estimates. This use of terms calls attention to the inherent underlying uncertainties of predictions that extrapolate from what is known to what cannot be empirically validated.

Ionizing radiation and chemical/bacteriological weapons effects typically induce varying degrees of performance degradation before death occurs. Also, the prophylactic drug regimens and antidotes for chemical weapons are themselves suspect as degrading influences. If military personnel can and will continue to perform their duties under such conditions, then design evaluations should consider such factors in assessing performance adequacy for those combat conditions. What is needed is a methodology for speculating about the magnitude of the performance decrement, given some absorbed dose of drug, weapon effect, ionizing radiation, etc.

To make such computations practical, a two-step process was devised. The Criterion Task Set (CTS) was proposed by the Armstrong Aerospace Medical Research Laboratory (AAMRL) as a suitable basis for describing human abilities. The first step in the process was to estimate (by expert judgment) how much each CTS factor contributed to the nominal (baseline or no exposure activity condition) task duration parameters (average and standard deviation of task time). The second step in the process was to speculate how much each of these CTS factors degraded under the dose level being evaluated.

A regression equation was fit to the expert judgments such that nominal task durations could be estimated as a function of the CTS weighting provided by subject matter experts (SMEs). Then the CTS factors could be adjusted as a function of the speculated impact of the imposed dose level. The regression equation would then be used to predict the inflation in task duration as a function of the dose induced impairment (as reflected in degraded CTS abilities, per speculations by SMEs).

Subsequent attempts to validate this methodology empirically were inconclusive due to the conservative nature of the treatment conditions used. Only non-significant performance decrements were observed in the criterion data that were to be used to validate model predictions. Consequently, the methods developed are basically sound but arbitrary and can only be used to speculate about possible consequences, not predict actual behavior.

OTHER OBSERVATIONS

The SPAMSS use of the IDEAL concept was done manually. Using Business Filevision, SofTech has mechanized portions of the IDEAL concept to provide a Computer Aided Engineering (CAE) tool for our own use. The conversion of the performance data base attribute values into SAINT or C-SAINT input data records has not been accomplished. Methods are available for doing so.

REFERENCES

1. Bachert, R.F., K.H. Evers, C.M. Hoyland, and E.P. Rolek. "IDEFo/ SAINT SAM Simulation: Hardware/Human Submodels." Proceedings of the IEEE 1983 National Aerospace and Electronics Conference (NAECON). Dayton, OH. May, 1983. pp 1066-1070.

2. Bachert, R.F., K.H. Evers, C.M. Hoyland, and E.P. Rolek. "Static and Dynamic Model Integration for a SAM C^3 Simulation via IDEF$_0$ / SAINT." Proceedings of 1982 IEEE International Conference on Cybernetics and Society. Seattle, WA. 1982. pp 144-148.

3. Bachert, R.F., K.H. Evers, and P.R. Santucci. "SADT/SAINT Simulation Technique." Proceeding of the IEEE 1981 National Aerospace and Electronics Conference (NAECON). Dayton, OH. 1981. pp 4-9.

4. Chubb, G.P. "Human Factors in Systems Engineering: New Analysis Requirements." Proceedings International Topical Meeting on Advances in Human Factors in Nuclear Power Systems. Oak Ridge National Laboratories, TN. 20-23 April, 1986.

5. Chubb, G.P. "SAINT, A Digital Simulation Language for the Study of Manned Systems." In J. Moraal and K.F. Kraiss (eds.), Manned System Design: Methods, Equipment, and Applications. Plenum Press, NY. 1981.

6. Chubb, G.P, and C.M. Hoyland. "Systematic Modeling of Multioperator Systems to Evaluate Design Concepts." Proceedings of the NATO Workshop on Human Operator Modeling, G. Mcmillan (et al.) editors. Plenum Press, NY. 1988 (in press).

7. Chubb, G.P., N. Stodolski, W.C. Fleming, and J.A. Hassoun. "STALL: A Simple Model for Workload Analysis in Early System Development." Proceedings of the 31st Annual Human Factors Society Meeting. NY. 1987. pp 363-367.

8. Evers, K.H., C.M. Hoyland, and R.L. Hann. "Measuring and Integrating Environmental Effects into Preliminary Crew Station Design." Proceedings of the IEEE 1986 National Aerospace and Electronics Conference (NAECON). Dayton, OH. May, 1986. pp 1350-1354.

9. Hoyland, C.M., D. Ganote, and G.P. Chubb. "C-SAINT: A Simulation Modeling Tool Customized for Workload Analysis." Proceedings of the IEEE 1988 National Aerospace and Electronics Conference (NAECON). Dayton, OH. May, 1988. pp 823-830.

10. Hoyland, C.M., G.P. Chubb, and K.H. Evers. Software User's Manual for the C-SAINT Upgrade, Final Report. SofTech, Fairborn, OH. July, 1988.

11. Hoyland, C.M., K.H. Evers, and D.E. Snyder. "Incorporating Human Operator Considerations into Existing Weapon System Analysis and Quantification Capabilities." Proceedings of the IEEE 1985 National Aerospace and Electronics Conference (NAECON). Dayton, OH. May, 1985. pp 911-916.

12. Marca, D.A., and C.L. McGowan. SADT: Structured Analysis and Design Technique. McGraw-Hill, NY. 1987.

13. Pritsker, A.A.B. Introduction to Simulation and SLAM. 3rd edition. John Wiley and Sons, New York. 1986.

14. Ross, D.T. "Applications and Extensions of SADT." IEEE Computer. April, 1985. pp 25-34.

15. Seifert, D.J. and G.P. Chubb. SAINT: A Combined Simulation Language for Modeling Large, Complex Systems. AMRL-TR-78-48. Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH. October, 1978.

16. Wallace, R.H, J.H. Stockenberg, and R.N. Charette. A Unified Methodology for Systems Development. McGraw-Hill Book Company. 1987.

17. Wortman, D.B., S.D. Duket, D.J. Seifert, R.L. Ham, and G.P. Chubb. The SAINT User's Manual, AMRL-TR-7762. Aerospace Medical Resarch Laboratory, Wright-Patterson Air Force Base, OH. June 1978.

DISCUSSION OF "SAINT PERFORMANCE ASSESSMENT MODEL OF A SURFACE TO AIR MISSILE SYSTEM"

by G. Chubb and C. Hoyland

DISCUSSANT: Grant R. McMillan, Aerospace Medical Research Laboratory,
Wright Patterson Air Force Base, Ohio

The Chubb and Hoyland paper provides a good example of how human performance models might be used to generate data for high resolution combat simulations. As such, this paper could have been included in Session II on Human Performance Models and Applications. Their approach to modelling crew performance under degraded conditions should be of special interest to the combat modelling community. Although only briefly discussed in the paper, the approach uses a battery of basic perceptual, cognitive, and motor tasks to provide estimates of the combat stress effects on specific human abilities. The results of these tests are then used to modify the task performance data in the SAINT network. Running the SAINT model provides predictions ("speculations", to quote the authors) of stressor effects on the overall performance of the crew-served weapon system. Although this approach makes significant, unvalidated assumptions about our ability to generalize from individual task data to crew performance, it is one of the only means available to predict combat stressor effects on weapon system effectiveness. This approach should be especially useful in sensitivity analyses, where one is attempting to identify critical human factors variables that impact combat outcome. In such cases, one is more interested in relative rather than absolute effects.

In my opinion, there are at least three important challenges in network modelling:

(1) Determining the functional structure of the man-machine system.
(2) Constructing the network model of the decomposed system.
(3) Obtaining and managing the performance data required to exercise the network model.

In their recent work, Chubb and Hoyland have attempted to systematize the approach to challenges one and three. They show how the use of IDEF$_0$ techniques can reduce costs and eliminate errors when decomposing systems. These benefits are particularly noticeable with large systems that require the work of several analysts. Although they acknowledge that available data base management tools could further enhance their system, the current Performance Data Base templates (frames) are an important initial step for collecting and managing the data required to execute the network model.

There was some informal discussion at MORIMOC II on the possibility of using network modelling tools to develop new combat simulations. Whether or not this is a good idea is beyond my expertise; but if this avenue is pursued, some of the supporting tools described in this paper will be critically necessary. This paper should be carefully read by anyone contemplating a network approach to combat modelling.

Military Operations Research Society (MORS) Conference
February 22-24, 1989

JAMCSI (JOINT AGENCY MEETING ON COMBAT SIMULATION ISSUES) CONFERENCE SUMMARY

William W. Banks
Lawrence Livermore National Laboratory
Livermore, California

Thomas Berghage
RMC International
San Diego, California

The joint Navy/Lawrence Livermore meeting (JAMCSI) was held at Lawrence Livermore National Laboratory on November 30th and December 1st 1988. The meeting was intended to bring together individuals from two research communities: the combat simulation and war gaming community, and the human performance community.

It was intended that the meeting provide a forum for the exchange of ideas and concepts with the ultimate objective of improving the fidelity of combat simulation modeling.

As background for the meeting two publications were given out to each attendee:

"Review and analysis of the literature in the area of human performance modeling."UCID 21558,LLNL,November,1988.

"An inventory of wargaming models for special warfare: Candidate Applications for the Infusion of Human Performance Data."UCID 21551,Lawrence Livermore National Lab.,November, 1988.

The first day of the meeting was devoted to presentations on existing combat simulation models and their attempts to incorporate human performance information. Speakers provided an overview of some of the more widely used models and outlined the rational for the model's development, its current use, and the strategy being used to incorporate human performance information.

The list of topics and speakers were as follows:

Human Performance Research at the Naval Health Research	Captain Chaney NHRC Center
Overview on Navy Modelling Needs	Captain Jones NMRDC
Meeting Objectives and Organization	LCDR Kelleher NHRC
JANUS Model	Dr. Toms LLNL
Crew III	Mr. Anno Pacific-Sierra-Eaton
SEES Model	Dr. Terhune LLNL
TWSEAS Model	Maj. Anderson USMCB Pendelton
AURA Model	Dr. Kolpcic ABRL
NURA Model	Dr. Yencha & Dr. Kirk NSWC
Micro SAINT	Dr. Laughery Micro Analysis & Design
SHIPDAM	Mr. Hawkins DTRC
Human Factors Modelling Requirements	CDR Contreras ONT

The evening dinner speaker was Col. John Pickering (USAF Ret.) and he provided a historical perspective to the development of the military's interest in combat simulation modelling.

The second day of the meeting was to be devoted to future combat simulation models and work groups were requested to deal with design and specification issues. The original agenda had to be abandoned when it became apparent that there were several issues that needed to be addressed by the entire group. Before the meeting was opened to general discussion Commander Tom Contreras from the Office of Navy

Technology gave a presentation on the organization and structure of the research administrative environment and the hurdles that faced any new research initiative.

The open discussion was completed with a short presentation by Dr. Earl Alluisi from the Office of the Under Secretary of Defense. Dr. Alluisi indicated his strong support for computer modelling and the use of combat simulation for evaluating various defense alternatives. He said that he was supporting a major research thrust in this area because of the wide spread potential for this technology.

To bring you up-to-date on the Navy/LLNL meeting, I would like to summarize some of the major points made by the presenters at the meeting.

Captain Chaney outlined the numerous human performance programs currently underway at NHRC and indicated that these research programs, along with the individual researchers involved, were available to support the combat simulation effort. He felt that the human performance research that is going on in the Navy medical laboratories was an untapped resource and, that if appropriately applied, could enhance the fidelity of combat simulation models.

Captain Jones directed most of his comments to human performance databases. He indicated that several Navy laboratories were using the MICRO SAINT software product to organize and develop human performance information. He felt that the human performance modelling development in the past had been hampered by both hardware constraints and the lack of user friendly software, but that both of those barriers had now been overcome and that it was now time to develop the supporting data modelers will have to deal with:

1. The goal of the model.

   Modelers need to identify the users and bring them into the development process early.

2. How good is the model?

   The model has to make things easier for the user and the user has to be able to believe in the results. To do this we must make sure he knows what assumptions went into the model and he must know that the results are valid.

3. What databases are available?

   We need a major effort to bring together the various fragmented databases and make them generalizable for use in the various models.

4. How do you get fleet support?

Captain Jones suggested that there are four elements in getting fleet support. They are:

a. Existence of a valid requirement.
b. An identified user.
c. Integration within existing technology base.
d. A transition plan for moving the technology along.

LCDR KELLEHER

In outlining the objectives of the meeting LCDR Kelleher indicated that one of the main purposes of the meeting was to bring together the various organizations and agencies that have been working independently on computer simulation models and develop a dialogue among researchers. This objective was certainly obtained.

LCDR Kelleher made the point that he considered himself a user for the modelling effort. He felt that combat simulation models are very much a research tool in addition to their other uses. He suggested that combat models could be used to guide and structure research efforts in the future. He also questioned whether or not we had fully utilized the data that is currently available. Maybe we need to have a major effort to organize and make available the existing human performance data before we go out and collect more data.

For detailed information regarding the presentations on the individual models you should pick up a copy of the meeting proceedings. We will, however, summarize some of the general comments that were made by the speakers.

Dr. Toms talked about Janus and suggested four essential items for a good model.

1. Openness - full disclosure of the models structure and content - good documentation.

2. Usefulness - The model should be used by individuals other than the developer. It should not be a clever laboratory game.

3. Limitations - The limitations of the model should be clearly spelled out and made available to the users.

4. Validation - The model should reflect what really goes on in the real world.

Dr. Toms indicated that Janus had been changed over to a distributed data processing architecture to enhance processing speed and allow the running of the model in remote locations. Dr. Toms reported that Janus included some basic human performance information, but that he considered the lack of this type of information one of the biggest short-falls in the model.

Mr. Anno described the multidisciplinary development of Crew III and how it was a continuation of the IDP work done for the Defense Nuclear Agency. This program like many others has used symptomology descriptions to tie stress variables to performance.

The Crew III model is now being used in Janus to handle some of the human performance information.

Dr. Terhune talked about the SEES model which is a modified version of Janus that was developed for the Office of Security Evaluations. It is unique in its ability to handle combat simulation in an urban type environment. Its primary purpose is to model the problem of armed intrusion against a secure site. Dr. Terhune felt one of the main reasons for their success has been the close working relationship they have had with the security guard users. They have been intimately involved from the start.

Major Anderson is a user of computer models. He manages the Tactical Warfare Simulation (TWSEAS) unit at Camp Pendleton. His system, like the others presented, includes limited human performance information. TWSEAS is used for staff training and as such needs to be as realistic as possible. Maj. Anderson stressed, however, that the human performance inputs did not have to be perfect, an approximation of the human element would be better than what he currently has.

Dr. Klopcic described the AURA (Army Resiliency Analysis) model. Unlike the other models, AURA is a one sided model that looks at the functioning of a unit over time including times following hostile attack. AURA is designed to be a framework into which existing models can be incorporated. Dr. Klopcic referred to AURA as a methodology rather than a model. One point that Dr. Klopcic made that needs to be emphasized is that by not considering a given variable in a model does not mean that you have not included its effect. It just means you have either consciously or unconsciously assigned it a value of one.

Dr. Yencha and Dr. Kirk spoke on the NURA model which is the Navy's version of AURA. The model is primarily designed for assessing the impact of chemical attacks on naval vessels. They are looking for good human performance information for their model, more specifically, they are looking for information on the affects of MOPP gear (chemical warfare protective clothing) on performance.

Dr. Laughery devoted most of his time to "task network modelling" which is the structural technique used in the MICRO SAINT software. Dr. Laughery feels that this software can act as the bridge between human performance modelling and the combat simulation model. MICRO SAINT is a commercial product that was developed under government contract. It has been used extensively for modelling human performance, but it is general enough to be used to build any network simulation. The developers of MICRO SAINT think that it will do for modelling what the spread sheet programs did for financial analysis. Its user friendly nature will eliminate the need for a modelling specialist and bring modelling capability down to the user level.

Mr. Hawkins described the SHIPDAM model which is a modified version of the ship vulnerability model that has been developed at David Taylor Research Center over the last 15 to 20 years. It is a Monte Carlo model designed to handle probabilistic events. The model is not yet complete or documented, but is being used for several projects.

Commander Contreras spent considerable time detailing the R&D system and outlining the research project review process. He indicated that knowledge of this system was important for getting new initiatives such as combat simulation and human performance modeling funded. He emphasi the importance of identifying the user community early on in the development so that a transition plan can be put into place. He closed his presentation by listing the four criteria he uses in evaluating new programs. They are:

1. A documented need or requirement and an indication that the researcher has taken the time to learn about and understand the problem.

2. A well laid-out plan with achievable goals.

3. An indication that the researcher is aware of and using all available resources; both equipment and information resources. Not just in his organization, but throughout the R&D community.

4. An understanding and support of the R&D system. For a project to develop smoothly the researcher has to know the steps involved in the R&D System.

Finally Cmdr. Contreras suggested that researchers working on computer systems that potentially could be used in operational medicine be aware of DAMSEA the Defense Medical Systems Support Center. It is the DOD organization that oversees the implementation of new computer systems. He was not sure what they were doing in the modeling arena, but suggested that they be contacted.

In the general discussion session Dr. Alluisi suggested that the human performance modeling had to be focused on those tasks that make a difference in the outcome of combat engagements. He suggested that we should conduct some sensitivity studies using the combat simulation models to determine what tasks we need to model. He indicated that this information was going to be extremely important in the evaluation of weapon systems and that the science of combat simulation was going to take on increased importance in the future because it is one of the only ways you can systematically evaluate the importance of various system components.

CONCLUSIONS AND RECOMMENDATIONS

1. We really don't have a good idea of what human performance data is available for modelers and/or how useful it will be.
2. We need a list of problem areas from the modelers to help focus the human performance research effort.

3. There seem to be two points of view regarding how we should attack the modelling problem: one suggests a bottom up approach while the other feels an top down approach is more appropriate.

   Bottom Up - Each lab has its own particular needs and as models are built to meet these needs, they can be used as building blocks to build larger combat simulation models.

   Top Down - If we wait for the development of all of these individual models we will never get to the overall model that will meet user needs. Lets get a rough cut model up and running and use it to identify information that is needed. Let the large overall model drive the research effort.

4. Hardware no longer appears to be a constraint. By using distributed data processing along with 32 bit intelligent terminals modelers seem to be able to do just about everything they currently want to do. The choke point now appears to be the quality and quantity of human performance data available.

5. There appears to be a need to bring in funding from a number of different sources rather than relying solely on the Medical R&D Command.

6. It was suggested that we use a matrix of independent and dependent variables to relate the impact of various stressors to various performance variables. The relationships could be developed from the scientific literature and specifically designed studies. There seems to be some concern as to whether such a matrix could be translated into combat performance.

7. There appears to be a need for some sort of clearing-house for the exchange of information regarding the modelling effort. It was suggested that an electronic network be set up that includes bulletin boards and electronic mail for the dissemination of information and ideas.

8. It was suggested that a directory of modelers be developed and that it include the electronic addresses for those on the ARPA Net.

9. Several investigators felt that we need to confine, at least initially, the human performance modelling effort to the small combat unit rather than trying to introduce human performance data at the Division or Brigade level.

10. There appears to be a need to get some type of human performance information in the existing models right now. We can refine the development after we get something out there operating.

11. A publication needs to be developed that reviews the human performance literature and evaluates the models that are currently available. Something similar to what was done by LLNL for the combat simulation models.

RECOMMENDATIONS / SUGGESTIONS

1. Within the US Navy, a combat simulation integration review function is needed to provide specifications and guidance regarding the existing Navy modelling efforts. This function is needed so that greater utility for existing models can be generated. It would also allow for greater integration of existing models. Currently, there are many different machines and programs running models which cannot be easily joined together if needed. A Modelling oversight group could insure that new models conform to criteria and standards which will allow them to be integrated in the future or at least be "modular and transportable" to other systems/models.

2. A Handbook for Combat Simulation Model Development from a multidisciplinary point of view could be very valuable in banding together professionals from the OR, Physical Sciences, Behavioral Sciences, and Engineering Communities. Lawrence Livermore National Laboratory and BDM have developed a straw-man table of contents and a suggested author list of over 30 individuals to contribute to this effort if sufficient funding from the joint services is obtained.

Work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract W-7405-Eng-48

SESSION III: PREDICTING HUMAN PERFORMANCE/ AVAILABILITY IN COMBAT ENVIRONMENTS

SESSION CHAIR: Sally J. Van Nostrand
US Army Concepts Analysis Agency

I think that Session III is the most exciting session in MORIMOC II - not because it is the longest, but because of the papers in it. I'm delighted that this is the session that I'm chairing. Since one of the goals of MORIMOC II is to bring back a multidisciplinary orientation to operations research, we've specifically invited (and have present) people from diverse fields that we have not previously had at MORS meetings. As you look at the education and experience of the group of authors and discussants in this session, you'll see that about all that they do have in common is an interest in human performance and how that performance can be represented in combat models.

The morning papers. The papers this morning center on predicting soldier performance. Most combat models now assume that soldiers are similar to binary computer bits; they're totally on (perform at 100 percent proficiency) or totally off (when they become a casualty and are totally removed, at least temporarily, from the system). These papers, from the human research community, show the results of research projects that were specifically designed to measure combat performance when it is less than 100 percent, but not necessarily zero percent. If we are to add the human dimension to combat models, we must include the factors that influence human performance, and the measures in the model that represent human performance must be more than binary on or off functions.

The morning discussants. One of the main reasons given for why soldier performance has not been included in combat models is that modelers tend to say, "There isn't any data," while the human researcher says, "They won't use the data I give them." I believe that there has been a communication problem--the two sets of people don't speak quite the same language. I've chosen the discussants for the morning session to test that theory, and if the problem exists, to see if we can help find a solution. We'll have all of the papers first. Then the first discussant is an experienced, distinguished combat modeler who will tell us whether he thinks these research results are immediately usable in combat models, and if not, why not. He will be followed by an equally distinguished and experienced researcher in military psychology who will provide either a rebuttal for the whole group or further critique, whichever he feels is appropriate at that point.

The afternoon section. The papers this afternoon are application or modeling oriented. Most of the papers seem to have a foot in both human research and operations research. In order to continue to test whether human researchers and operations researchers are understanding each other, I've tried to choose discussants who know something about the area addressed by the paper, but who are not from the same discipline as the authors. I've asked all the discussants to address the question of whether the data or technique presented is directly usable in today's combat modeling. Where they say, "No," I've asked them to tell us what they feel is needed to make it usable.

EFFECTS OF IONIZING RADIATION ON THE PERFORMANCE OF SELECTED TACTICAL COMBAT CREWS

George H. Anno and Michael A. Dore
Pacific-Sierra Research Corporation
12340 Santa Monica Boulevard
Los Angeles, California 90025

ABSTRACT

A model is developed to assess the expected performance of four selected types of tactical army combat crews when the individual crewmembers function at degraded performance levels due to acute exposure to ionizing radiation. The model is also applicable to other situations that degrade individual crewmember performance. The results provide performance data for larger scale U.S. Army models that simulate battlefield conflicts where nuclear weapons might be employed. Performance-level data are generated as a function of dose and time after exposure for each crew type.

INTRODUCTION

This article reports on a continuing effort of study to quantify the acute effects of intermediate level gamma and neutron doses (0.75 Gy to 45 Gy free-in-air) on the combat effectiveness of tactical combat forces. Our effort builds on prior work of the Defense Nuclear Agency (DNA) Intermediate Dose Program (IDP), where the expected performance level of individual soldiers was determined as a function of dose and time after exposure [Anno, Wilson, and Dore, 1983]. Here we extend the prior IDP work to small army combat crew units to assess combined crew performance. Four types of crews selected for this study are shown in Table 1, along with a brief description of the specific combat engagement action of each. These crews are all associated with Army combat vehicles representing a significant portion of the tactical ground force.

The methodology followed in developing combat crew performance levels is outlined in Fig. 1. There were two specific objectives of this study. The first was to determine a relationship for the collective performance based upon the performance levels of each of the crewmembers. For a typical crew of four crewmembers, the relationship, indicated in Fig. 1 under "Four-dimensional functional fits," required eight parameters, two each for four independent variables (independent crewmember performances). This relationship can be used in combat simulation codes such as Ballistic Research Laboratory's (BRL) AURA [Klopcic and Roach, 1984] to predict battlefield performance of a crew for an arbitrary mix of degraded crewmembers. It applies regardless of the underlying cause of degraded performance, whether due to radiation sickness or some other physical or psychological stressor.

Utilizing the above crew performance algorithm, the second objective was to determine crew performance as a function of acute dose

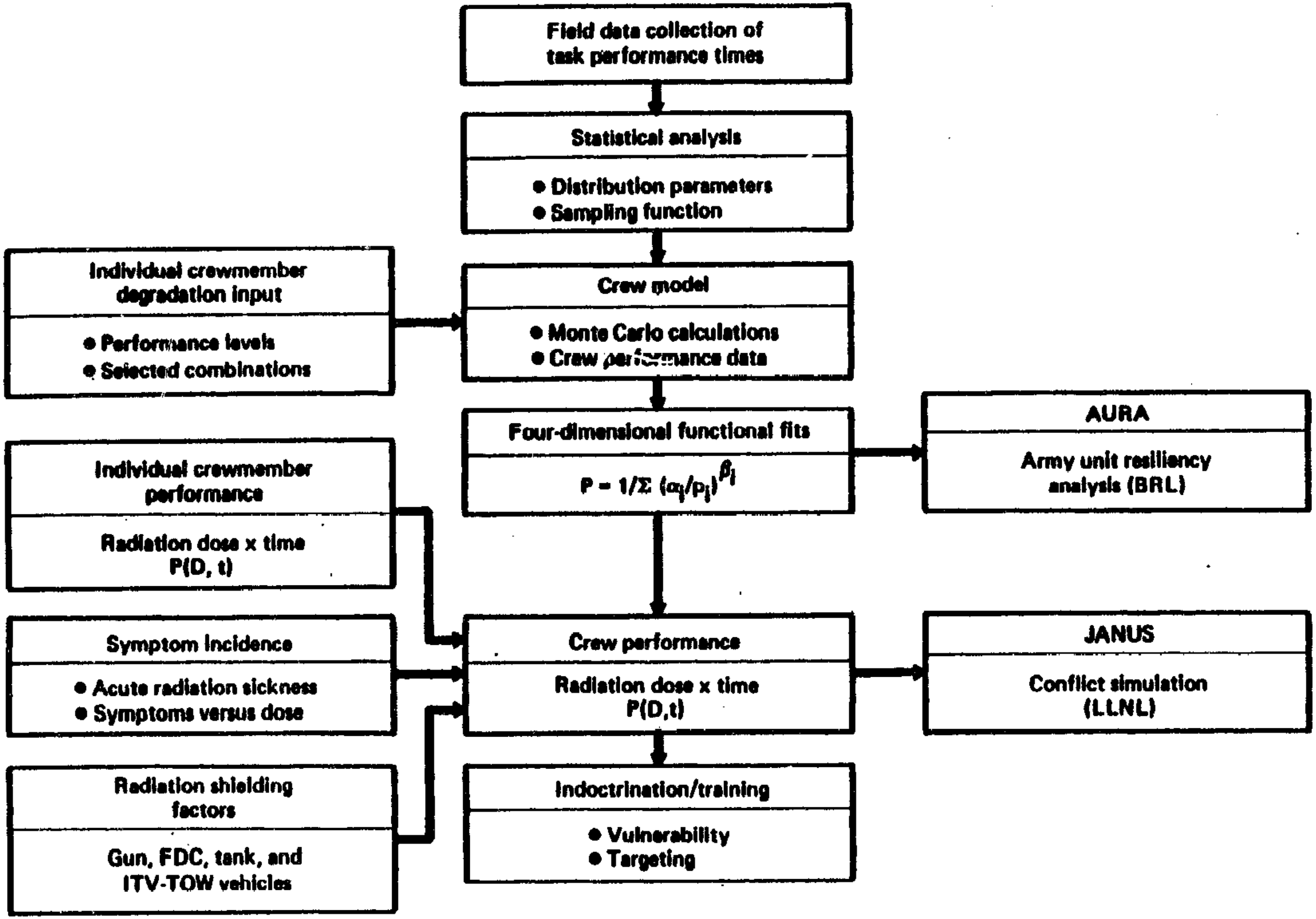


Figure 1. Combat crew performance--methodology.

Table 1. Selected Army combat crews.

Crew Type	No. of Tasks	Crewmembers	No. of Tasks	Brief Description of Engagement Actions	Normal Time (s)
M109 howitzer Gun crew	31[a]	1 Chief of section	8	Set-up, aim, load, and fire 2 rounds.	103.2
		2 Gunner	3		
		3 Assistant gunner	2		
		4 No. 1 cannoneer	18		
Fire direction center (FDC)	26	1 Fire direction officer	8	Acquire, calculate, and transmit target aiming data to battery.	93.4
		2 Computer	3		
		3 Horiz. control operator	7		
		- (observer & battery)[b]	(8)		
M60A3 tank crew	49[a]	1 Tank commander	19	Site target, emerge from cover, fire, reload 2 rounds, and resume cover.	28.1
		2 Gunner	18		
		3 Loader	6		
		4 Driver	4		
M901 ITV-TOW crew	24	1 Squad leader	7	Move to site, sequentially fire and guide both missiles and reload.	157.8
		2 Gunner	8		
		3 Driver[c]	2		
		4 Loader	7		

[a] Includes 2 recoils.
[b] Fixed times.
[c] Inferred--no data.

external to the crew vehicle and time after exposure. The methodology depicted in the lower portion of Fig. 1, combines a number of factors to yield the desired crew unit performance. Among these factors are: radiation shielding by the vehicle, the variability among individuals in expressing symptoms of radiation sickness, the crew performance algorithm, and results from the IDP studies of dose and time-dependent individual crewmember performance.

Dose and time-dependent performance data were generated for each crew type assuming typical vehicle shielding factors and averaging over the distribution of symptom incidence. These performance data, illustrated as a set of surfaces, are suitable for implementation on higher level combat simulation codes, such as the JANUS code at Lawrence Livermore National Laboratory (LLNL) [Blumenthal et al., 1984]^3.

Performance is expressed as the ratio of the time normally required to complete the specified combat engagement with crewmembers unaffected by radiation (the normal mission time) to the time required when the performance level of one or more crewmembers is degraded (i.e., slower) from the effects of radiation. This definition was adopted in the earlier IDP study of individual crewmembers [Anno, Wilson, and Dore, 1983][1] and has the property of yielding a value between zero and one as well as being amenable to empirical field exercise measurements.

A combat engagement includes a number of tasks shared among the individual crewmembers that must be completed according to a predetermined sequence. The performance of each individual crewmember is expressed quantitatively according to the time taken to complete assigned tasks. The simulated combat engagements are all very brief, ranging from about 30 s to 1.5 min, and represent a critical period of combat where crewmembers are required to execute a large number of often split-second tasks in a coordinated team effort. Accordingly, such intense combat engagements represent events crucial to the outcome of tactical battlefield encounters.

Below we discuss the methodology (summarized in Fig. 1) developed for determining degraded performance levels for each of the four combat crew units. The appendix contains tables and diagrams referred to in the discussion that provide some detailed information and data.

ANALYSIS OF EMPIRICAL DATA

We have utilized data assembled from a large number of field measurements of the time required by trained military personnel to perform the tasks comprising the engagements described in Table 1. The M60A3 tank data [Moyer and Lam, 1987][4] were combined with data for three other crew types: the M109 howitzer, fire direction center (FDC), and M901 Improved TOW (tube-launched, optically tracked, wire-guided) Vehicle (ITV-TOW) crews [Moyer, O'Donoghue, and Feinberg, 1984][5]. These data were taken from multiple repetitions of standard operational engagements by trained soldiers. A brief description of each engagement is included in Table 1. More detailed descriptions of the specific tasks are contained in a DNA report [Dore and Anno, 1988][6].

The goal was to generate an accurate, dynamic simulation model for each crew type. The empirical data gathered were suitable for statistical evaluation to provide the following critical inputs to such a model:

1. Representative mean times for each task
2. Variances about the mean times for crewmembers
3. Detailed chronological sequence and hierarchy of task executions to allow construction of task interaction diagrams

The scope and extent of the measurements for all four crew types is summarized in Table 2. Most tasks for most crew types were performed five or more times by several different crews comprised of different individuals. However, because only a limited number of

Table 2. Structure of empirical task-time measurements.

Crew Type		No. of Tasks	No. of Separate Crew Units	No. of Separate Individuals	No. of Trials Each
	Gun crew	2	--	1	55[a]
		25	11	33	5
	FDC	23	10	30	5
Tank crew	COFT	18	15	30	7 to 12
	Live	4	10	10	1
		2	4	4	10
	ITV-TOW crew	7	--	2	20[a] & 30[a]
		12	5	10	10[a]

[a] Only the last 5 trials were used.

qualified personnel were available for certain positions, they were required to occupy those positions for all of the trials. For example, for the gun crew a single individual (chief of section) participated in each of the 11 separate crew units. This crewmember performed two of the same tasks five times each for a total of 55 trials. Also, because certain tasks involved the actual firing of expensive, live ordinance, those tasks were not repeated as often. In Table 2, two rows of numbers are used for all crew types but the FDC. The upper numbers represent situations in which the number of qualified personnel was limited.

The entry in Table 2 for the M60A3 tank crew is divided in two because task time measurements of the tank crewmembers were made in two separate environments. The many rapidly occurring tasks for the tank commanders and gunners required that timings be taken in a standard tank simulator, the "conduct of fire trainer" (COFT). This simulator is stationary and does not fire live rounds; therefore, the loader and driver tasks were possibly not realistic representations of an actual tank environment in some respects. However, a limited number of loader and driver task timings were taken in actual tanks participating in exercise maneuvers. These exercises included live firing of the gun. The two subdivisions given for the tank crews in Table 2 delineate the tasks and number of trials applicable to each set of measurements.

The final mean times and variances for the empirical data, except for the tank crew, result from five independent trials by each individual. Where more trials were available, only the last five were used. Occasionally, an individual would have only four valid trials; in these cases, the standard statistical practice was employed of introducing an artificial fifth trial equal to the mean of the four valid trials. For the tank crews, however, all available data were used.

Of concern was the possibility of a "learning curve" effect in which a trend towards shorter times would develop with successive trials by an individual. Correlation tests were performed to determine if any systematic trend could be found progressing from the first to the fifth trial. Neither a trend toward longer times (due to boredom with repetitions of the same exercise) nor toward shorter times (learning curve) was discernible.

The complete lack of any appropriate data on crewmembers in degraded states (due to radiation injury or any other applicable insult) was one of our concerns. In particular, a method was needed for estimating the variance about the mean task times for degraded crewmembers. Given this deficiency, our approach was to model the variance as a function of the mean, assuming the same functional form to hold for both the degraded and undegraded situations.

Let T_i be the normal (undegraded) mean time for completing a given task i. Then the performance of a crewmember for the task is defined as $p_i = T_i/t_i$, where t_i is the longer time required to accomplish the same task when the crewmember is in a degraded state. So for a given performance level p_i, the mean time for a degraded crewmember to perform task i is just $\mu_i = T_i/p_i$. A model for the variance about μ_i is desired. If, for healthy crewmembers, a functional relationship of the form $\sigma^2_{norm} = f(\mu)$ were found, then we could approximate the degraded variance by assuming the relationship:

$$\frac{\sigma^2_{deg}}{\sigma^2_{norm}} = \frac{f(\mu/p)}{f(\mu)} . \qquad (1)$$

A scatter plot of σ_i versus μ_i was made for each crew type, and a combined plot for all four crew types is shown in Fig. 2, which shows that the upward, linear trend is nebulous.

Grouping tasks according to the type of activity was then considered. The types chosen were:

1. cognitive,
2. physically demanding,
3. normal.

This partitioning of tasks was suggested by the previous IDP study [Anno, Wilson, and Dore, 1983]^1, where trends were observed indicating that certain physically demanding tasks deviated substantially from the norm. A physically demanding task is one that would more effectively fatigue a crewmember compared to other tasks and thus

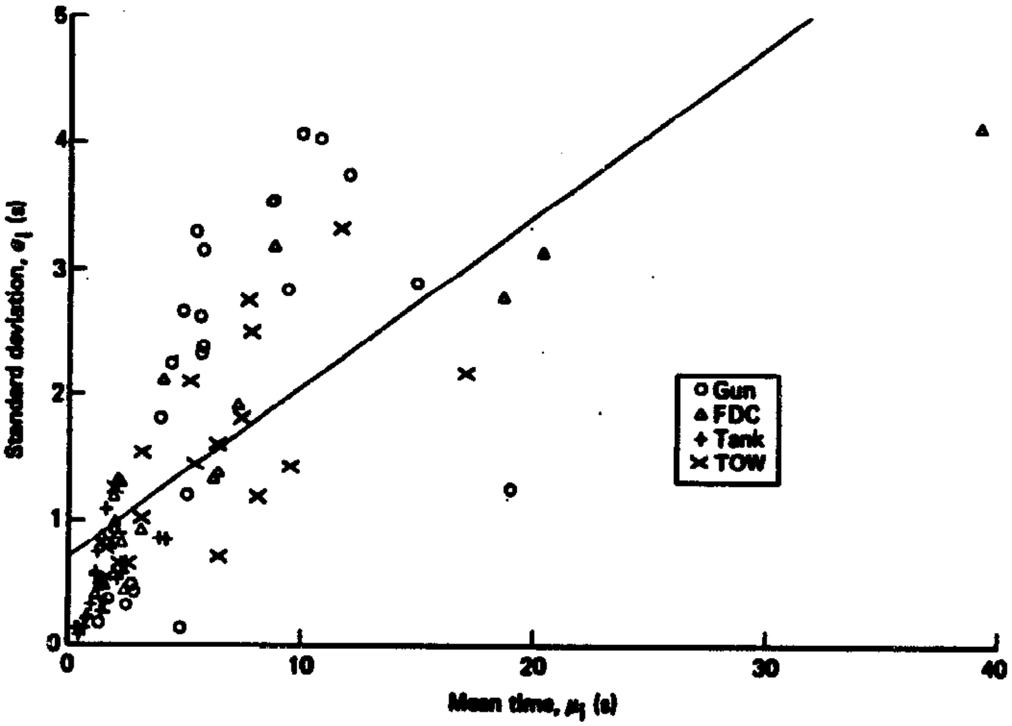
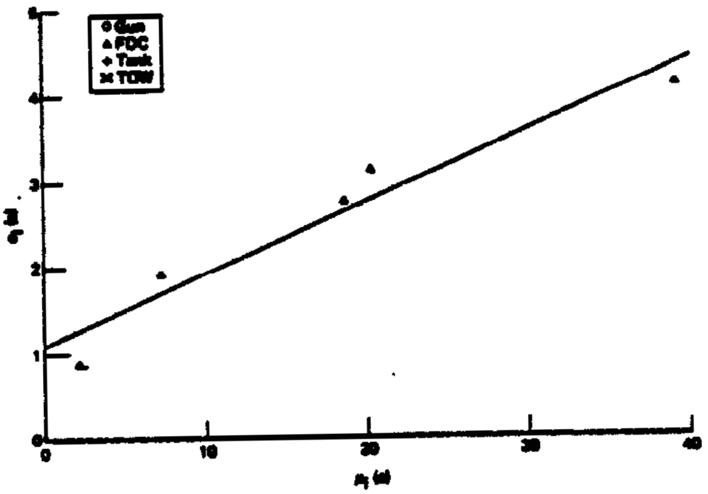


Figure 2. Scatter plot of α versus σ for all tasks.

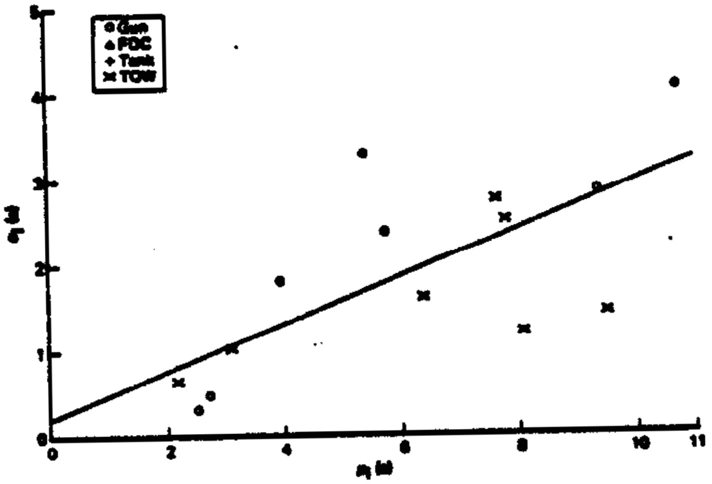
lengthen task performance time. Similar deviations were also observed for certain FDC tasks which require intense mental concentration. These "cognitive" tasks are indicated by a "C" in the column titled "Task Type" in Table A2 (Appendix). The physically demanding tasks are similarly marked with a "P" in Tables A1, A3, and A4 (Appendix). The remainder of the tasks are normal and are unmarked.

Examples of scatter plots of σ_i versus μ_i for the three groupings listed above are shown in Fig. 3. Again, only a very diffuse trend is evident. To assess these results, several statistical tests were applied, including the student-t and standard error. The results are summarized in Table 3, where the A and B parameters refer to a best fit of the form, $\sigma = A + B\mu$. The results indicate that the activity groupings are not sufficiently distinct from one another to justify distinguishing between them. Figure 4 is a scatter plot of σ_i^* versus μ_i^*, where

a. Cognitive.



b. Demanding.



c. Normal.

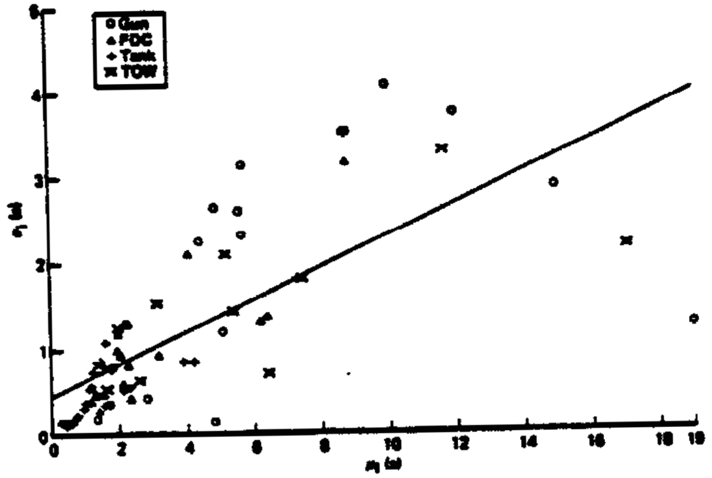


Figure 3. Scatter plot of σ versus μ by task group.

Table 3. Regression results.

	Crew Type	Tasks	SD	A	B	SD(A)	SD(B)	COV(A,B)
1	Gun crew	23	1.1702	1.0741	0.1589	0.4576	0.0572	-0.0222
2	FDC crew	23	0.4979	0.7873	0.0999	0.1271	0.0120	-0.0009
3	Tank crews	20	0.1982	0.1640	0.1943	0.0740	0.0427	-0.0025
4	TOW crew	19	0.5916	0.7269	0.1312	0.2411	0.0344	-0.0068

Standard Error and T-Statistics

Comparisons	Intercept (A)			Slope (B)		
	Standard Error	Student-T	DF	Standard Error	Student-T	DF
Gun crew/FDC crew	0.0990	2.8957	44	0.0122	4.8362	44
Gun crew/Tank crew	0.0968	9.3975	41	0.0153	2.3204	41
Gun crew/TOW crew	0.1103	3.1476	40	0.0143	1.9345	40
FDC crew/Tank crew	0.0312	19.9523	41	0.0099	9.5576	41
FDC crew/TOW crew	0.0613	0.9845	40	0.0083	3.7845	40
Tank crew/TOW crew	0.0577	9.7523	37	0.0124	5.0963	37

Stress Type	Tasks	SD	A	B	SD(A)	SD(B)	COV(A,B)
Cognitive	5	0.3567	1.1004	0.0832	0.2702	0.0125	-0.0027
Demanding	14	0.8310	0.2209	0.2725	0.5364	0.0804	-0.0392
Normal	66	0.7443	0.4671	0.1859	0.1295	0.0231	-0.0021
All tasks	85	0.8077	0.7029	0.1354	0.1170	0.0152	-0.0012

Standard Error and T-Statistics

Comparisons	Intercept (A)			Slope (B)		
	Standard Error	Student-T	DF	Standard Error	Student-T	DF
Cognitive/Demanding	0.1875	4.6909	17	0.0222	8.5278	17
Cognitive/Normal	0.1219	5.1957	69	0.0063	16.4060	69
Cognitive/All tasks	0.1215	3.2715	88	0.0058	8.9665	88
Demanding/Normal	0.1442	1.7071	78	0.0217	3.9950	78
Demanding/All tasks	0.1439	3.3492	97	0.0215	6.3651	97
Normal/All tasks	0.0204	11.5718	149	0.0033	15.3807	149

NOTE: DF = degrees of freedom.

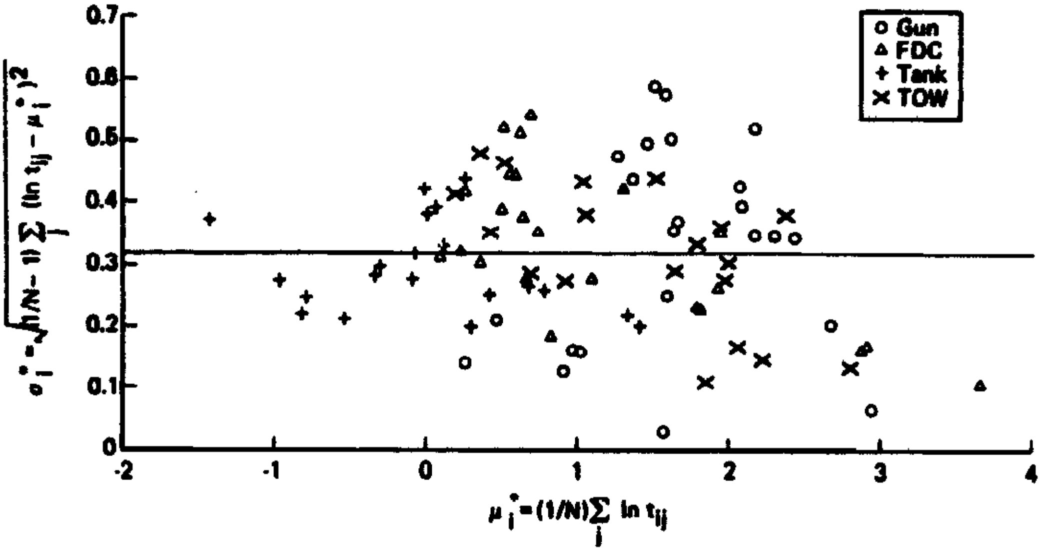


Figure 4. Scatter plot of σ^* versus μ^* for all tasks.

$$\mu_i^* = \frac{1}{N}\sum_{j=1}^{N} \ln t_{ij} \;, \tag{2a}$$

and

$$\sigma_i^* = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(\ln t_{ij} - \mu_i^*\right)^2} \;, \tag{2b}$$

and where $j = 1,2,\ldots,N$ ranges over trials for task i, and the t_{ij} are measured times. A best-fit, straight horizontal line was fit to this pattern, and we concluded that such a curve (with slope = 0) was as good a fit as any other. Thus, the algorithm for sampling for degraded times becomes:

> Sample random variables, τ, from a normal distribution with mean, $\mu_i^* - \ln p$, and standard deviation σ_i^*; then, $t = \exp(\tau)$.

This is equivalent to sampling t from a lognormal distribution with mean,

$$\mu(p) = \mu(p=1)/p = \exp(\mu^* - \ln p + \sigma^{*2}/2) \;, \tag{3a}$$

and standard deviation (SD),

$$\sigma(p) = \sqrt{e^{2(\mu^* - \ln p) + \sigma^{*2}} \left(e^{\sigma^{*2}} - 1\right)} \quad , \tag{3b}$$

i.e.,

$$\begin{aligned} \sigma(p) &= \sigma(p=1) \sqrt{e^{-2 \ln p}} \\ &= 1/p \cdot \sigma(p=1) \quad . \end{aligned} \tag{3c}$$

This is the desired relationship for $\sigma(\mu)$. Accordingly, an alternative form of sampling for t, which is equivalent, is:

> Sample random variables, t, from a lognormal distribution mean, $\mu(p=1)/p$, and standard deviation, $\sigma(p=1)/p$.

TASK DIAGRAM DEVELOPMENT

The decomposition of an engagement involving several interacting crewmembers into a specific set of modular tasks depends on the level of resolution required. The particular structures chosen for collecting the empirical data were designed to facilitate accurate timings between natural breakpoints in the action. This is illustrated by considering a representative specific task for the M60A3 tank crew, where the task commences immediately after the tank commander (TC) orders "backup." The full description states that the task consists of the driver backing up until the TC orders "stop," whereupon the driver stops and parks the tank. Here two different individuals are involved, and three distinct actions must be performed in sequence, i.e., an action by the driver, then an order from the TC, followed by a final action by the driver. For an accurate simulation in which the TC and driver may be in very different degradation states, the distinction between the three components of this "task" can be significant. A similar situation arises in several of the other empirical tasks.

As a consequence, many of the empirical tasks were subdivided into two or more simulation tasks for incorporation into the final task diagrams. This, in turn, necessitated distributing the mean task time and the associated variance among the subdivided tasks without altering their combined values. Although this could only be done by careful intuitive judgment, the net mean time and variance for the set, which must remain as measured, were preserved. Since for the original measurements these appeared as a single, continuous activity, the addition of detailed internal structure to this set of tasks had negligible effect.

The procedure for redistributing the variances among subdivided tasks results from requiring that three conditions be met. For an empirical task with mean time t, the subtasks (a and b) are assigned mean times such that

$$t_a + t_b = t \ . \tag{4}$$

Assuming that the tasks are independent, the standard deviations are then constrained such that

$$\frac{\sigma_a}{\sigma_b} = \frac{t_a}{t_b} \qquad \text{and} \qquad \sigma_a^2 + \sigma_b^2 = \sigma^2 \ , \tag{5}$$

and are then given by

$$\sigma_b^2 = \frac{\sigma^2}{1 + (t_a/t_b)^2} \qquad \text{and} \qquad \sigma_a^2 = \sigma^2 - \sigma_b^2 \ . \tag{6}$$

The resulting input values used for the simulation are tabulated in Tables A1, A2, A3, and A4 of the Appendix, along with the final task lists for the four crew types. These correspond to the task diagrams given in Figs. A1, A2, A3, and A4 of the Appendix. In Tables A1 through A4, the column marked "data task" contains the original empirical task number from the measurement data, and is followed by a letter (A, B, or C) if the task was subdivided. Empirical tasks 21 through 24 for the M109 gun crew were in fact copies of tasks 8 through 11, i.e., no independent measurements were taken for the repeat of the loading operations.

MONTE CARLO SIMULATION

The CREW-III Monte Carlo simulation code [Dore, 1986][7] was utilized to predict the mean overall engagement time for each crew type. This involved computing 100 independent histories (realizations) of the simulated engagement, resulting in a predicted mean time for each type of engagement. Reconciliations of the mean overall engagement times obtained from this simulation with measured ones are shown in Table 4. Since some of the simulated engagements contained additional tasks not included in earlier measured engagements, correction terms were needed to make a meaningful comparison between the predicted and measured times. For example, the initial six tasks for the tank crew, during which the tank commander and gunner are searching for a target, were included to conform to the earlier engagement scenario. These tasks consume an average of about 1.7 s which were not included in the empirical measurements.

The corrections for the tank crew were suggested by the team who performed timing measurements in the field [Moyer and Lam, 1987][4] after they noticed loaders and drivers taking longer in the field exercises than had been allowed for the COFT simulations.

The right-hand column of Table 4 summarizes the relative accuracy with which the Monte Carlo simulation (addressed in the following section) reproduces the empirical results. Except for the FDC crew, for which many empirical measurements are very doubtful, the agreement is satisfactory.

Table 4. Reconciliation of overall engagement times.

	CREW-III Results			Empirical Data			
Crew Type	Raw Time	Correction	Net	Raw Time[a]	Correction	Net	Percent
Gun	103.18		103.18	116.43		116.43	-12.8
FDC	93.38		93.38	69.69	+6.86[b]	76.55	+18.0
Tank	28.06	-1.70[c]	26.36	22.79	+2.71[d]	25.50	+3.26
ITV-TOW	157.82	-28.00[e]	129.82	141.57		141.57	-9.05

[a] Empirical values for total engagement time reflect the less than ideal conditions under which field data measurements were gathered.
[b] Tasks 11 through 14 were removed.
[c] Initial target acquisition.
[d] Loader and driver too fast in COFT.
[e] Driver tasks.

The CREW-III code was then applied to selected combinations of degraded crewmembers. The degradation states chosen and the final mean engagement times found are summarized in Tables A5 through A8 of the Appendix. For some crew types, additional points were included to ensure that the entire four-space, i.e., four crewmembers, was sufficiently represented.

FOUR-DIMENSIONAL FUNCTIONAL FIT

A simplified expression for overall crew performance as a function of the performance levels of the individual crewmembers was derived considering a crew consisting of four crewmembers, each responsible for a specific subset of the full set of tasks making up the engagement. Assuming that all tasks are performed sequentially, the total time for the engagement would simply be the sum of the time taken by each crewmember. Let T_i be the sum of the normal (undegraded) times for the tasks of crewmember i, i = 1, 2, 3, 4, so that the total mission time is

$$T_{tot} = \sum_{i=1}^{4} T_i \quad . \tag{7a}$$

The performance of a crewmember is defined as $p_i = T_i/t_i$, where t_i is the longer time required to accomplish the same set of tasks when the crewmember is degraded. Thus, p_i is always between 0 and 1. Similarly, the overall crew performance is given by $P = T_{tot}/t_{tot}$, where $t_{tot} = \sum t_i$. It then follows that

$$P = \frac{1}{\sum\left(\frac{\alpha_i}{P_i}\right)} \quad , \tag{7b}$$

where,

$$\alpha_i = T_i/T_{tot} \ . \tag{8}$$

Equation 7b shows that each of the crewmember performances (α_i) is reciprocally weighted by the fraction of the total mission time contributed by each crewmember when all are undegraded. The assumption of a series of sequential tasks consistent with this expression is true only specifically for the ITV-TOW crew. The other three crew types require the crewmembers to perform tasks in a series-parallel fashion, where at various times during the engagement tasks are performed simultaneously. However, by using Eq. 7b as a guide, the Monte Carlo results for these other three crew types were found to be remarkably well fit by a more generalized form of Eq. 7b,

$$P = \frac{1}{\sum_{i=1}^{4}\left(\frac{\alpha_i}{P_i}\right)^{\beta_i}} \quad . \tag{9}$$

The β exponent allows four more degrees of freedom to account for interaction effects between crewmembers. Both the α's and β's were determined from a simplex algorithm which mimimized the standard error over all points. The starting guesses for the α's were determined from Eq. 8 and the β's were initialized to 1.0; in all cases, the results converged rapidly to a distinct minimum. The resulting values are tabulated in Table 5, along with the correspondence between the indices and crewmember positions.

The CREW-III simulation for the ITV-TOW crew incorporated only three crewmembers: squad leader, gunner, and loader. The fourth crewmember, the driver, was not included since the empirical data for the ITV-TOW crew was taken without a driver present. As already mentioned, the ITV-TOW tasks are performed sequentially, and therefore the β's are all 1.0. The accuracy of the CREW-III simulation and subsequent fit to Eq. 9 is evident by noting the near unity values of the β's for the ITV-TOW crew in Table 5.

The previous IDP study [Anno, Wilson, and Dore, 1983][1] did include a driver, whose total task time was normally 28.0 s for maneuvering and parking at the launch site prior to the other crewmembers beginning their tasks. For both consistency with the previous IDP study and because an actual ITV-TOW crew includes a driver, a modification was made to the results to incorporate the effects of a driver into the results.

This modification consisted of adding an additional 28 s to the engagement time for vehicle motion with an undegraded driver. Then, if T_{tot} = 129.824 s is the total engagement time for the three man

Table 5. Fit parameters.

Parameter	Gun Crew	FDC Crew	Tank Crew	TOW Crew
α_1	0.073953	0.23466	0.24249	0.13111
α_2	0.074018	0.37234	0.22197	0.47759
α_3	0.24954	0.44990	0.31879	0.17742
α_4	0.59286	--	0.31198	0.21546
β_1	1.2502	1.6017	0.98866	1.0389
β_2	0.67995	0.81154	1.1181	1.0137
β_3	1.0037	0.98212	1.2815	1.0000
β_4	1.1755	--	1.0046	0.96201

Crew Type	Crewmember Index (i)			
	1	2	3	4
M109 howitzer gun crew	Chief of Section	Gunner	Assistant Gunner	No. 1 Cannoneer
Fire Direction center (FDC)	Fire Direction Officer	Computer	Horizontal Control Operator	--
M60A3 tank crew	Tank Commander	Gunner	Loader	Driver
M901 ITV-TOW crew	Squad Leader	Gunner	Driver	Loader

crew, the new total time (to accommodate the driver) will be $T'_{tot} = T_{tot} + 28 = 157.824$ s. Also, the total engagement time with degraded crewmembers becomes

$$T' = T + 28/p_D, \quad \text{where} \quad T = T_{tot} \cdot \sum_{i=1}^{3} \left(\frac{\alpha_i}{p_i}\right)^{\beta_i} \tag{10}$$

is the total engagement time for the three-man crew, and p_D is the performance level of the driver. Thus,

$$\frac{1}{P'} = \frac{T'}{T'_{tot}} = \frac{T_{tot}}{T_{tot} + 28} \cdot \sum_{i=1}^{3} \left(\frac{\alpha_i}{p_i}\right)^{\beta_i} + \frac{28}{T_{tot} + 28} \left(\frac{1}{p_D}\right) \tag{11}$$

$$= \frac{T_{tot}}{T_{tot} + 28} \left[\sum_{i=1}^{3} \left(\frac{\alpha_i}{p_i}\right)^{\beta_i} + \left(\frac{28/T_{tot}}{p_D}\right) \right] .$$

so we find that

$$P' = \frac{K}{\sum_{i=1}^{4} \left(\frac{\alpha_i}{p_i}\right)^{\beta_i}} , \quad \text{with } \alpha_4 = 28/T_{tot} , \quad \beta_4 = 1, \quad p_4 = p_D, \text{ and } K = (T_{tot} + 28)/T_{tot} . \tag{12}$$

This is the same form as Eq. 9 except for the constant, K, which is easily removed by incorporating it into the α's.

Comparisons of the simulation results and the values predicted by Eq. 9 are included in Tables A5 through A8 of the Appendix. Figure 5 gives sample plots of crew performance as a function of the performance of each individual crewmember. The plots also illustrate the functional fit of the Monte Carlo generated data (points) based on Eq. 9; the full set of data used in the parameter fits is tabulated in Tables A5 through A8.

DOSE AND TIME DEPENDENCE

The crew performance formula Eq. 9 is suitable for many applications involving degraded crewmembers. One such application is the Army Unit Resiliancy Analysis (AURA) code at BRL [Klopcic and Roach, 1984]^2. AURA allows reconstitution of a tactical unit with any mix of degraded personnel, regardless of the underlying cause of the degradation. For instance, the limitations on crew performance caused

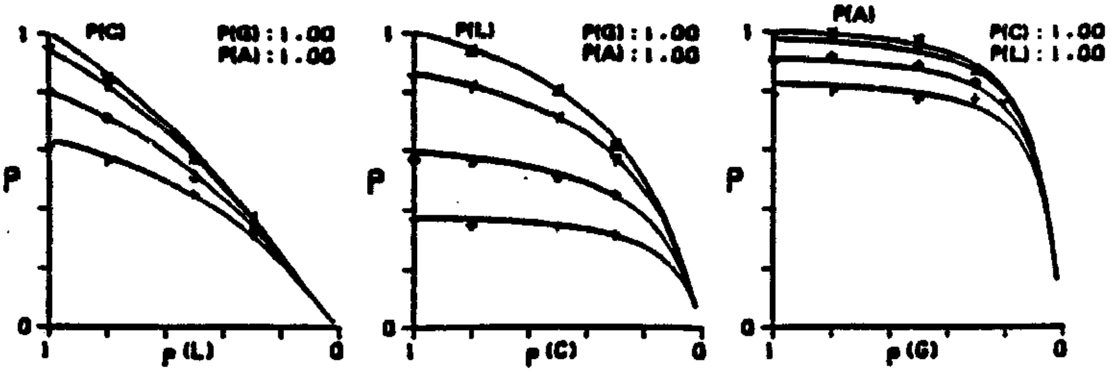


Figure 5. 4-D functional-fit of crew performance.
[M109 gun crew performance (P); chief of section (C); gunner (G); assistant gunner (A); and loader (L)]

by wearing mission-oriented protection posture (MOPP) gear in a chemical attack environment can be simulated by applying Eq. 9.

Other battlefield simulation codes are primarily concerned with effects on a tactical unit directly, rather than indirectly through the performance of the crewmembers. One such is the JANUS conflict simulation code at LLNL [Blumenthal, et al., 1984][3]. One requirement of JANUS is to simulate the time-dependent performance of combat crew units after acute exposure to an intermediate dose of ionizing radiation from a tactical nuclear weapon detonation.

Providing such a model extends naturally from the earlier IDP study [Anno, Wilson, and Dore, 1983][1] in which the performance of individuals was characterized for acute exposure as a function of dose and time. A given external dose to a unit (e.g., a tank) is reduced by an appropriate shielding factor for each crew position in the vehicle, resulting in a dose to each crewmember. The results of the IDP study were then applied to determine the performance level of each crewmember at some postexposure time, and Eq. 9 was then applied to yield the overall unit performance at that time.

A modification of the above procedure is necessary because not all individuals develop any or all of the disabling symptoms of acute radiation sickness, the cause of degraded performance. The individual crewmember performance values reported in the IDP study [Anno, Wilson, and Dore, 1983][3] assume an individual expresses these symptoms; however, there is no degradation for those not affected, i.e., their performance level is 1.0. This individual variability in response to radiation exposure was accounted for where the expected performance is determined for the unit by combining the sixteen possible degraded/nondegraded situations of four crewmembers, each either expressing or not expressing symptoms; symptom incidence fractions are weighting factors applied for the affected states (ones in Table 7) in determining the expected performance.

Estimates of incidence fractions for each of the major symptoms (e.g., nausea, vomiting, fatigability) of acute exposure to ionizing radiation have been determined as a function of dose [Anno, Wilson, and Baum, 1985][8]. While these incidence fractions are specific to each type of symptom, they are highly correlated, and the actual performance degradation of an individual is usually dominated by the initial and most severe symptom developed. This allows an overall incidence curve to be determined by assuming an individual develops the characteristic degradation with the onset of any one symptom. Curves of incidence fraction versus dose for the major initial symptoms of acute radiation sickness are plotted in Fig. 6 based on both lognormal and logistic forms of the distributions with dose. The overall incidence curve consists of the leftmost of these symptom curves at any incidence level. The resulting three-part relationship is given in Table 6.

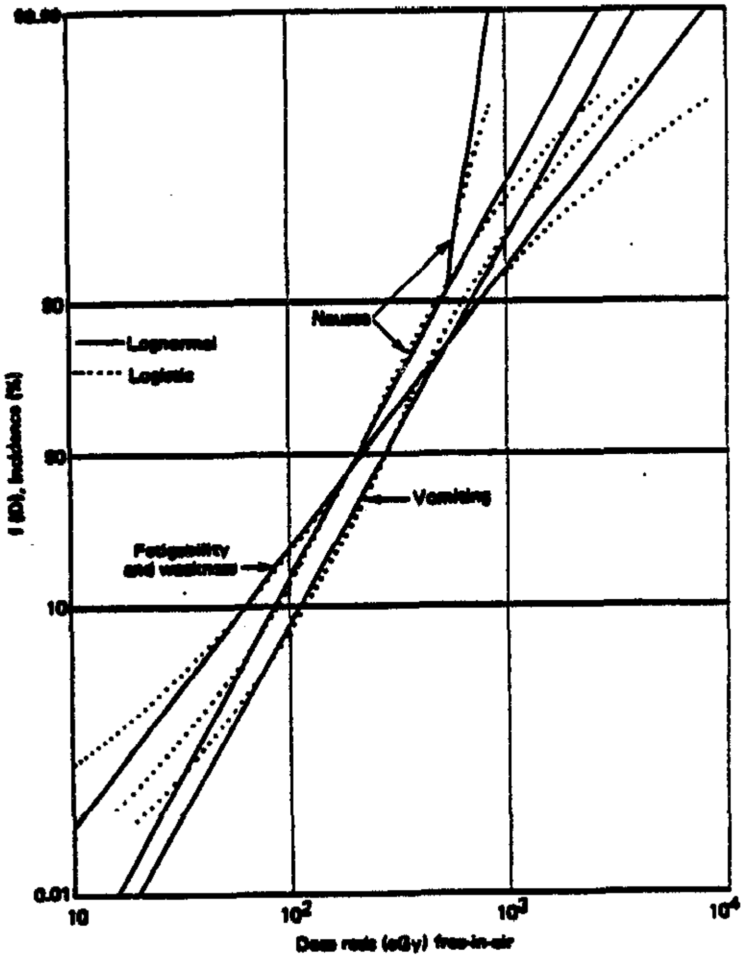


Figure 6. Symptom incidence.

Table 6. Symptom incidence relationship.

Principal Symptom	Dose Range (rads)	a	b
Fatigability and Weakness	0 - 185	8.75	1.63
Nausea	185 - 530	12.2	2.30
Nausea	> 530	49.8	8.27

Symptom incidence:

$$f(D) = \frac{1}{1 - e^{a - b \ln D}} \quad .$$

The expected performances were then calculated for the four crew types. First the protection factors from Table 7 were applied to reduce the exposure external to the vehicle by an amount appropriate for inside the type of vehicle. Note that only the M60A3 tank varied in mass sufficiently to afford different degrees of radiation protection dependent upon where in the vehicle the crewmember is located. The weights and performances of each crewmember were then calculated for all 16 combinations of symptom incidence and then combined together with Eq. 9, in the manner given as follows: Let D_o be an acute dose of ionizing radiation external to the vehicle which occurs at time $t = 0$. Then, for the ith crewmember ($i = 1, 2, 3, 4$), define:

PF_i = protection factor,
$D_i = D_o/PF_i$,
$p_i(D_i, t)$ = conditional performance level of ith crewmember,
$f_i(D_i)$ = probability that symptoms are expressed by ith crewmember.

Table 7. Radiation shielding protection factors.

Crew and Position	Factor
M109 howitzer gun crew	
Chief of section	1.3
Gunner	1.3
Assistant gunner	1.3
No. 1 cannoneer	1.3
FDC	
Fire direction officer	1.0
Computer	1.0
Horizontal control operator	1.0
M60A3 tank crew	
Tank commander	1.4
Gunner	2.0
Loader	1.9
Driver	3.8
M901 ITV-TOW crew	
Squad leader	1.4
Gunner	1.4
Driver	1.4
Loader	1.4

The 16 combinations can then be represented by taking:

0 => Unaffected: $p_i^* = 1.0$, $w_i = 1 - f_i$,
1 => Affected: $p_i^* = p_i$, $w_i = f_i$,

and arranging all combinations as a binary series, i.e.,

Case n:		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
Crew	2	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
member	3	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	4	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1

The weight assigned to the *n*th case is then given by

$$W_n = \prod_{i=1}^{4} w_i \ , \quad \textit{where} \quad \sum_{n=0}^{15} W_n = 1 \ , \tag{13}$$

and the crew performance (given by Eq. 9) for case n is

$$Q_n = \frac{1}{\sum_{i=1}^{4} \left(\frac{\alpha_i}{p_i^*} \right)^{\beta_i}} \ . \tag{14}$$

The combined average crew performance is then given by

$$P(D_o, t) = \sum_{n=0}^{15} W_n Q_n \ . \tag{15}$$

The crew unit performance valuess which result from the calculations given above are tabulated in Tables A9 through A12 (Appendix) as an evenly spaced grid of points in log dose (external to the unit) by log time. Figures 7 through 10 give three-dimensional plots of the performance data for the four crew types simulated; Figures 11 through 14 give the same performance data expressed in ten percent contour intervals.

DISCUSSION

The methodology followed in this study to develop crew unit performance demonstrates that a fairly complex procedure, such as that required for modeling the collective actions of individual crewmembers, can often be replaced by a fairly simple yet effective algorithm. This result embodies all the appropriate details and has the obvious advantage of reducing computation, especially for utilization in larger scale battlefield simulations such as AURA, which involve a great amount of other computations. Also, the algorithm

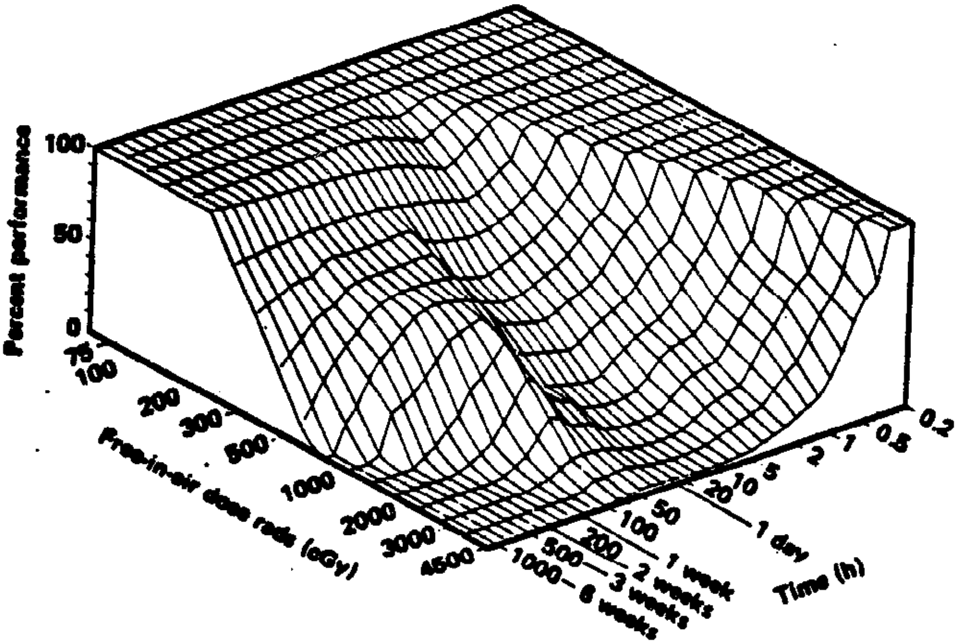


Figure 7. Performance for the gun crew.

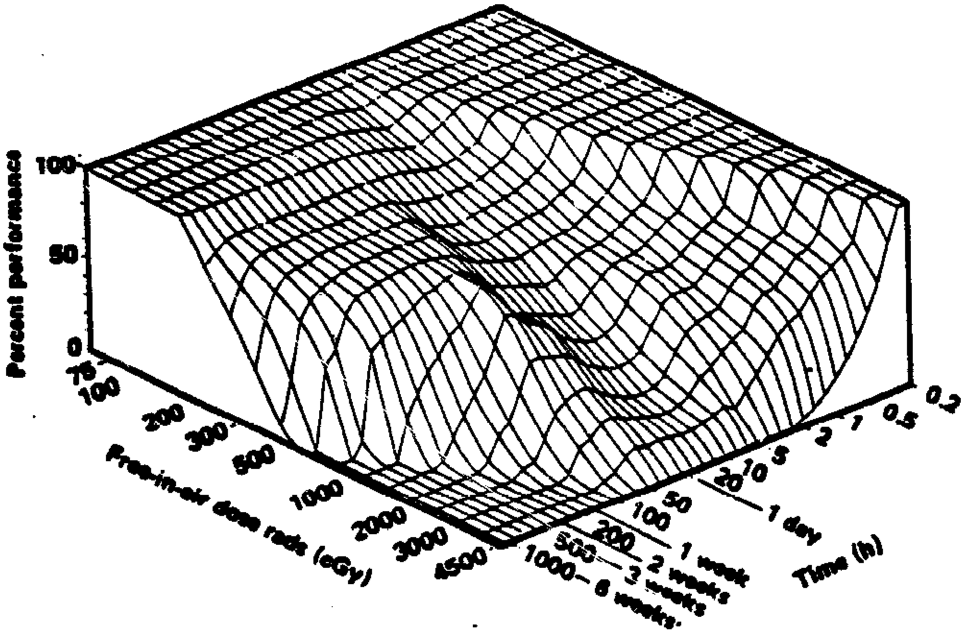


Figure 8. Performance for the FDC crew.

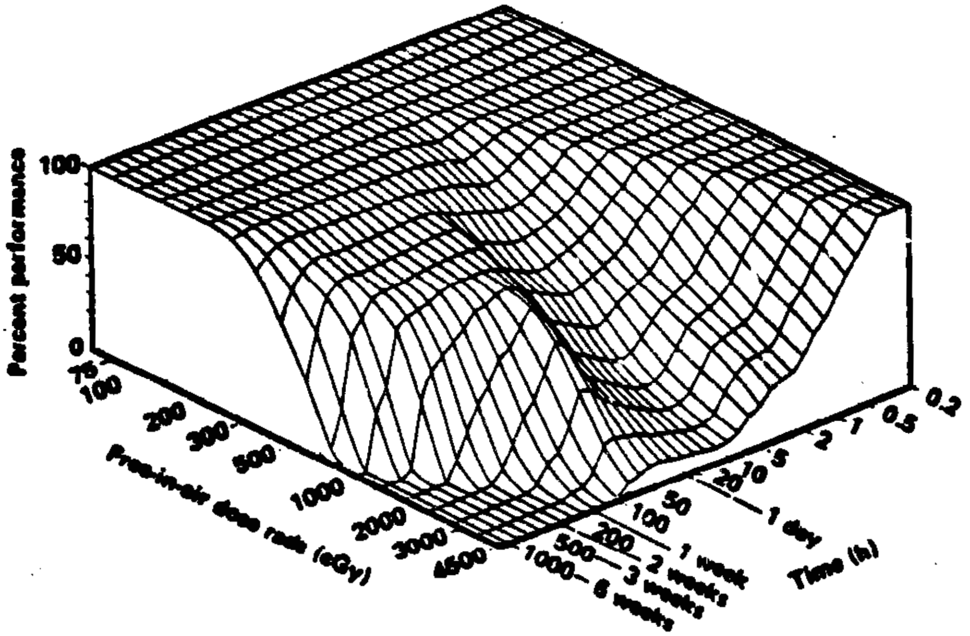


Figure 9. Performance for the tank crew.

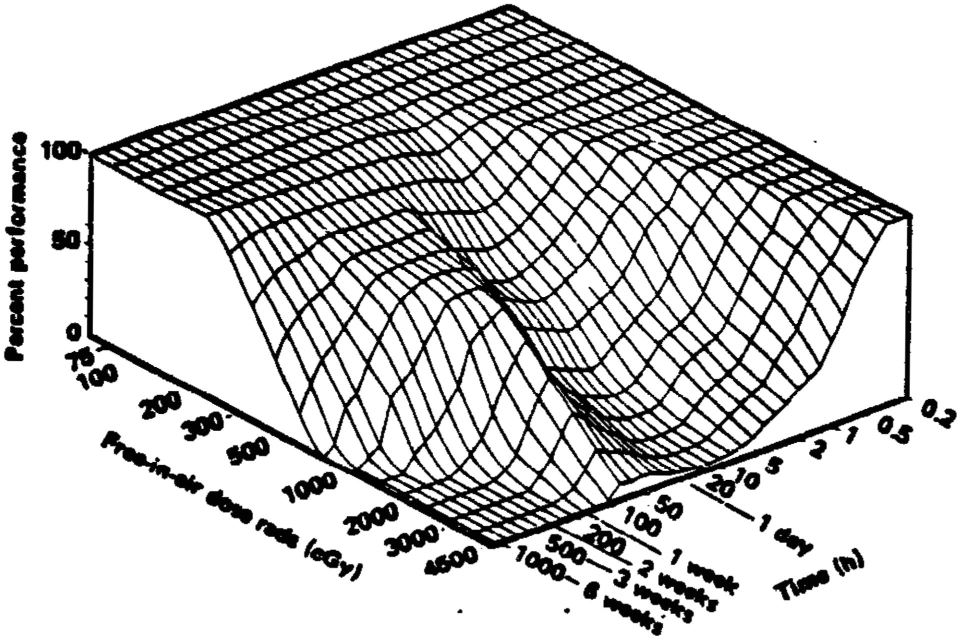


Figure 10. Performance for the ITV-TOW crew.

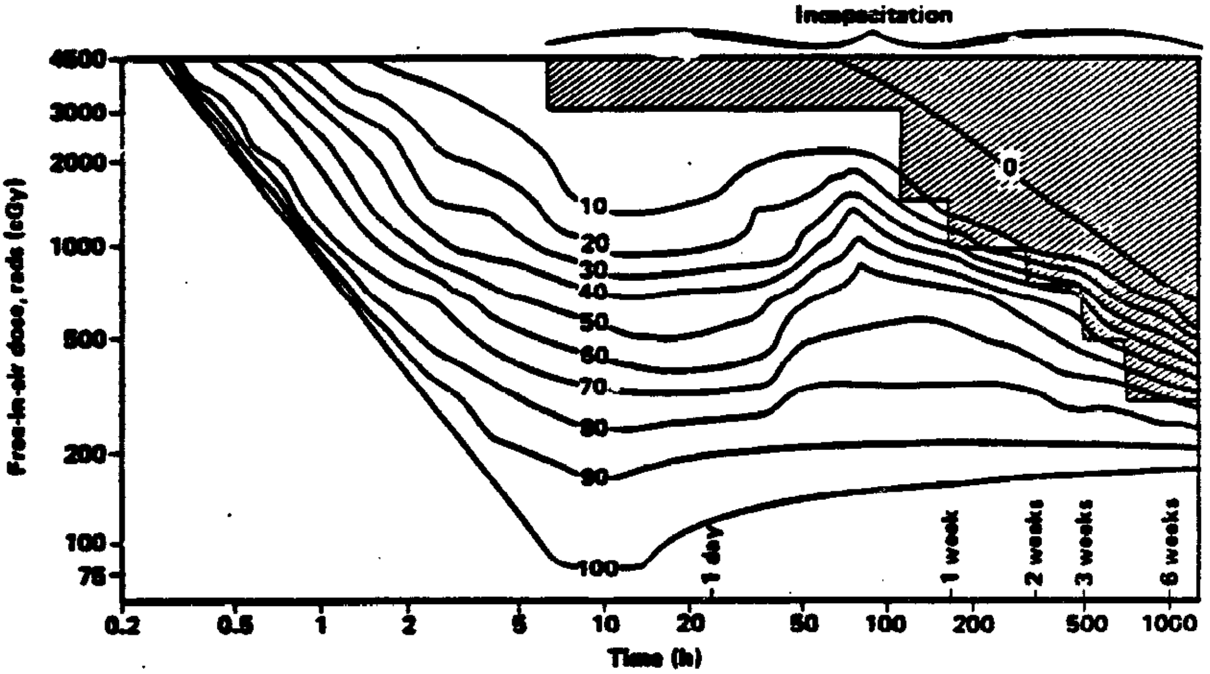


Figure 11. Gun crew--10 percent performance intervals.

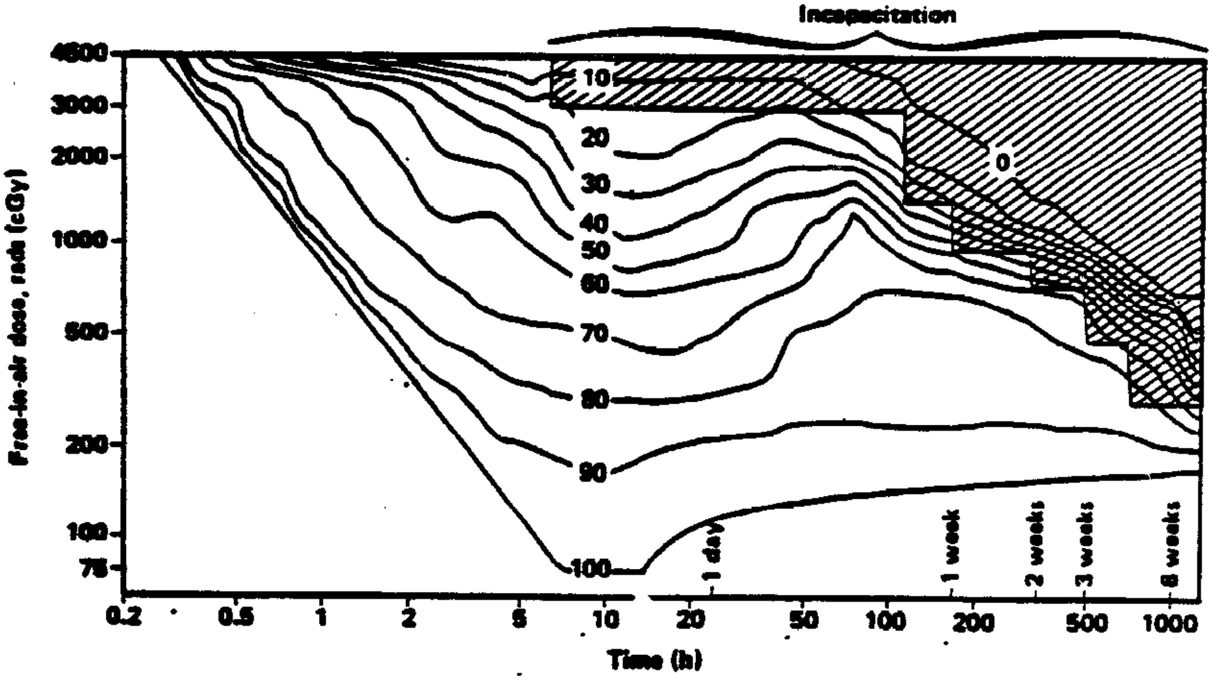


Figure 12. FDC crew--10 percent performance intervals.

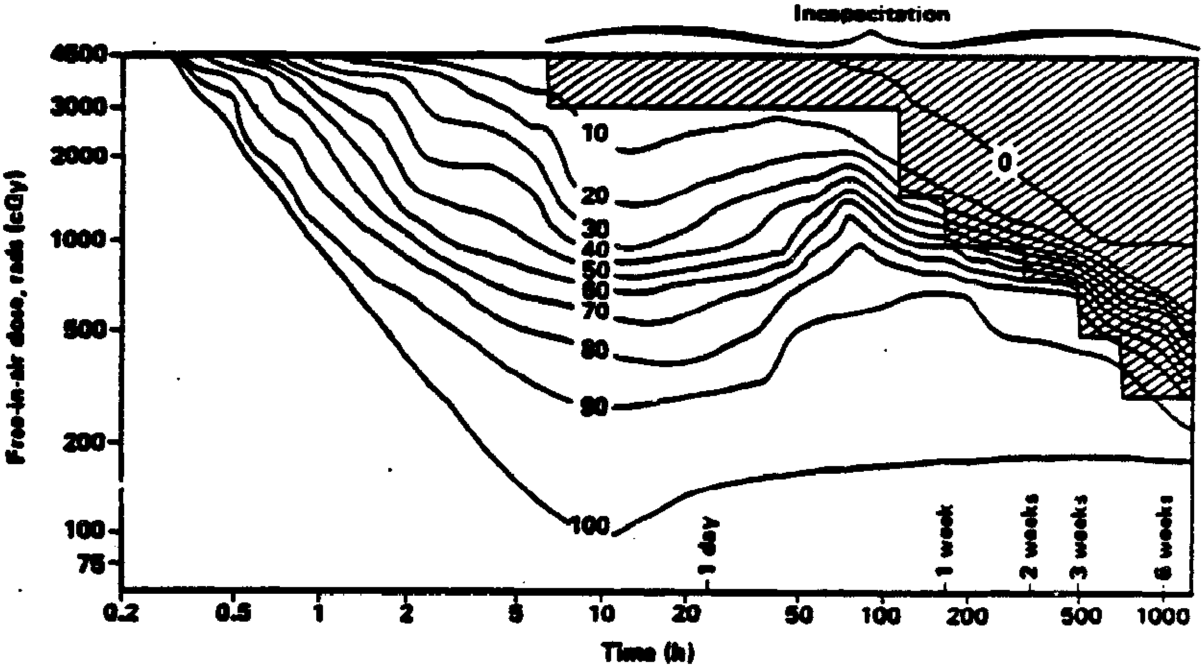


Figure 13. Tank crew--10 percent contour intervals.

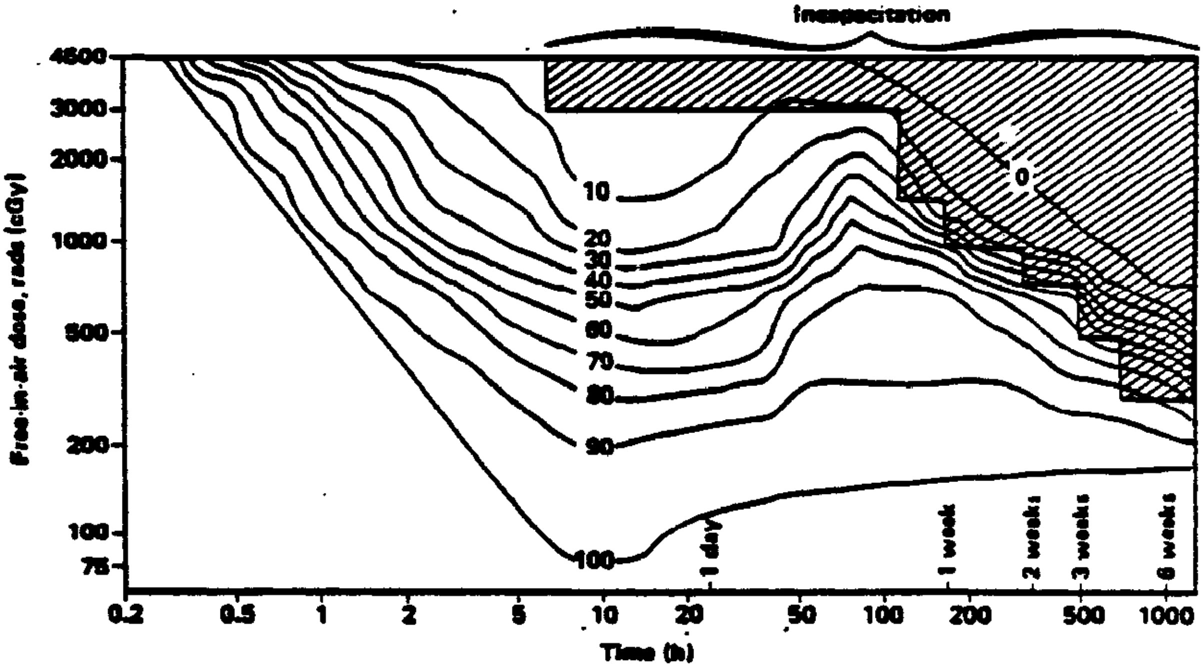


Figure 14. ITV-TOW crew--10 percent contour intervals.

provides the flexibility that enables crew unit performance estimates to be made for arbitrary levels of degraded states distributed among the various crewmembers, regardless of the cause.

The methodology demonstrated can also be extended to specifically address crew unit performance where the performance of individual crewmembers may be impaired by the effects of ionizing radiation exposure. Here the conditional crew unit performance algorithm (Eq. 9) is modified by taking into account both individual variability in response to radiation effects as well as extra- and intravehicular environmental dose levels among the crewmembers.

Crew unit performance data generated in this study pertains to fixed short combat engagement scenarios for the combat crews evaluated along a post-acute-exposure time frame (up to 6 weeks) for extravehicular environmental dose levels. These data provide input to other larger-scale battlefield simulation programs such as JANUS, which require prepackaged radiation effects embedded in crew performance values.

However, based on the approach described, discrete component crew performance estimates and associated probabilities can easily be generated for arbitrary dose and time, utilizing the shielding factors and symptom incidence data presented in this report and the previously developed dose-time individual crewmember performance data [Anno, Wilson, and Dore, 1983][1]. Such discrete probabalistic crew unit performance data, for example, would be more appropriate for battlefield simulations which treat combat crew unit performance stochastically.

Aside from application to larger-scale battlefield simulation or gaming codes, the graphical presentations of crew unit performance can be useful for training and education purposes in developing tactical guidelines. In this regard, some general observations can be made and caveats noted.

Crew unit performance generally follows the level of performance of the individual crewmembers over dose and time, although not necessarily in a linear fashion. For example, overall crew performance is not greatly affected by any one crewmember who may be significantly degraded compared to the others until that crewmember's performance level approaches around 50 percent. A plot of crew unit performance against decreasing performance values of one of the crewmembers, all others fixed, would initially appear with a relatively shallow slope and then rapidly turn down to smaller values at some point of the individual's decreasing performance.

Task time variability, derived from field measurements according to a detailed characterization of subtasks that comprise a combat encounter scenario, comes from a population of participants in a normal state of health. Since it is impossible to acquire the corresponding measure of variability from a population of individuals impaired by the effects of ionizing radiation, we must assume that the standard deviation of task-performance time varies monotonically with the increased mean task-performance time expected when an individual is impaired by the effects of radiation. The assumed inverse performance relationship is consistent with the results from the statistical analysis of the field measurement data.

Finally, the crew unit performance values are all based on short but intense actions that take place during the tactical combat engagements modeled. In this regard, we are uncertain how applicable these performance values are to other combat crew scenarios where the action levels are not as intense and/or where the duration of performance extends over much longer periods, for example, as in logistic or support operations, road marches, or other activities where protracted fatigue might become a factor.

ACKNOWLEDGEMENT

This work was sponsored by the Defense Nuclear Agency (DNA) under contract DNA001-85-C-0352. The authors wish to express their appreciation for the assistance and cooperation of S. Levin of Technico Southwest, Inc., Espanola, New Mexico; M. J. Moyer of Science Applications International Corporation, La Jolla, California; and for the direction and guidance of R. W. Young of the DNA.

LIST OF REFERENCES

1. Anno, G. H., D. B. Wilson, and M. A. Dore, *Acute Radiation Effects on Individual Crewmember Performance*, Pacific-Sierra Research Corporation, Note 572, October 1983.

2. Klopcic, J. T. and L. K. Roach, *An Introduction to the Use of the Army Unit Resiliency Analysis (AURA) Methodology*, Vol. I, U.S. Ballistics Research Laboratory, Aberdeen Proving Ground, Maryland, Memorandum Report BRL-MR-3384, September 1984.

3. Blumenthal, D. K., et al., *An Analysis of a Massed Versus a Dispersed Attacking Ground Force in Close Combat*, Lawrence Livermore National Laboratory, Livermore, California, UCID-20010, 31 January 1984.

4. Moyer, M. J., and S. Lam, *M60A3 Combat Tank Data*, Science Applications International Corporation, unpublished, forwarded in a letter dated 6 March 1987 to R. W. Young, Defense Nuclear Agency, Washington, DC.

5. Moyer, M. J., T. K. O'Donoghue, and M. L. Feinberg, *The Final Report of the Task Timeline Analysis Project*, Science Applications International Corporation, La Jolla, California, Report 1-262-03-509-02, 12 September 1984.

6. Dore, M. A., and G. H. Anno, *Effects of Ionizing Radiation on the Performance of Selected Tactical Combat Crews*, Pacific-Sierra Research Corporation, Report 1846 (DNA-TR-8817), 1 July 1988.

7. Dore, M. A., *CREW III User's Guide*, Pacific-Sierra Research Corporation, Report 1660, 1 September 1986.

8. Anno, G. H., D. B. Wilson, and S. J. Baum, *Severity Levels and Symptom Complexes for Acute Radiation Sickness: Description and Quantification*, Pacific-Sierra Research Corporation, Report 1597, 30 November 1985.

APPENDIX

Figures A1 through A4
Tables A1 through A12

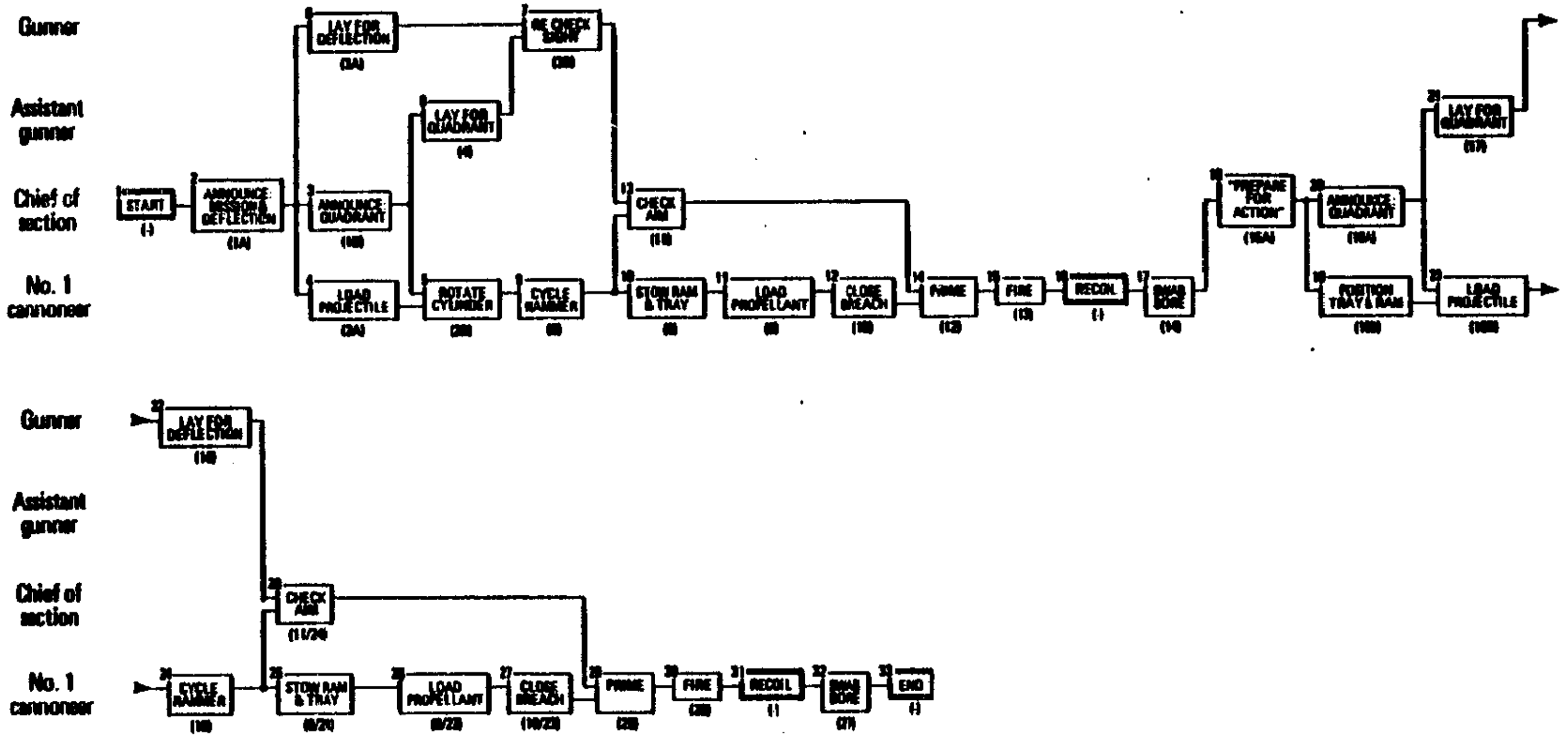


Figure A1. Task schematic for gun crew.

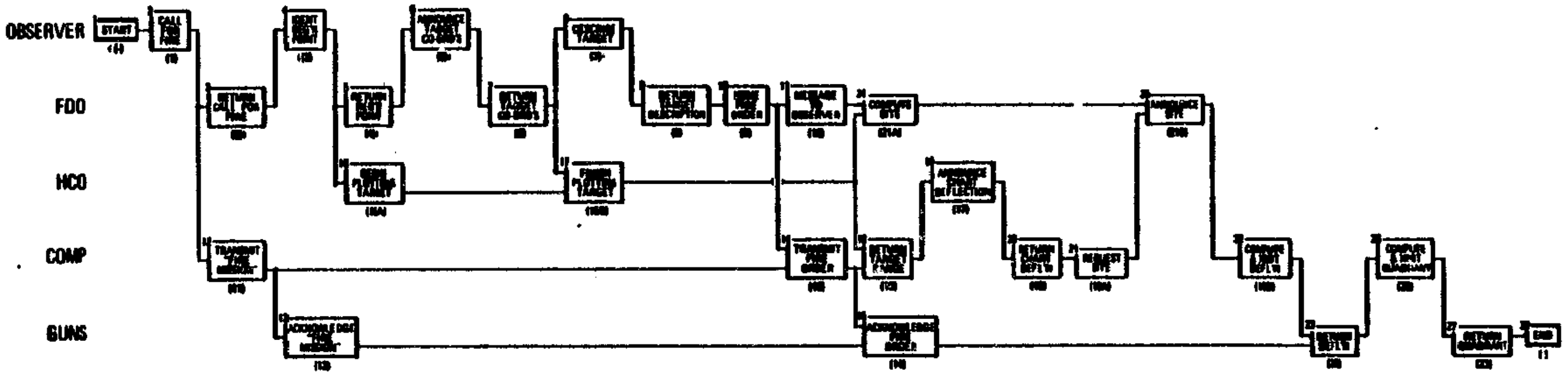


Figure A2. Task schematic for FDC crew.

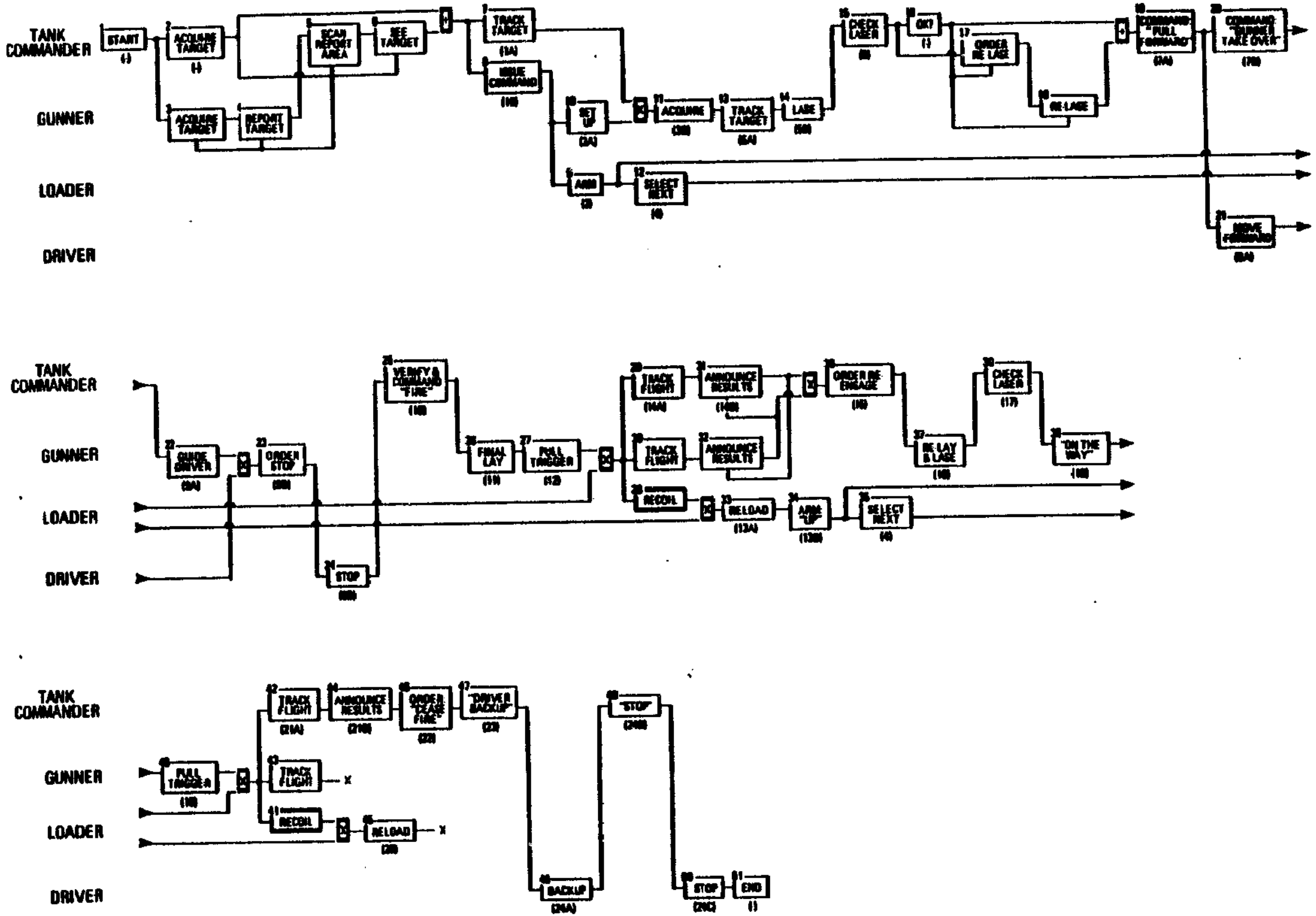


Figure A3. Task schematic for tank crew.

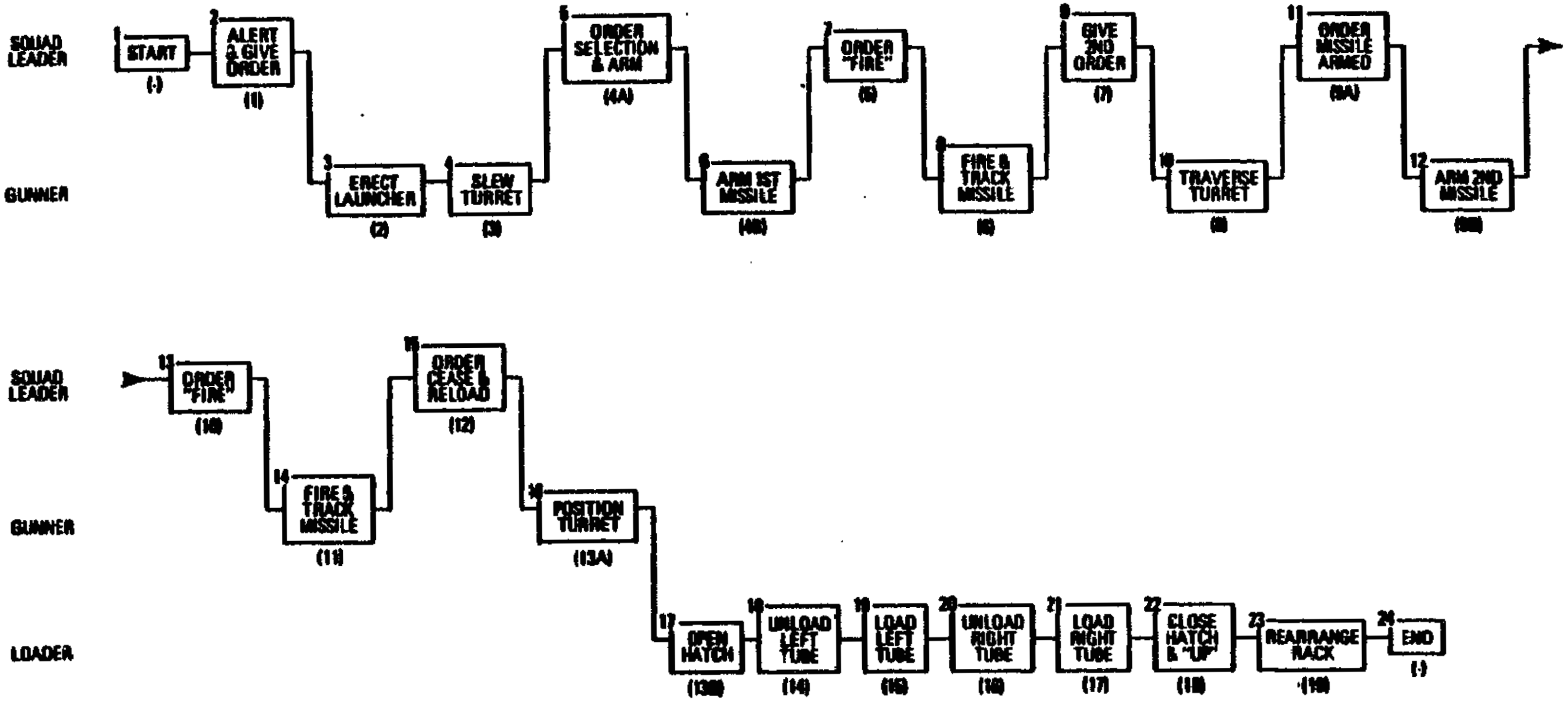


Figure A4. Task schematic for ITV-TOW crew.

Table A1. Input data for gun crew.

Task No.	Description	Data Task[a]	Task Type	Mean (sec)	SD (sec)	Crewmember	Precursor Task(s)
1	START	--		0.0	0.0	--	--
2	AN:MISSION&DEFL.	1A		15.99	1.226	C.SECT.	1
3	AN: QUADRANT	1B		3.00	0.230	C.SECT.	2
4	LOAD PROJECTILE	2A	P	4.74	2.926	NO. 1 C.	3
5	ROTATE CYLINDER	2B	P	1.00	0.617	NO. 1 C.	3 4
6	LAY FOR DEFL.	3A		8.44	4.446	GUNNER	2
7	RECHECK SIGHT	3B		1.50	0.790	GUNNER	6 8
8	LAY FOR QUAD.	4		11.96	3.725	AST.GNR.	3
9	CYCLE RAMMER	6		5.08	1.223	NO. 1 C.	5
10	STOW RAM & TRAY	8	P	2.72	0.718	NO. 1 C.	9
11	LOAD PROPELLANT	9		2.82	0.509	NO. 1 C.	10
12	CLOSE BREACH	10	P	2.51	0.443	NO. 1 C.	11
13	CHECK AIM	11		4.81	0.143	C.SECT.	7 9
14	PRIME	12		4.86	2.580	NO. 1 C.	12 13
15	FIRE	13		5.70	3.128	NO. 1 C.	14
16	RECOIL	--		1.00	0.0	--	15
17	SWAB BORE	14		8.75	3.553	NO. 1 C.	16
18	AN: NEXT	15A		1.00	0.583	C.SECT.	17
19	POS'N TRAY & RAM	15B	P	3.94	1.811	NO. 1 C.	18
20	AN: QUAD	16A		1.00	0.748	C.SECT.	18
21	LAY FOR QUAD	17		1.34	0.345	AST.GNR.	20
22	LAY FOR DEFL	18		1.72	0.689	GUNNER	21
23	LOAD PROJECTILE	16B	P	5.39	3.368	NO. 1 C.	19 20
24	CYCLE RAMMER	19		5.67	2.329	NO. 1 C.	23
25	STOW RAM & TRAY	(8/21)	P	2.72	0.718	NO. 1 C.	24
26	LOAD PROPELLANT	(9/22)		2.82	0.509	NO. 1 C.	25
27	CLOSE BREACH	(10/23)	P	2.51	0.443	NO. 1 C.	26
28	CHECK AIM	(11/24)		4.81	0.143	C.SECT.	22 24
29	PRIME	25		4.40	2.302	NO. 1 C.	27 28
30	FIRE	26		5.59	2.661	NO. 1 C.	29
31	RECOIL	--		1.00	0.0	--	30
32	SWAB BORE	27		8.67	3.552	NO. 1 C.	31
33	END	--		0.0	0.0	--	32

---

[a] Task numbers in this column refer to empirical task numbers.

Table A2. Input data for FDC crew.

Task No.	Description	Data Task[a]	Task Type	Mean (sec)	SD (sec)[b]	Crewmember	Precursor Task(s)	
1	START	--		0.0	0.0	--	--	
2	CALL FOR FIRE	1		2.35	(0.547)	--	1	
3	RETURN CFF	2		2.12	0.915	FDO	2	
4	ID REG'N POINT	3		3.16	(1.080)	--	3	
5	RETURN REG'N PT	4		2.26	1.202	FDO	4	
6	AN: TARGT COORDS	5		6.21	(1.509)	--	5	
7	RETURN TGT CORDS	6	C	7.22	2.343	FDO	6	
8	DESCRIBES TARGET	7		4.06	(2.058)	--	7	
9	RETURN DESCRIPTN	8		2.03	1.085	FDO	8	
10	ISSUE FIRE ORDER	9	C	2.16	1.292	FDO	9	
11	MESSAGE TO OBS'R	10		1.98	1.217	FDO	10	
12	XMIT FIRE MIS'N	11		1.95	1.079	COMPUTER	2	
13	XBAK FIRE MIS'N	12		1.33	(0.476)	--	12	
14	XMIT FIRE ORDER	13		1.52	0.548	COMPUTER	10	12
15	XBAK FIRE ORDER	14		1.81	(0.859)	--	13	14
16	BEGIN PLOTTING	15A	C	16.14	3.051	HCO	4	
17	FINSH PLOT & AN:	15B	C	23.00	4.348	HCO	7	16
18	RETURN RANGE	16		2.28	0.963	COMPUTER	14	17
19	AN: DEFLECTION	17		6.39	1.717	HCO	18	
20	RETURN DEFL.	18		2.18	1.341	COMPUTER	19	
21	REQUEST SITE	19A	C	10.16	6.627	COMPUTER	20	
22	COMP & XMIT DEFL	19B	C	10.16	6.627	COMPUTER	25	
23	RETURN DEFL	20		1.46	(0.922)	--	15	22
24	COMP & AN: SITE	21A	C	17.64	6.315	FDO	11	17
25	REPLY WITH SITE	21B	C	1.00	0.359	FDO	21	24
26	COMP & XMIT QUAD	22		8.76	5.806	COMPUTER	23	
27	RETURN QUAD	23		1.17	(0.460)	--	26	
28	END	--		0.0	0.0	--	27	

---

[a] Task numbers in this column refer to empirical task numbers.

[b] Standard deviations indicated in parentheses ( ) were set to 0.0, making these fixed time tasks.

Table A3. Input data for tank crew.

Task No.	Description	Data Task[a]	Task Type	Mean (sec)	SD (sec)	Crewmember	Precursor Task(s)	
1	START	--		0.000	0.000	--	--	
2	ACQUIRE TARGET	--		0.100	0.050	T.CMDR.	1	
3	ACQUIRE TARGET	--		0.200	0.100	GUNNER	-2	1
4	REPORT TARGET	--		1.000	0.400	GUNNER	-2	3
5	SCAN REPORT AREA	--		1.500	1.000	T.CMDR.	-2	4
6	SEE TARGET	--		0.100	0.050	T.CMDR.	-2	5
7	TRAVERSE TURRET	1A		0.274	0.158	T.CMDR.	2	6
8	ISSUE COMMAND	1B		0.274	0.105	T.CMDR.	2	6
9	ARM - "UP"	2		1.916	0.807	LOADER	8	
10	SET COMP.	3A		0.500	0.298	GUNNER	8	
11	ACQUIRE TARGET	3B		0.731	0.360	GUNNER	7	10
12	SELECT ROUND	4		5.745	2.427	LOADER	9	
13	TRACK TARGET	5A		0.592	0.421	GUNNER	11	
14	LASER RANGE	5B		0.592	0.421	GUNNER	13	
15	CHECK RANGE	6		1.142	0.343	T.CMDR.	14	
16	RANGE OK?	--		0.000	0.000	T.CMDR.	15	
17	ORDER: RE-LASE	--		3.000	1.000	T.CMDR.	-16	15
18	RE-LASE	--		3.000	1.000	GUNNER	-16	17
19	ORDER: MOVE OUT	7A		0.365	0.086	T.CMDR.	16	18
20	ORDER: TAKE OVER	7B		0.400	0.090	T.CMDR.	19	
21	MOVE FORWARD	8A		3.603	0.354	DRIVER	19	
22	GUIDE DRIVER	9A		2.023	0.619	GUNNER	20	
23	ORDER: STOP	9B		0.300	0.238	GUNNER	21	22
24	STOP	8B		1.000	0.186	DRIVER	23	
25	COMMAND FIRE	10		2.127	0.721	T.CMDR.	24	
26	FINAL LAY	11		1.288	0.584	GUNNER	25.	
27	PULL TRIGGER	12		1.034	0.382	GUNNER	26	
28	RECOIL TIME	--		0.100	0.000	--	9	27
29	TRACK FLIGHT	14A		0.500	0.289	T.CMDR.	9	27
30	TRACK FLIGHT	--		0.500	0.289	GUNNER	9	27
31	REPORT IMPACT	14B		1.081	0.246	T.CMDR.	-32	29
32	REPORT IMPACT	--		1.500	0.500	GUNNER	-31	30
33	RELOAD	13A	P	5.682	2.712	LOADER	12	28
34	ARM - "UP"	13B		0.500	0.200	LOADER	33	
35	SELECT NEXT	4		5.745	2.427	LOADER	34	
36	ORDER: RE-ENGAGE	15		0.480	0.109	T.CMDR.	31	32
37	RE-LAY & LASE	16		0.968	0.205	GUNNER	36	
38	CHECK RANGE	17		1.601	0.873	T.CMDR.	37	
39	FINAL LAY	18		0.610	0.105	GUNNER	38	
40	PULL TRIGGER	19		0.747	0.158	GUNNER	39	
41	RECOIL TIME	--		0.100	0.000	--	19	34
42	TRACK FLIGHT	21A		0.500	0.289	T.CMDR.	19	34
43	TRACK FLIGHT	--		0.500	0.289	GUNNER	19	34
44	REPORT IMPACT	21B		0.912	0.242	T.CMDR.	42	
45	RELOAD	20	P	5.699	2.262	LOADER	35	41
46	"CEASE FIRE"	22		0.401	0.071	T.CMDR.	44	
47	"BACK UP"	23		0.453	0.060	T.CMDR.	46	
48	START BACK	24A		3.616	0.116	DRIVER	47	
49	"STOP"	24B		0.300	0.032	T.CMDR.	48	
50	STOP	24C		1.000	0.061	DRIVER	49	
51	END	--		0.000	0.000	--	50	

---

[a] Task numbers in this column refer to empirical task numbers.

Table A4. Input data for ITV-TOW crew.

Task No.	Description	Data Task[a]	Task Type	Mean (sec)	SD (sec)	Crewmember	Precursor Task(s)
1	START	--		0.0	0.0	--	--
2	ALERT/GIVE ORDER	1		5.32	0.029	SQD.LDR.	1
3	ERECT LAUNCHER	2		6.36	0.594	GUNNER	2
4	SLEW TURRET	3		11.29	1.845	GUNNER	3
5	AN: SELECT & ARM	4A		0.75	0.006	SQD.LDR.	4
6	ARM 1ST MISSILE	4B		1.76	0.009	GUNNER	5
7	AN: "FIRE"	5		2.02	0.597	SQD.LDR.	6
8	FIRE/TRACK MIS'L	6	P	19.37	0.843	GUNNER	7
9	GIVE 2ND ORDER	7		7.91	0.396	SQD.LDR.	8
10	TRAVERSE TURRET	8		4.80	1.679	GUNNER	9
11	AN: ARM MISSILE	9A		0.44	0.067	SQD.LDR.	10
12	ARM 2ND MISSILE	9B		1.02	0.103	GUNNER	11
13	AN: "FIRE"	10		1.80	0.763	SQD.LDR.	12
14	FIRE/TRACK MIS'L	11	P	18.13	0.961	GUNNER	13
15	AN: CEASE/RELOAD	12		2.17	0.639	SQD.LDR.	14
16	POSITION TURRET	13A		13.20	2.239	GUNNER	15
17	OPEN HATCH	13B		3.05	1.076	LOADER	16
18	UNLOAD LEFT TUBE	14	P	3.22	0.957	LOADER	17
19	LOAD LEFT TUBE	15	P	7.61	1.791	LOADER	18
20	UNLOAD RGHT TUBE	16	P	2.06	0.381	LOADER	19
21	LOAD RIGHT TUBE	17	P	8.12	1.281	LOADER	20
22	CLOSE HATCH/"UP"	18		3.24	0.981	LOADER	21
23	REARRANGE RACK	19	P	6.39	0.778	LOADER	22
24	END	--		0.0	0.0	--	23

---

[a] Task numbers in this column refer to empirical task numbers.

Table A5. Best-fit results for gun crew (undegraded mission time: 103.182 s).

Crewmember Performance				Mean	SD	Crew	Predicted	Fractional
G	AG	CS	L	(s)	(s)	Performance	by Fit	Difference
0.80	1.00	1.00	1.00	103.654	9.607	0.99545	0.98972	-0.00575
1.00	0.80	1.00	1.00	105.171	9.475	0.98109	0.97477	-0.00644
0.80	0.80	1.00	1.00	105.818	8.913	0.97509	0.96312	-0.01227
0.50	1.00	1.00	1.00	106.426	10.119	0.96952	0.95136	-0.01873
1.00	0.50	1.00	1.00	114.289	11.424	0.90282	0.90867	0.00648
0.80	0.50	1.00	1.00	112.799	11.392	0.91474	0.89854	-0.01771
0.50	0.80	1.00	1.00	110.656	12.036	0.93246	0.92676	-0.00611
0.50	0.50	1.00	1.00	116.808	12.745	0.88335	0.86681	-0.01872
0.30	1.00	1.00	1.00	119.403	17.929	0.86415	0.88254	0.02128
1.00	0.30	1.00	1.00	131.091	17.151	0.78710	0.82385	0.04669
0.80	0.30	1.00	1.00	128.766	14.542	0.80131	0.81552	0.01773
0.20	0.80	1.00	1.00	136.763	29.263	0.75446	0.78392	0.03905
0.50	0.30	1.00	1.00	133.613	15.137	0.77225	0.78929	0.02207
0.30	0.50	1.00	1.00	125.489	17.824	0.82224	0.80931	-0.01573
0.30	0.30	1.00	1.00	133.808	14.324	0.77112	0.74133	-0.03863
0.80	1.00	0.80	1.00	109.192	9.229	0.94496	0.93222	-0.01348
1.00	0.80	1.00	0.80	123.090	12.510	0.83826	0.84168	0.00407
0.80	0.80	0.80	0.80	129.511	11.256	0.79670	0.79187	-0.00607
0.50	1.00	0.50	1.00	133.467	11.907	0.77309	0.76884	-0.00550
1.00	0.50	1.00	0.50	183.749	17.943	0.56154	0.56138	-0.00028
0.80	0.50	0.80	0.50	188.181	15.897	0.54831	0.53878	-0.01739
0.50	0.80	0.50	0.80	150.590	13.356	0.68519	0.67078	-0.02102
0.50	0.50	0.50	0.50	206.819	19.175	0.49890	0.47984	-0.03820
0.30	1.00	0.30	1.00	179.919	16.619	0.57349	0.58273	0.01610
1.00	0.30	1.00	0.30	292.827	31.478	0.35237	0.34482	-0.02141
0.80	0.30	0.80	0.30	297.240	33.308	0.34713	0.33616	-0.03162
0.20	0.80	0.30	0.80	208.957	23.501	0.49380	0.49484	0.00212
0.50	0.30	0.50	0.30	311.130	31.322	0.33164	0.31223	-0.05851
0.30	0.50	0.30	0.50	242.088	18.255	0.42622	0.40009	-0.06129
0.30	0.30	0.30	0.30	349.141	33.313	0.29553	0.27638	-0.06479
1.00	1.00	0.80	1.00	109.788	9.748	0.93983	0.94313	0.00351
1.00	1.00	1.00	0.80	122.170	12.503	0.84458	0.86192	0.02053
1.00	1.00	0.80	0.80	126.276	12.861	0.81711	0.81798	0.00105
1.00	1.00	0.50	1.00	128.549	9.952	0.80267	0.80159	-0.00134
1.00	1.00	1.00	0.50	181.440	19.213	0.56868	0.59566	0.04744
1.00	1.00	0.80	0.50	183.383	19.604	0.56266	0.57434	0.02077
1.00	1.00	0.50	0.80	145.418	12.698	0.70955	0.70935	-0.00029
1.00	1.00	0.50	0.50	201.946	18.327	0.51094	0.51858	0.01496
1.00	1.00	0.30	1.00	168.178	9.884	0.61353	0.63253	0.03097
1.00	1.00	1.00	0.30	281.705	28.332	0.36628	0.37254	0.01711
1.00	1.00	0.80	0.30	301.321	32.496	0.34243	0.36409	0.06325
1.00	1.00	0.30	0.80	183.812	11.908	0.56135	0.57366	0.02195
1.00	1.00	0.50	0.30	306.564	32.007	0.33658	0.34086	0.01272
1.00	1.00	0.30	0.50	230.123	20.680	0.44838	0.44213	-0.01393
1.00	1.00	0.30	0.30	333.926	30.359	0.30900	0.30607	-0.00947

Table A6. Best-fit results for FDC crew.
(undegraded mission time: 93.375 s).

Crewmember Performance			Mean	SD	Crew	Predicted	Fractional
F	C	H	(s)	(s)	Performance	by Fit	Difference
0.80	1.00	1.00	98.055	13.402	0.95227	0.95680	0.00476
0.50	1.00	1.00	116.395	15.013	0.80223	0.83152	0.03651
0.30	1.00	1.00	156.814	24.770	0.59545	0.63306	0.06316
1.00	0.80	1.00	100.629	13.182	0.92791	0.91571	-0.01315
0.80	0.80	1.00	102.055	14.023	0.91495	0.88168	-0.03636
0.50	0.80	1.00	123.434	18.138	0.75648	0.77419	0.02342
0.30	0.80	1.00	165.834	25.010	0.56306	0.59928	0.06431
1.00	0.50	1.00	124.391	25.575	0.75066	0.74533	-0.00710
0.80	0.50	1.00	125.777	22.691	0.74239	0.72263	-0.02661
0.50	0.50	1.00	141.272	25.788	0.66096	0.64880	-0.01840
0.30	0.50	1.00	180.216	33.306	0.51813	0.52129	0.00610
1.00	0.30	1.00	167.693	31.789	0.55682	0.57271	0.02853
0.80	0.30	1.00	171.653	31.460	0.54398	0.55921	0.02801
0.50	0.30	1.00	184.069	36.961	0.50728	0.51395	0.01315
0.30	0.30	1.00	213.197	38.895	0.43798	0.43053	-0.01700
1.00	1.00	0.80	102.417	11.930	0.91171	0.89700	-0.01614
0.80	1.00	0.80	107.952	12.956	0.86497	0.86433	-0.00074
0.50	1.00	0.80	123.335	17.658	0.75708	0.76078	0.00488
0.30	1.00	0.80	162.743	24.930	0.57376	0.59121	0.03041
1.00	0.80	0.80	115.333	17.410	0.80961	0.83065	0.02599
0.80	0.80	0.80	117.738	16.379	0.79307	0.80256	0.01196
0.50	0.80	0.80	131.961	16.900	0.70760	0.71251	0.00695
0.30	0.80	0.80	168.845	25.147	0.55302	0.56164	0.01558
1.00	0.50	0.80	137.610	21.255	0.67855	0.68799	0.01391
0.80	0.50	0.80	136.008	18.057	0.68654	0.66860	-0.02613
0.50	0.50	0.80	144.619	22.038	0.64566	0.60491	-0.06311
0.30	0.50	0.80	186.496	30.659	0.50068	0.49258	-0.01619
1.00	0.30	0.80	176.535	32.935	0.52893	0.53824	0.01760
0.80	0.30	0.80	184.951	36.285	0.50486	0.52630	0.04246
0.50	0.30	0.80	189.537	36.164	0.49265	0.48602	-0.01345
0.30	0.30	0.80	221.583	42.459	0.42140	0.41076	-0.02526
1.00	1.00	0.50	133.984	15.341	0.69691	0.69054	-0.00914
0.80	1.00	0.50	138.912	14.346	0.67219	0.67102	-0.00174
0.50	1.00	0.50	150.724	15.297	0.61951	0.60689	-0.02038
0.30	1.00	0.50	184.610	25.399	0.50580	0.49388	-0.02355
1.00	0.80	0.50	145.915	16.508	0.63993	0.65054	0.01659
0.80	0.80	0.50	145.589	18.634	0.64136	0.63318	-0.01275
0.50	0.80	0.50	156.978	20.520	0.59483	0.57577	-0.03204
0.30	0.80	0.50	189.092	24.921	0.49381	0.47308	-0.04198
1.00	0.50	0.50	171.036	23.184	0.54594	0.55965	0.02512
0.80	0.50	0.50	167.775	21.257	0.55655	0.54676	-0.01759
0.50	0.50	0.50	177.242	24.738	0.52682	0.50341	-0.04443
0.30	0.50	0.50	202.601	27.660	0.46088	0.42311	-0.08195
1.00	0.30	0.50	212.758	41.553	0.43888	0.45637	0.03985
0.80	0.30	0.50	209.611	33.765	0.44547	0.44776	0.00514
0.50	0.30	0.50	220.650	37.541	0.42318	0.41826	-0.01162
0.30	0.30	0.50	237.129	40.228	0.39377	0.36129	-0.08249
1.00	1.00	0.30	199.124	21.458	0.46893	0.49128	0.04767
0.80	1.00	0.30	202.856	22.735	0.46030	0.48132	0.04566
0.50	1.00	0.30	208.158	23.009	0.44858	0.44741	-0.00261
0.30	1.00	0.30	225.230	26.479	0.41458	0.38283	-0.07657
1.00	0.80	0.30	205.803	22.368	0.45371	0.47069	0.03743
0.80	0.80	0.30	207.053	21.473	0.45097	0.46154	0.02343
0.50	0.80	0.30	214.312	27.083	0.43570	0.43027	-0.01246
0.30	0.80	0.30	239.124	29.276	0.39049	0.37021	-0.05192
1.00	0.50	0.30	227.842	27.096	0.40982	0.42120	0.02776
0.80	0.50	0.30	231.984	31.477	0.40251	0.41385	0.02819
0.50	0.50	0.30	238.886	27.625	0.39088	0.38853	-0.00600
0.30	0.50	0.30	254.987	36.194	0.36620	0.33889	-0.07456
1.00	0.30	0.30	270.894	34.346	0.34469	0.35990	0.04412
0.80	0.30	0.30	281.183	44.969	0.33208	0.35452	0.06758
0.50	0.30	0.30	279.526	45.571	0.33405	0.33578	0.00517
0.30	0.30	0.30	290.646	39.646	0.32127	0.29805	-0.07228

Table A7. Best-fit results for tank crew (undegraded mission time: 28.062 s).

Crewmember Performance				Mean	SD	Crew	Predicted	Fractional
TC	G	L	D	(s)	(s)	Performance	by Fit	Difference
0.80	1.00	1.00	1.00	29.269	3.447	0.95876	0.97058	0.01232
1.00	0.80	1.00	1.00	28.129	3.209	0.99762	0.97807	-0.01960
0.80	0.80	1.00	1.00	30.154	2.902	0.93062	0.92338	-0.00778
0.50	1.00	1.00	1.00	33.957	3.489	0.82640	0.82517	-0.00149
1.00	0.50	1.00	1.00	33.567	3.986	0.83600	0.84226	0.00749
0.80	0.50	1.00	1.00	35.255	3.659	0.79597	0.80139	0.00680
0.50	0.80	1.00	1.00	35.574	3.889	0.78883	0.79080	0.00250
0.50	0.50	1.00	1.00	40.308	4.510	0.69619	0.69960	0.00489
0.30	1.00	1.00	1.00	42.989	4.989	0.65277	0.65170	-0.00164
1.00	0.30	1.00	1.00	41.624	6.690	0.67418	0.66757	-0.00981
0.80	0.30	1.00	1.00	44.522	6.196	0.63030	0.64163	0.01799
0.20	0.80	1.00	1.00	56.800	6.425	0.49405	0.50250	0.01710
0.50	0.30	1.00	1.00	49.791	6.321	0.56360	0.57468	0.01967
0.30	0.50	1.00	1.00	50.158	5.374	0.55947	0.57079	0.02023
0.30	0.30	1.00	1.00	59.920	7.150	0.46832	0.48481	0.03521
0.80	1.00	0.80	1.00	31.113	5.100	0.90194	0.90350	0.00173
1.00	0.80	1.00	0.80	30.269	2.751	0.92709	0.90876	-0.01977
0.80	0.80	0.80	0.80	34.007	4.068	0.82518	0.80812	-0.02068
0.50	1.00	0.50	1.00	43.562	10.171	0.64419	0.64829	0.00637
1.00	0.50	1.00	0.50	41.343	4.524	0.67876	0.66685	-0.01754
0.80	0.50	0.80	0.50	44.862	6.059	0.62552	0.61101	-0.02319
0.50	0.80	0.50	0.80	44.430	8.751	0.63160	0.59767	-0.05372
0.50	0.50	0.50	0.50	53.471	8.504	0.52481	0.48255	-0.08053
0.30	1.00	0.30	1.00	64.747	15.327	0.43341	0.41940	-0.03232
1.00	0.30	1.00	0.30	60.571	6.247	0.46329	0.44888	-0.03112
0.80	0.30	0.80	0.30	64.655	7.502	0.43403	0.42286	-0.02572
0.20	0.80	0.30	0.80	75.645	16.959	0.37097	0.34271	-0.07619
0.50	0.30	0.50	0.30	73.515	8.361	0.38172	0.35707	-0.06456
0.30	0.50	0.30	0.50	74.043	12.731	0.37900	0.34315	-0.09458
0.30	0.30	0.30	0.30	92.668	14.796	0.30282	0.27455	-0.09338
1.00	1.00	0.80	1.00	29.669	3.953	0.94584	0.95579	0.01052
1.00	1.00	1.00	0.80	29.127	2.757	0.96344	0.95443	-0.00935
1.00	1.00	0.80	0.80	31.219	4.046	0.89888	0.88949	-0.01044
1.00	1.00	0.50	1.00	37.789	8.959	0.74260	0.76899	0.03554
1.00	1.00	1.00	0.50	36.257	5.445	0.77397	0.77999	0.00778
1.00	1.00	0.80	0.50	38.315	4.906	0.73240	0.73608	0.00501
1.00	1.00	0.50	0.80	39.167	10.617	0.71647	0.72548	0.01258
1.00	1.00	0.50	0.50	44.562	7.492	0.62973	0.62008	-0.01533
1.00	1.00	0.30	1.00	53.948	16.740	0.52017	0.54955	0.05649
1.00	1.00	1.00	0.30	48.399	4.738	0.57981	0.58838	0.01479
1.00	1.00	0.80	0.30	50.247	5.262	0.55848	0.56304	0.00816
1.00	1.00	0.30	0.80	52.559	12.942	0.53391	0.52697	-0.01301
1.00	1.00	0.50	0.30	55.807	8.206	0.50284	0.49256	-0.02045
1.00	1.00	0.30	0.50	60.075	16.473	0.46712	0.46905	0.00414
1.00	1.00	0.30	0.30	68.389	15.890	0.41033	0.39224	-0.04409
0.90	1.00	1.00	1.00	27.814	3.002	1.00892	1.00000	-0.00884
0.80	1.00	1.00	1.00	28.817	3.499	0.97380	0.97058	-0.00331
0.70	1.00	1.00	1.00	30.011	3.638	0.93506	0.93148	-0.00382
0.60	1.00	1.00	1.00	32.005	5.784	0.87680	0.88402	0.00823
0.50	1.00	1.00	1.00	34.209	3.629	0.82031	0.82517	0.00592
0.40	1.00	1.00	1.00	37.263	3.822	0.75308	0.75026	-0.00374
0.30	1.00	1.00	1.00	42.925	4.607	0.65374	0.65170	-0.00312
0.20	1.00	1.00	1.00	54.915	6.985	0.51101	0.51615	0.01007
0.10	1.00	1.00	1.00	91.753	14.124	0.30584	0.31793	0.03952
1.00	0.90	1.00	1.00	27.957	3.207	1.00376	1.00000	-0.00374
1.00	0.80	1.00	1.00	28.319	2.977	0.99092	0.97807	-0.01297
1.00	0.70	1.00	1.00	29.723	3.112	0.94412	0.94266	-0.00154
1.00	0.60	1.00	1.00	31.247	3.455	0.89807	0.89855	0.00054
1.00	0.50	1.00	1.00	33.858	4.465	0.82881	0.84226	0.01622
1.00	0.40	1.00	1.00	36.448	4.839	0.76992	0.76830	-0.00211
1.00	0.30	1.00	1.00	43.332	7.335	0.64760	0.66757	0.03083
1.00	0.20	1.00	1.00	52.941	8.036	0.53006	0.52424	-0.01098
1.00	0.10	1.00	1.00	91.170	14.283	0.30780	0.31029	0.00809
1.00	1.00	0.90	1.00	28.505	4.193	0.98446	0.99684	0.01258
1.00	1.00	0.80	1.00	28.947	4.132	0.96943	0.95579	-0.01407
1.00	1.00	0.70	1.00	30.769	4.567	0.91202	0.90607	-0.00652
1.00	1.00	0.60	1.00	33.068	6.157	0.84862	0.84504	-0.00421
1.00	1.00	0.50	1.00	35.749	8.215	0.78497	0.76899	-0.02036
1.00	1.00	0.40	1.00	41.624	9.507	0.67418	0.67278	-0.00207
1.00	1.00	0.30	1.00	55.140	16.834	0.50892	0.54955	0.07983
1.00	1.00	0.20	1.00	76.463	22.704	0.36700	0.39120	0.06595
1.00	1.00	0.10	1.00	132.740	44.313	0.21141	0.19391	-0.08274
1.00	1.00	1.00	0.90	28.310	3.580	0.99124	0.99561	0.00441
1.00	1.00	1.00	0.80	29.879	3.262	0.93919	0.95443	0.01623
1.00	1.00	1.00	0.70	31.282	3.160	0.89707	0.90622	0.01020
1.00	1.00	1.00	0.60	33.192	3.494	0.84544	0.84900	0.00420
1.00	1.00	1.00	0.50	36.114	3.504	0.77704	0.77999	0.00380
1.00	1.00	1.00	0.40	41.305	4.185	0.67939	0.69516	0.02322
1.00	1.00	1.00	0.30	49.094	4.714	0.57160	0.58838	0.02936
1.00	1.00	1.00	0.20	64.482	7.109	0.43519	0.44993	0.03387
1.00	1.00	1.00	0.10	109.445	11.716	0.25640	0.26346	0.02753

Table A8. Best-fit results for ITV-TOW crew (undegraded mission time: 157.824 s).

Crewmember Performance				Mean	SD	Crew	Predicted	Fractional
SL	G	D[a]	L	(s)	(s)	Performance	by Fit	Difference
0.80	1.00	1.00	1.00	162.926	4.935	0.96869	0.96960	0.00094
0.50	1.00	1.00	1.00	177.587	5.522	0.88871	0.88690	-0.00204
0.30	1.00	1.00	1.00	205.723	5.816	0.76717	0.76817	0.00131
1.00	0.80	1.00	1.00	176.789	5.384	0.89273	0.89306	0.00037
0.80	0.80	1.00	1.00	180.935	5.678	0.87227	0.86854	-0.00428
0.50	0.80	1.00	1.00	196.637	5.682	0.80262	0.80159	-0.00128
0.30	0.80	1.00	1.00	223.738	7.682	0.70540	0.70334	-0.00292
1.00	0.50	1.00	1.00	234.026	8.866	0.67439	0.67497	0.00086
0.80	0.50	1.00	1.00	239.131	8.594	0.65999	0.66087	0.00133
0.50	0.50	1.00	1.00	254.489	9.256	0.62016	0.62138	0.00197
0.30	0.50	1.00	1.00	281.568	10.536	0.56052	0.56067	0.00026
1.00	0.30	1.00	1.00	337.737	14.396	0.46730	0.46968	0.00509
0.80	0.30	1.00	1.00	341.848	14.035	0.46168	0.46281	0.00245
0.50	0.30	1.00	1.00	353.790	15.868	0.44610	0.44309	-0.00674
0.30	0.30	1.00	1.00	384.025	15.719	0.41097	0.41133	0.00086
1.00	1.00	1.00	0.80	166.522	5.424	0.94777	0.94837	0.00064
0.80	1.00	1.00	0.80	171.849	5.230	0.91839	0.92077	0.00260
0.50	1.00	1.00	0.80	186.606	6.219	0.84576	0.84588	0.00014
0.30	1.00	1.00	0.80	213.642	6.392	0.73873	0.73720	-0.00207
1.00	0.80	1.00	0.80	184.608	6.093	0.85491	0.85147	-0.00403
0.80	0.80	1.00	0.80	190.765	5.519	0.82732	0.82915	0.00221
0.50	0.80	1.00	0.80	205.345	6.459	0.76858	0.76792	-0.00085
0.30	0.80	1.00	0.80	234.372	6.348	0.67339	0.67728	0.00578
1.00	0.50	1.00	0.80	242.031	9.018	0.65208	0.65094	-0.00175
0.80	0.50	1.00	0.80	248.679	8.452	0.63465	0.63782	0.00499
0.50	0.50	1.00	0.80	264.326	9.542	0.59708	0.60096	0.00649
0.30	0.50	1.00	0.80	290.175	8.912	0.54389	0.54398	0.00017
1.00	0.30	1.00	0.80	345.089	14.501	0.45734	0.45792	0.00125
0.80	0.30	1.00	0.80	349.241	15.161	0.45191	0.45138	-0.00116
0.50	0.30	1.00	0.80	366.428	16.578	0.43071	0.43261	0.00440
0.30	0.30	1.00	0.80	392.880	15.756	0.40171	0.40228	0.00141
1.00	1.00	1.00	0.50	192.296	7.700	0.82073	0.82218	0.00177
0.80	1.00	1.00	0.50	196.299	6.867	0.80400	0.80136	-0.00328
0.50	1.00	1.00	0.50	212.653	7.363	0.74217	0.74402	0.00250
0.30	1.00	1.00	0.50	239.250	7.978	0.65966	0.65862	-0.00157
1.00	0.80	1.00	0.50	210.520	7.340	0.74969	0.74835	-0.00179
0.80	0.80	1.00	0.50	215.862	8.208	0.73113	0.73105	-0.00011
0.50	0.80	1.00	0.50	231.317	8.797	0.68228	0.68304	0.00110
0.30	0.80	1.00	0.50	258.199	9.577	0.61125	0.61038	-0.00142
1.00	0.50	1.00	0.50	267.275	9.079	0.59049	0.58890	-0.00270
0.80	0.50	1.00	0.50	273.892	9.744	0.57623	0.57814	0.00332
0.50	0.50	1.00	0.50	288.085	10.372	0.54784	0.54769	-0.00027
0.30	0.50	1.00	0.50	314.944	10.910	0.50112	0.49997	-0.00229
1.00	0.30	1.00	0.50	369.950	14.509	0.42661	0.42632	-0.00067
0.80	0.30	1.00	0.50	370.212	14.026	0.42631	0.42065	-0.01326
0.50	0.30	1.00	0.50	391.071	14.022	0.40357	0.40430	0.00181
0.30	0.30	1.00	0.50	417.531	16.399	0.37799	0.37769	-0.00081
1.00	1.00	1.00	0.30	238.261	10.797	0.66240	0.66727	0.00736
0.80	1.00	1.00	0.30	240.860	11.032	0.65525	0.65349	-0.00269
0.50	1.00	1.00	0.30	256.597	10.614	0.61507	0.61485	-0.00035
0.30	1.00	1.00	0.30	283.046	12.202	0.55759	0.55534	-0.00403
1.00	0.80	1.00	0.30	255.373	9.775	0.61801	0.61780	-0.00034
0.80	0.80	1.00	0.30	259.813	8.873	0.60745	0.60597	-0.00244
0.50	0.80	1.00	0.30	276.243	9.695	0.57132	0.57260	0.00224
0.30	0.80	1.00	0.30	304.458	12.528	0.51838	0.52065	0.00438
1.00	0.50	1.00	0.30	311.031	10.873	0.50742	0.50494	-0.00490
0.80	0.50	1.00	0.30	318.653	11.660	0.49528	0.49701	0.00347
0.50	0.50	1.00	0.30	331.547	11.507	0.47602	0.47433	-0.00355
0.30	0.50	1.00	0.30	360.829	12.542	0.43739	0.43812	0.00166
1.00	0.30	1.00	0.30	413.889	16.507	0.38132	0.38052	-0.00211
0.80	0.30	1.00	0.30	417.046	15.533	0.37843	0.37599	-0.00645
0.50	0.30	1.00	0.30	433.200	17.662	0.36432	0.36287	-0.00397
0.30	0.30	1.00	0.30	465.502	18.800	0.33904	0.34129	0.00664

---

[a] No variation in driver performance.

Table A9. Performance data grids for gun crew.

Grid Line	Dose (rads)	0.20	0.25	0.32	0.40	0.50	0.63	0.79	1.00	1.26	1.58	2.00	2.51	3.16
		Time Grid Lines (h)												
18	4529.9	1.0000	1.0000	0.8081	0.6118	0.4845	0.3462	0.2513	0.1772	0.1198	0.0765	0.0435	0.0186	0.0012
17	3516.3	1.0000	1.0000	1.0000	0.7285	0.6682	0.5583	0.4173	0.2870	0.2036	0.1706	0.1388	0.1065	0.0845
16	2729.6	1.0000	1.0000	1.0000	0.9820	0.7538	0.6248	0.5411	0.4348	0.3383	0.2832	0.2181	0.1482	0.1278
15	2118.8	1.0000	1.0000	1.0000	1.0000	0.8425	0.6814	0.6453	0.5558	0.4558	0.3863	0.2838	0.1830	0.1727
14	1644.7	1.0000	1.0000	1.0000	1.0000	1.0000	0.8208	0.6781	0.6102	0.5148	0.4307	0.3353	0.2590	0.2448
13	1276.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7464	0.6884	0.5745	0.4800	0.3808	0.3355	0.3154
12	991.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8112	0.6748	0.5983	0.5095	0.4208	0.3683
11	769.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8088	0.6858	0.6188	0.5616	0.4817
10	597.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8000	0.7261	0.6689	0.5901
9	463.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8337	0.7200	0.6753
8	359.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9810	0.8203	0.7545
7	279.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9790	0.8568
6	216.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9655
5	168.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	130.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	101.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Grid Line	Dose (rads)	3.98	5.01	6.31	7.94	10.00	12.59	15.85	19.95	25.12	31.62	39.81	50.12	63.10
		Time Grid Lines (h)												
18	4529.9	0.0008	0.0004	0.0005	0.0004	0.0004	0.0003	0.0003	0.0004	0.0004	0.0003	0.0004	0.0003	0.0001
17	3516.3	0.0838	0.0407	0.0368	0.0306	0.0282	0.0224	0.0232	0.0252	0.0278	0.0253	0.0313	0.0273	0.0177
16	2729.6	0.1068	0.0824	0.0635	0.0428	0.0351	0.0313	0.0313	0.0345	0.0411	0.0461	0.0604	0.0603	0.0474
15	2118.8	0.1521	0.1285	0.0814	0.0557	0.0443	0.0408	0.0388	0.0441	0.0550	0.0723	0.0814	0.1002	0.1024
14	1644.7	0.2270	0.1822	0.1280	0.0812	0.0681	0.0658	0.0671	0.0757	0.0882	0.1307	0.1541	0.1886	0.2239
13	1276.7	0.3000	0.2375	0.1827	0.1288	0.1078	0.1017	0.1075	0.1183	0.1334	0.1861	0.2144	0.2500	0.3457
12	991.0	0.3367	0.3081	0.2665	0.2215	0.1883	0.1480	0.1602	0.1723	0.1846	0.2036	0.2183	0.3388	0.4299
11	769.3	0.4207	0.3833	0.3510	0.3154	0.2742	0.2517	0.2664	0.2814	0.2961	0.3113	0.3513	0.4502	0.5173
10	597.2	0.5128	0.4637	0.4311	0.4042	0.3721	0.3467	0.3548	0.3687	0.3858	0.4035	0.4652	0.5541	0.5832
9	463.5	0.6061	0.5482	0.5082	0.4818	0.4512	0.4230	0.4098	0.4157	0.4343	0.4552	0.5250	0.6265	0.6470
8	359.8	0.6820	0.6285	0.5800	0.5503	0.5240	0.4883	0.4767	0.4788	0.4864	0.5168	0.5839	0.6792	0.6827
7	279.3	0.7591	0.7112	0.6615	0.6301	0.6087	0.6004	0.5886	0.6087	0.6220	0.6366	0.6736	0.7367	0.7469
6	216.8	0.8342	0.7881	0.7515	0.7182	0.7038	0.7174	0.7367	0.7544	0.7652	0.7754	0.7864	0.7998	0.8068
5	168.3	0.9626	0.9090	0.8668	0.8180	0.7875	0.8277	0.8525	0.8712	0.8973	0.9087	0.9186	0.9272	0.9347
4	130.6	1.0000	0.9764	0.9408	0.8868	0.8751	0.9075	0.9328	0.9518	0.9755	0.9864	0.9853	1.0000	1.0000
3	101.4	1.0000	1.0000	0.9854	0.9584	0.9440	0.9638	0.9861	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	0.9964	0.9950	0.9869	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Grid Line	Dose (rads)	79.43	100.00	125.89	158.49	199.53	251.19	316.23	398.11	501.19	630.96	794.33	1000.00	1258.93
		Time Grid Lines (h)												
18	4529.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
17	3516.3	0.0131	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
16	2729.6	0.0521	0.0138	0.0014	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15	2118.8	0.0958	0.0649	0.0274	0.0153	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14	1644.7	0.2429	0.1603	0.0869	0.0508	0.0142	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
13	1276.7	0.3938	0.2905	0.1972	0.1110	0.0703	0.0360	0.0110	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
12	991.0	0.5245	0.4441	0.3742	0.3157	0.2028	0.1431	0.0782	0.0404	0.0000	0.0000	0.0000	0.0000	0.0000
11	769.3	0.5838	0.5670	0.5424	0.5132	0.4331	0.3810	0.2810	0.2334	0.2245	0.0279	0.0000	0.0000	0.0000
10	597.2	0.6145	0.6276	0.6330	0.6247	0.5873	0.5440	0.4844	0.4318	0.3452	0.2098	0.1383	0.0000	0.0000
9	463.5	0.6539	0.6582	0.6598	0.6560	0.6442	0.6204	0.5809	0.5134	0.4585	0.3888	0.3068	0.1910	0.1013
8	359.8	0.6908	0.6878	0.6866	0.6854	0.6821	0.6788	0.6530	0.5881	0.5640	0.5310	0.4653	0.4019	0.3311
7	279.3	0.7401	0.7352	0.7339	0.7344	0.7362	0.7434	0.7295	0.6798	0.6749	0.6648	0.6254	0.5982	0.5610
6	216.8	0.8090	0.8125	0.8147	0.8142	0.8110	0.8111	0.8068	0.7984	0.7919	0.7885	0.7763	0.7729	0.7652
5	168.3	0.9414	0.9474	0.9527	0.9572	0.9612	0.9654	0.9698	0.9743	0.9781	0.9840	0.9893	0.9952	1.0000
4	130.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	101.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table A10. Performance data grids for FDC crew.

Grid Line	Dose (rads)	Time Grid Lines (h)												
		0.20	0.25	0.32	0.40	0.50	0.63	0.79	1.00	1.26	1.58	2.00	2.51	3.16
18	4529.9	1.0000	1.0000	0.8177	0.6337	0.4901	0.3737	0.2782	0.2023	0.1386	0.0907	0.0521	0.0222	0.0021
17	3516.3	1.0000	1.0000	1.0000	0.7988	0.7462	0.6863	0.6548	0.5855	0.5201	0.4875	0.4378	0.3552	0.2786
16	2729.6	1.0000	1.0000	1.0000	0.9836	0.8107	0.7343	0.7121	0.6577	0.5978	0.5880	0.5360	0.4388	0.3987
15	2118.8	1.0000	1.0000	1.0000	1.0000	0.8828	0.7706	0.7409	0.6987	0.6462	0.6282	0.5734	0.4981	0.4800
14	1644.7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9371	0.7859	0.7281	0.6881	0.6523	0.6008	0.5497	0.5491
13	1276.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8144	0.7571	0.7251	0.6767	0.6307	0.5980	0.5956
12	991.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8808	0.7607	0.7362	0.6937	0.6478	0.6214
11	769.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8478	0.7857	0.7362	0.6976	0.6619
10	597.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8507	0.7844	0.7434	0.7091
9	463.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8535	0.7871	0.7642
8	359.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9891	0.8835	0.8182
7	279.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9799	0.8905
6	216.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9706
5	168.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	130.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	101.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Grid Line	Dose (rads)	Time Grid Lines (h)												
		3.98	5.01	6.31	7.94	10.00	12.59	15.85	19.95	25.12	31.62	39.81	50.12	63.10
18	4529.9	0.0012	0.0005	0.0010	0.0009	0.0007	0.0008	0.0008	0.0007	0.0007	0.0005	0.0007	0.0004	0.0001
17	3516.3	0.2154	0.1437	0.1788	0.1487	0.1287	0.1120	0.1149	0.1231	0.1320	0.1185	0.1385	0.1082	0.0526
16	2729.6	0.3386	0.2966	0.2888	0.1886	0.1673	0.1519	0.1583	0.1711	0.1910	0.2076	0.2409	0.2227	0.1686
15	2118.8	0.4813	0.4194	0.3431	0.2467	0.2046	0.1616	0.1980	0.2194	0.2491	0.2821	0.3385	0.3363	0.3110
14	1644.7	0.5478	0.5109	0.4128	0.3158	0.2785	0.2718	0.2831	0.3144	0.3498	0.4070	0.4521	0.4631	0.4841
13	1276.7	0.6020	0.5723	0.4823	0.4038	0.3828	0.3528	0.3884	0.4014	0.4417	0.5087	0.5542	0.5814	0.6446
12	991.0	0.6147	0.6045	0.5682	0.5188	0.4502	0.4278	0.4518	0.4774	0.5033	0.5311	0.5589	0.6420	0.7146
11	769.3	0.6349	0.6278	0.6130	0.5800	0.5465	0.5272	0.5498	0.5722	0.5932	0.6137	0.6434	0.7031	0.7501
10	597.2	0.6739	0.6610	0.6546	0.6453	0.6329	0.6143	0.6240	0.6418	0.6611	0.6807	0.7182	0.7602	0.7707
9	463.5	0.7330	0.7111	0.7058	0.6988	0.6800	0.6718	0.6848	0.6734	0.6934	0.7152	0.7616	0.8068	0.8113
8	359.8	0.7819	0.7558	0.7437	0.7350	0.7282	0.7189	0.7080	0.7131	0.7312	0.7513	0.7989	0.8342	0.8344
7	279.3	0.8287	0.8050	0.7835	0.7715	0.7707	0.7749	0.7754	0.7865	0.7992	0.8119	0.8424	0.8596	0.8601
6	216.8	0.8782	0.8582	0.8287	0.8182	0.8189	0.8329	0.8481	0.8644	0.8738	0.8806	0.8862	0.8892	0.8896
5	168.3	0.9884	0.9316	0.9049	0.8755	0.8886	0.8901	0.9088	0.9235	0.9418	0.9442	0.9528	0.9564	0.9594
4	130.6	1.0000	0.9791	0.9548	0.9280	0.9128	0.9348	0.9529	0.9676	0.9830	0.9889	0.9937	1.0000	1.0000
3	101.4	1.0000	1.0000	0.9855	0.9870	0.9561	0.9702	0.9861	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	0.9937	0.9926	0.9941	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Grid Line	Dose (rads)	Time Grid Lines (h)												
		79.43	100.00	125.89	158.49	199.53	251.19	316.23	398.11	501.19	630.96	794.33	1000.00	1258.93
18	4529.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0030	0.0000
17	3516.3	0.0272	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
16	2729.6	0.1388	0.0278	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15	2118.8	0.2828	0.1489	0.0525	0.0357	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14	1644.7	0.4778	0.3483	0.2018	0.1483	0.0447	0.0014	0.0000	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000
13	1276.7	0.6768	0.5352	0.4001	0.3168	0.2244	0.1233	0.0234	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
12	991.0	0.7708	0.6898	0.6065	0.5658	0.4845	0.4183	0.3273	0.2123	0.0000	0.0000	0.0000	0.0000	0.0000
11	769.3	0.7812	0.7711	0.7452	0.7306	0.6903	0.6565	0.6135	0.5806	0.5325	0.0794	0.0000	0.0000	0.0000
10	597.2	0.7917	0.8004	0.8022	0.7988	0.7828	0.7828	0.7680	0.7438	0.6788	0.5189	0.4301	0.0001	0.0001
9	463.5	0.8142	0.8172	0.8177	0.8168	0.8141	0.8111	0.8015	0.7813	0.7585	0.7122	0.6269	0.3317	0.1003
8	359.8	0.8341	0.8323	0.8316	0.8324	0.8318	0.8320	0.8274	0.8112	0.8040	0.7892	0.7240	0.5902	0.3412
7	279.3	0.8594	0.8549	0.8541	0.8550	0.8558	0.8549	0.8568	0.8456	0.8459	0.8383	0.8011	0.7173	0.5741
6	216.8	0.8894	0.8881	0.8879	0.8880	0.8883	0.8887	0.8895	0.8871	0.8884	0.8864	0.8719	0.8301	0.7739
5	168.3	0.9622	0.9649	0.9678	0.9704	0.9731	0.9759	0.9787	0.9816	0.9845	0.9873	0.9897	0.9923	1.0000
4	130.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	101.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table A11. Performance data grids for tank crew.

Grid Line	Dose (rads)	Time Grid Lines (h) 0.20	0.25	0.32	0.40	0.50	0.63	0.79	1.00	1.26	1.58	2.00	2.51	3.16
18	4529.9	1.0000	1.0000	0.8224	0.6220	0.4685	0.3489	0.2527	0.1779	0.1219	0.0783	0.0463	0.0205	0.0023
17	3516.3	1.0000	1.0000	1.0000	0.8837	0.7958	0.7020	0.5878	0.4338	0.3781	0.3515	0.2883	0.2282	0.1730
16	2729.6	1.0000	1.0000	1.0000	1.0000	0.8647	0.7883	0.7135	0.5883	0.5015	0.4551	0.3552	0.2808	0.2759
15	2118.8	1.0000	1.0000	1.0000	1.0000	0.9348	0.8388	0.7808	0.7005	0.5885	0.5222	0.4083	0.3503	0.3488
14	1644.7	1.0000	1.0000	1.0000	1.0000	1.0000	0.8785	0.8267	0.7588	0.6638	0.5782	0.4821	0.4140	0.4081
13	1276.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8865	0.8134	0.7374	0.6250	0.5283	0.4618	0.4378
12	991.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9157	0.8235	0.7382	0.6682	0.5863	0.5021
11	769.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9177	0.8186	0.7738	0.6981	0.6188
10	597.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9086	0.8538	0.8000	0.7243
9	463.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9252	0.8487	0.8051
8	359.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9178	0.8782
7	279.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9443
6	216.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9862
5	168.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	130.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	101.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Grid Line	Dose (rads)	Time Grid Lines (h) 3.98	5.01	6.31	7.94	10.00	12.59	15.85	19.95	25.12	31.62	39.81	50.12	63.10
18	4529.9	0.0014	0.0007	0.0010	0.0008	0.0008	0.0006	0.0006	0.0006	0.0006	0.0004	0.0004	0.0002	0.0000
17	3516.3	0.1253	0.0701	0.0688	0.0688	0.0587	0.0484	0.0488	0.0487	0.0502	0.0433	0.0448	0.0330	0.0183
16	2729.6	0.2485	0.1817	0.1557	0.0884	0.0787	0.0708	0.0754	0.0812	0.0838	0.0881	0.0885	0.0868	0.0883
15	2118.8	0.3430	0.2841	0.2128	0.1298	0.0883	0.0825	0.1032	0.1142	0.1184	0.1318	0.1528	0.1858	0.1882
14	1644.7	0.4010	0.3874	0.2882	0.1713	0.1418	0.1374	0.1484	0.1708	0.1858	0.2243	0.2577	0.2873	0.3533
13	1276.7	0.4414	0.4180	0.3307	0.2308	0.1824	0.1831	0.1888	0.2284	0.2717	0.3148	0.3588	0.4188	0.5427
12	991.0	0.4731	0.4428	0.3875	0.3228	0.2477	0.2228	0.2472	0.2751	0.3044	0.3358	0.3882	0.5288	0.6542
11	769.3	0.5513	0.5157	0.4787	0.4388	0.3888	0.3828	0.3824	0.4228	0.4518	0.4814	0.5358	0.6544	0.7372
10	597.2	0.6435	0.5878	0.5888	0.5442	0.5188	0.4478	0.5075	0.5353	0.5656	0.5887	0.6700	0.7645	0.7878
9	463.5	0.7378	0.6812	0.6481	0.6248	0.5883	0.5855	0.5578	0.5743	0.6064	0.6418	0.7236	0.8228	0.8484
8	359.8	0.8103	0.7538	0.7128	0.6828	0.6605	0.6488	0.6286	0.6432	0.6713	0.7020	0.7748	0.8551	0.8867
7	279.3	0.8721	0.8288	0.7781	0.7433	0.7303	0.7316	0.7458	0.7677	0.7888	0.8054	0.8412	0.8862	0.8843
6	216.8	0.9282	0.8852	0.8453	0.8085	0.8046	0.8281	0.8536	0.8777	0.8826	0.9032	0.9114	0.9162	0.9228
5	168.3	0.9858	0.9631	0.9274	0.8872	0.8816	0.9102	0.9323	0.9320	0.9456	0.9720	0.9778	0.9820	0.9854
4	130.6	1.0000	0.9883	0.9780	0.9570	0.9411	0.9837	0.9787	0.9830	0.9881	1.0000	1.0000	1.0000	1.0000
3	101.4	1.0000	1.0000	0.9888	0.9820	0.9841	0.9838	0.9888	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Grid Line	Dose (rads)	Time Grid Lines (h) 79.43	100.00	125.89	158.49	199.53	251.19	316.23	398.11	501.19	630.96	794.33	1000.00	1258.93
18	4529.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
17	3516.3	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
16	2729.6	0.0541	0.0118	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15	2118.8	0.1387	0.0688	0.0300	0.0221	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14	1644.7	0.3840	0.2188	0.1251	0.0851	0.0331	0.0025	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
13	1276.7	0.5884	0.4140	0.2812	0.2141	0.1407	0.0820	0.0242	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
12	991.0	0.7475	0.6312	0.5221	0.4018	0.3547	0.3080	0.2208	0.1344	0.0900	0.0000	0.0000	0.0000	0.0000
11	769.3	0.7876	0.7704	0.7334	0.7342	0.6480	0.5884	0.5532	0.5107	0.4781	0.0592	0.0000	0.0000	0.0000
10	597.2	0.8133	0.8245	0.8338	0.8368	0.8253	0.7830	0.7683	0.7588	0.6481	0.4208	0.3481	0.0000	0.0000
9	463.5	0.8454	0.8445	0.8420	0.8370	0.8285	0.8048	0.7828	0.7862	0.7564	0.6832	0.5883	0.2881	0.0883
8	359.8	0.8699	0.8627	0.8517	0.8418	0.8328	0.8288	0.8180	0.8148	0.8237	0.8105	0.7226	0.5842	0.3098
7	279.3	0.8963	0.8885	0.8781	0.8736	0.8702	0.8706	0.8672	0.8683	0.8844	0.8781	0.8272	0.7173	0.5470
6	216.8	0.9231	0.9254	0.9283	0.9278	0.9290	0.9309	0.9323	0.9343	0.9380	0.9370	0.9189	0.8530	0.7868
5	168.3	0.9884	0.9910	0.9833	0.9853	0.9870	0.9885	0.9884	0.9888	1.0000	1.0000	1.0000	1.0000	1.0000
4	130.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	101.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table A12. Performance data grids for ITV-TOW crew.

Grid	Dose	Time Grid Lines (h)												
Line	(rads)	0.20	0.25	0.32	0.40	0.50	0.63	0.79	1.00	1.26	1.58	2.00	2.51	3.16
18	4529.9	1.0000	1.0000	0.8152	0.6317	0.4887	0.3721	0.2755	0.2008	0.1411	0.0947	0.0578	0.0282	0.0042
17	3516.3	1.0000	1.0000	1.0000	0.8714	0.8157	0.7017	0.5128	0.4033	0.3054	0.2663	0.2340	0.1820	0.1613
16	2729.6	1.0000	1.0000	1.0000	0.9852	0.8705	0.7744	0.6853	0.5887	0.4457	0.3827	0.3228	0.2408	0.2039
15	2118.8	1.0000	1.0000	1.0000	1.0000	0.9280	0.8466	0.8021	0.7004	0.5678	0.4703	0.3814	0.2883	0.2465
14	1644.7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9598	0.8349	0.7619	0.6474	0.5356	0.4255	0.3467	0.3145
13	1276.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9817	0.8233	0.7283	0.6082	0.4778	0.4181	0.3828
12	991.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9105	0.8326	0.7536	0.6452	0.5225	0.4419
11	769.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9213	0.8217	0.7786	0.6854	0.5979
10	597.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9034	0.8609	0.8186	0.7345
9	463.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9187	0.8577	0.8135
8	359.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9851	0.9129	0.8782
7	279.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9880	0.9363
6	216.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9439
5	168.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	130.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	101.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Grid	Dose	Time Grid Lines (h)												
Line	(rads)	3.98	5.01	6.31	7.94	10.00	12.59	15.85	19.95	25.12	31.62	39.81	50.12	63.10
18	4529.9	0.0032	0.0018	0.0014	0.0013	0.0011	0.0009	0.0010	0.0011	0.0012	0.0013	0.0017	0.0014	0.0007
17	3516.3	0.1257	0.0770	0.0583	0.0480	0.0406	0.0344	0.0362	0.0412	0.0477	0.0540	0.0747	0.0725	0.0567
16	2729.6	0.1866	0.1231	0.0811	0.0604	0.0487	0.0432	0.0434	0.0488	0.0643	0.0875	0.1209	0.1421	0.1483
15	2118.8	0.2081	0.1603	0.1237	0.0726	0.0568	0.0520	0.0505	0.0584	0.0810	0.1282	0.1865	0.2081	0.2543
14	1644.7	0.2829	0.2417	0.1676	0.1019	0.0848	0.0822	0.0846	0.1008	0.1286	0.1787	0.2220	0.2881	0.3622
13	1276.7	0.3554	0.3122	0.2254	0.1552	0.1281	0.1212	0.1304	0.1512	0.1786	0.2343	0.2731	0.3318	0.4716
12	991.0	0.4014	0.3583	0.3052	0.2506	0.1867	0.1640	0.1838	0.2044	0.2270	0.2542	0.2783	0.4516	0.5818
11	769.3	0.5162	0.4586	0.4175	0.3742	0.3284	0.3081	0.3330	0.3583	0.3831	0.4086	0.4681	0.6043	0.6805
10	597.2	0.6360	0.5671	0.5231	0.4877	0.4517	0.4284	0.4467	0.4717	0.4968	0.5280	0.6141	0.7261	0.7823
9	463.5	0.7378	0.6644	0.6094	0.5720	0.5340	0.5036	0.4963	0.5114	0.5416	0.5753	0.6644	0.7759	0.7875
8	359.8	0.8087	0.7445	0.6852	0.6431	0.6104	0.5841	0.5735	0.5880	0.6145	0.6454	0.7182	0.8151	0.8280
7	279.3	0.8682	0.8218	0.7645	0.7200	0.6940	0.6828	0.7066	0.7286	0.7478	0.7665	0.8023	0.8563	0.8645
6	216.8	0.9226	0.8812	0.8411	0.7981	0.7826	0.8027	0.8280	0.8504	0.8643	0.8757	0.8864	0.8977	0.9030
5	168.3	0.9619	0.9504	0.9166	0.8877	0.8708	0.8828	0.9138	0.9301	0.9436	0.9512	0.9562	0.9619	0.9667
4	130.6	1.0000	0.9867	0.9651	0.9457	0.9329	0.9494	0.9645	0.9765	0.9875	0.9828	0.9870	1.0000	1.0000
3	101.4	1.0000	1.0000	0.9831	0.9617	0.9747	0.9839	0.9833	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	0.9986	0.9876	0.9868	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Grid	Dose	Time Grid Lines (h)												
Line	(rads)	79.43	100.00	125.89	158.49	199.53	251.19	316.23	398.11	501.19	630.96	794.33	1000.00	1258.93
18	4529.9	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
17	3516.3	0.0508	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
16	2729.6	0.1790	0.0880	0.0161	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15	2118.8	0.3028	0.2161	0.1158	0.0534	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14	1644.7	0.4280	0.3350	0.2147	0.1032	0.0330	0.0037	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
13	1276.7	0.5557	0.4676	0.3413	0.1711	0.1201	0.0651	0.0231	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
12	991.0	0.6855	0.6295	0.5725	0.4863	0.2818	0.1801	0.1079	0.0513	0.0006	0.0000	0.0000	0.0000	0.0000
11	769.3	0.7541	0.7426	0.7311	0.6963	0.5862	0.4635	0.3682	0.3108	0.2704	0.0352	0.0000	0.0000	0.0000
10	597.2	0.7821	0.7800	0.7828	0.7828	0.7501	0.6883	0.6010	0.5178	0.4015	0.2281	0.1782	0.0000	0.0000
9	463.5	0.8050	0.8077	0.8086	0.8044	0.7818	0.7712	0.7189	0.6156	0.5361	0.4387	0.3855	0.2520	0.1111
8	359.8	0.8272	0.8250	0.8256	0.8255	0.8286	0.8212	0.7920	0.7082	0.6821	0.6173	0.5611	0.4818	0.3416
7	279.3	0.8600	0.8600	0.8600	0.8612	0.8606	0.8600	0.8552	0.8087	0.7878	0.7681	0.7257	0.6849	0.5685
6	216.8	0.9066	0.9081	0.9105	0.9114	0.9126	0.9137	0.9118	0.9053	0.8986	0.8874	0.8647	0.8238	0.7696
5	168.3	0.9705	0.9737	0.9764	0.9788	0.9811	0.9834	0.9854	0.9875	0.9883	0.9810	0.9826	0.9853	1.0000
4	130.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	101.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	78.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	61.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

PREDICTION MODELING OF PHYSIOLOGICAL RESPONSES AND SOLDIER PERFORMANCE IN THE HEAT

Kent B. Pandolf, Leander A. Stroschein,
Richard R. Gonzales and Michael N. Sawka
US Army Research Institute of Environmental Medicine
Natick, Massachusetts 01760-5007

ABSTRACT

Over the last two decades, our laboratory has been establishing the data base and developing a series of predictive equations for deep body temperature, heart rate and sweat loss responses of clothed soldiers performing physical work at various environmental extremes. Individual predictive equations for rectal temperature, heart rate and sweat loss as a function of the physical work intensity, environmental conditions and particular clothing ensemble have been published in the open literature. In addition, important modifying factors such as energy expenditure, state of heat acclimation and solar heat load have been evaluated and appropriate predictive equations developed. Currently, we have developed a comprehensive model which is programmed on a Hewlett Packard 41 CV hand-held calculator. The primary physiological inputs are deep body (rectal) temperature and sweat loss while the predicted outputs are the expected physical work-rest cycle, the maximum single physical work time if appropriate, and the associated water requirements. This paper presents the mathematical basis employed in the development of the various individual predictive equations of our heat stress model. In addition, our current heat stress prediction model as programmed on the HP 41 CV is discussed from the standpoint of propriety in meeting the Army's needs and therefore assisting in military mission accomplishment.

Key words: clothing; computer prediction modeling; environmental factors; exercise; physiological responses.

INTRODUCTION

Over the last two decades, the Military Ergonomics Division of the US Army Research Institute of Environmental Medicine has been establishing the data base and developing a series of predictive equations for deep body temperature, heart rate and sweat loss responses of clothed soldiers performing physical work in various environmental extremes. Individual predictive equations for rectal temperature [5], heart rate [6] and sweat loss [17] as a function of the physical work intensity, environmental conditions and particular clothing ensemble have been published in the open literature. In addition, important modifying factors such as energy expenditure [10], state of heat acclimation [7] and solar heat load [1] have been evaluated and appropriate predictive equations developed. Suitable data bases to evaluate the predictive importance of cardiorespiratory physical fitness [11,15], gender [12,14,16] and state of hydration [12,13] have been established. Our upper physiological limits during experimentation (rectal temperature $\leq 39.5°C$, heart rate ≤ 180 beats•min^{-1}) are within safe bounds and any errors in associated predictions within these limits should not materially endanger soldier performance.

Over this same time period, our Division has also attempted to program these predictive equations on various desk-top and hand-held calculators with the express purpose of developing a comprehensive heat stress model for predicting soldier performance to physical work, clothing and the environment. The initial computer program was written on a Hewlett Packard 9810A desk-top calculator with the outputs being the predicted rectal temperature and heart rate responses. As the technology advanced, we adapted these computer programs for the Hewlett Packard 65 hand-held calculator with similar outputs to those of the desk-top version. Currently, we have developed a more comprehensive model which is programmed on a Hewlett Packard 41CV hand-held calculator. The current model deals with the interaction of various multi-disciplinary factors such as (a) the theoretical physics of heat transfer, (b) the biophysics of clothing, (c) the physiology of metabolic heat production, distribution and elimination, and (d) related meteorological considerations. The primary physiological inputs are deep body (rectal) temperature and sweat loss while the predicted outputs are the expected physical work-rest cycle, the maximum single physical work time if appropriate, and the associated water requirements.

This paper presents the mathematical basis employed in the development of the various individual predictive equations of our heat stress model. In addition, our current heat stress prediction model as programmed on the HP 41CV is discussed from the standpoint of propriety in meeting the Army's needs and therefore assisting in military mission accomplishment.

MATHEMATICAL BASIS

Unless otherwise stated, all terminology for abbreviations and units of measurement follow the usage recommended by the Systéme international d'unites (SI units) and the International Union of Physiological Sciences.

Rectal Temperature Prediction

The general formula for predicting the final equilibrium rectal temperature (T_{ref}) as suggested by Givoni and Goldman [5] is

$$T_{ref}(^{0}C)=36.75+0.004(M-W_{ex})+0.0011\ H_{(r+c)}+0.8\ \exp[0.0047(E_{req}-E_{max})] \qquad [1]$$

(Metabolic) (Dry Heat) (Evaporative Heat)
(Exchange) (Exchange)

Equation 1 is comprised of three components

(1) the metabolic component $[36.75+0.004\ (M-W_{ex})]$

where $M\ =1.5W+2.0(W+L)(L/W)^{2}+\eta(W+L)[1.5(V_{w})^{2}+0.35GV_{w}]$ [2]

as originally published by Pandolf et al. [11]

and $W_{ex}\ =0.098\ G(W+L)V_{w}$ [3]

as suggested by Givoni and Goldman [5]

where M =metabolic rate, (watt)

W_{ex}=external work, (watt)

W =nude body weight, (kg)

L =clothing and equipment weight, (kg)

η =terrain factor

V_w =walking velocity, $(m \cdot s^{-1})$

G =grade, (%)

(2) the dry heat exchange component $[0.0011\ H_{(r+c)}]$

where $H_{(r+c)} = 6.45\ A_D(T_{db}-T_{sk})/I_T$ [4]

as inferred by Givoni and Goldman [5]

where A_D =body surface area, (m^2)

T_{db} =dry bulb temperature, (°C)

T_{sk} =average skin temperature, (°C)

I_T =total insulation including air layer (I_a) and intrinsic clothing, (I_{cl})

(3) the evaporative heat exchange component $\{0.8 \exp[0.0047(E_{req}-E_{max})]\}$

as indicated by Givoni and Goldman [5]

where E_{req} $=(M-W_{ex})+H_{(r+c)}$ [5]

and E_{max} $=14.21\ i_m/I_T\ A_{Deff}\ (P_{sk}-\phi_a P_a)$ [6]

where e = base of natural log

i_m = permeability index (N.D.)

A_{Deff} = effective surface area for evaporation, (m^2)

P_{sk} = water vapor pressure at the skin, (mm Hg)

ϕ_a = relative humidity, (%)

P_a = saturated water vapor pressure of air at T_{db}, (mm Hg)

and other abbreviations as described above.

In order to compute physical work-rest cycles, the time patterns of rectal temperature have been analyzed for three different conditions: (a) the time pattern for resting subjects under various heat stress conditions referred to as resting T_{ret} (resting rectal temperature at any time t); (b) the elevation pattern for rectal temperature during physical work at the given climatic conditions referred to a working T_{ret} (rectal temperature at any time t after beginning physical work); and (c) the recovery rectal temperature after cessation of physical work referred to as recovery T_{ret} (rectal temperature at any time t after completion of physical work). These three equations have been presented and discussed in detail elsewhere [5].

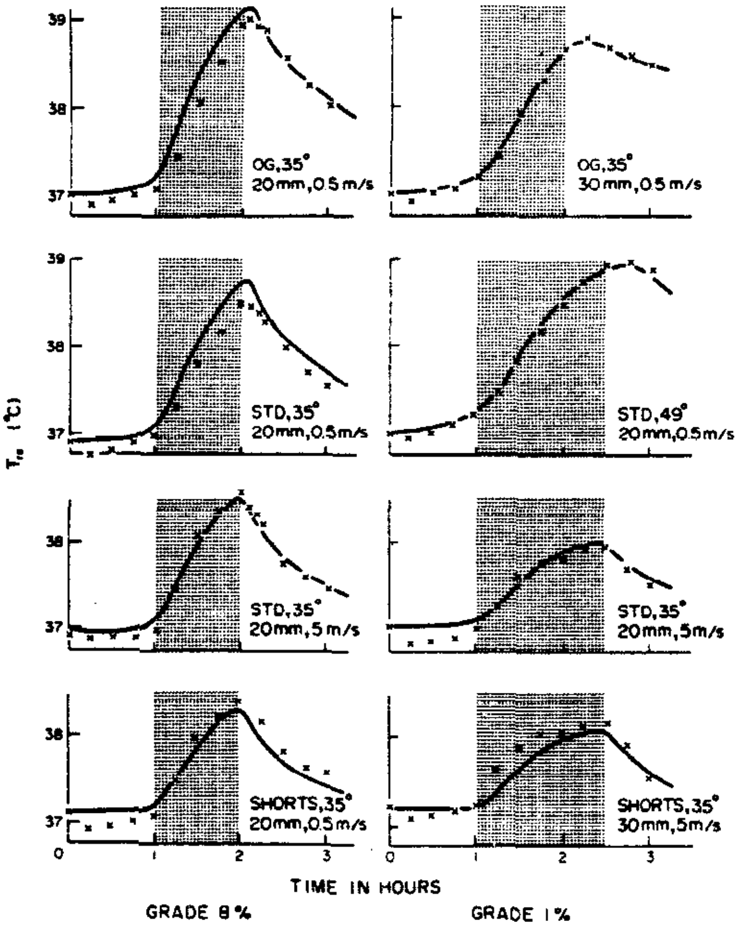


FIGURE 1. Comparison of predicted (lines) and measured (x) patterns of rectal temperature (T_{re}) during one hour cycles of rest, exercise and recovery for 12 soldiers as published by Givoni and Goldman [5]. Subjects wore shorts, standard fatigue uniforms (STD) or protective overgarments over fatigues (OG) in climatic conditions of either 35° or 49° C ambient temperature with vapor pressures of 20 or 30 mm Hg at wind speeds of 0.5 $m \cdot s^{-1}$.

Figure 1 presents a comparison for 12 volunteer male subjects of the predicted (lines) and measured (points) time patterns for rectal temperature during one hour cycles of rest, physical work and recovery as originally published by Givoni and Goldman [5]. These findings indicate that the prediction of rectal temperature from the proposed equations is in good agreement with the experimental observations covering a wide range of metabolic rates, climatic conditions and clothing properties.

All of the predictive formulae for rectal temperature presented and discussed above pertain to an exercise-heat acclimated individual. In order to characterize the non- and partially-acclimated individual, these equations were modified for the purpose of describing the acclimation process as the final equilibrium rectal temperature or for the general time pattern of rectal temperature as $\Delta\ T_{ref(accl)}$ and $\Delta\ T_{ret(accl)}$, respectively [7]. Figure 2 illustrates mean daily patterns of rectal temperature during seven days of exercise-heat acclimation with the points representing the average measured values for 24 subjects [7]. In general, there is good agreement between the measured and predicted patterns.

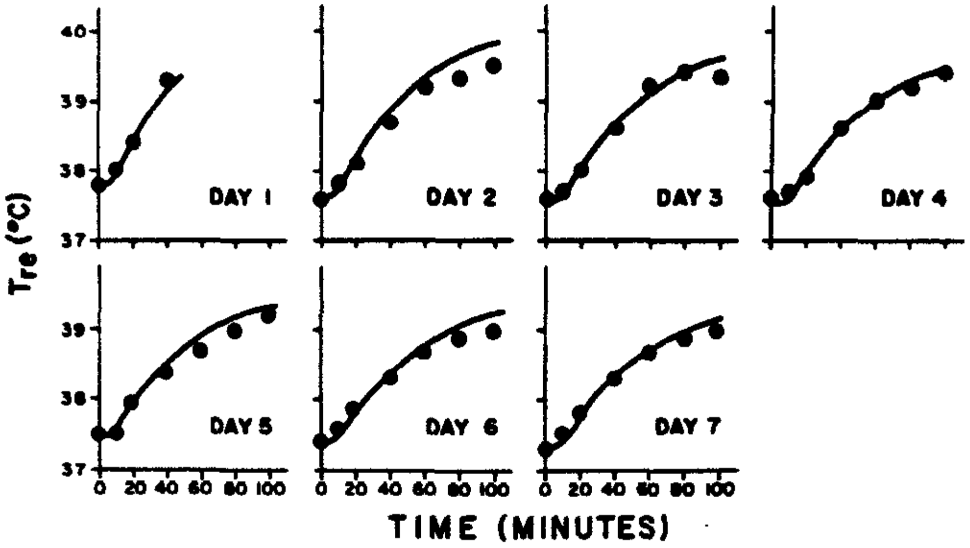


FIGURE 2. Comparison of predicted (lines) and measured (dots) patterns of rectal temperature as a function of day of acclimation for 24 soldiers walking for an attempted 100 min at 49°C, 20% rh as published by Givoni and Goldman [7].

Figure 3 shows the comparison of predicted and observed rectal temperature responses for 12 soldiers while wearing three different military clothing ensembles during tests under two different climatic conditions in Australia. These data which were collected by a group independent of our Institute are in quite good agreement with the predicted values, and in all but two instances, the observed responses are within ±1 S.D. of predicted.

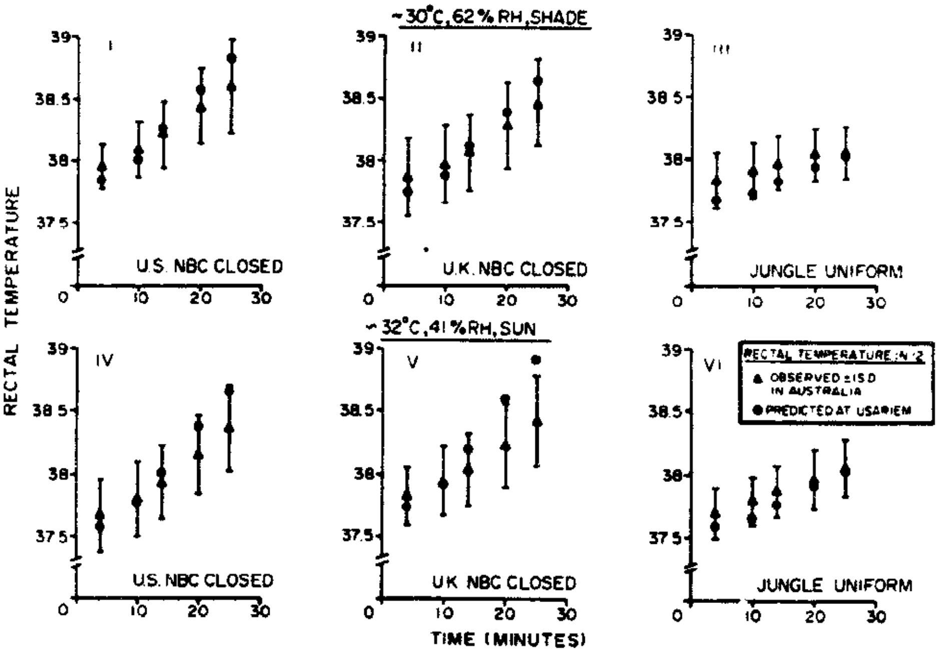


FIGURE 3. Comparison of predicted and observed rectal temperature responses of 12 soldiers while wearing three different military clothing systems each under two different climatic conditions.

Sweat Loss Prediction

The general equation for predicting sweat loss response (Δm_{sw}) as a function of exercise, environmental and clothing interactions as proposed by Shapiro et al. [17] is

$$\Delta m_{sw}(g\cdot m^{-2}\cdot h^{-1})=27.9\cdot E_{req}\cdot(E_{max})^{-0.455} \qquad [7]$$

where Δm_{sw}= change in body weight from sweat loss and other abbreviations as described earlier.

This prediction equation was derived from over 250 experimental exposures to a wide range of climatic conditions (ambient temperature, 20-54° C and relative humidity, 10-90%) while wearing various clothing ensembles (light clothing and heavy clothing of high permeability or low permeability) at different metabolic rates (rest to moderate physical work). Therefore, this formula can be employed over a wide range of E_{req} (50-360, $W\cdot m^{-2}$) and E_{max}

(20-525, $W \cdot m^{-2}$). In the present form, this formula is more applicable for predicting water requirements; however, it can be presented in appropriate units ($W \cdot m^{-2}$) for predicting the rate of sweat loss [17].

A comparison of predicted and measured Δm_{sw} for 111 individual exposures involving 24 soldiers is illustrated in Figure 4. These experiments considered ambient temperatures ranging from 35-49°C, relative humidities from 20-75%, different clothing ensembles and both resting and exercise evaluations. A correlation coefficient between the predicted and measured sweat loss of $r=0.94$ was observed over a wide range of sweating responses.

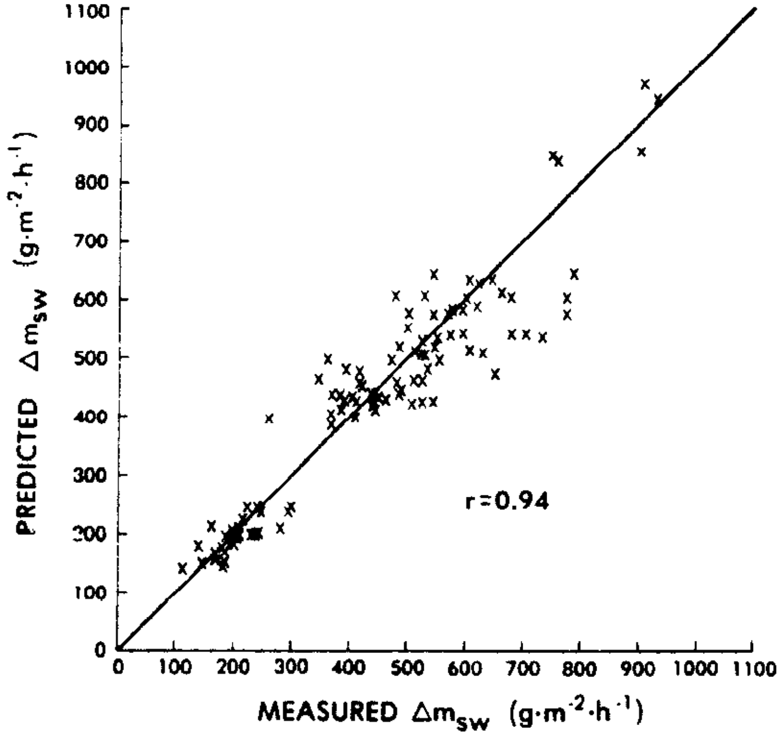


FIGURE 4. Relationship between predicted and measured sweat loss for 111 individual responses involving 24 soldiers as published by Shapiro et al. [17].

Figure 5 displays a comparison of four different methods for predicting sweat loss utilising the experimental findings of Shapiro et al. [17] derived from 34 soldiers. Lustinec's equation [8] employs a linear relationship between sweat rate and E_{req} for low skin wettedness but a non-linear relationship between sweat rate, E_{req} and E_{max} for high skin wettedness. Givoni and Berner-Nir [4] developed a prediction equation for expected sweat rate structured from the exponential function of the ratio E_{req}/E_{max}. Macpherson [9] developed the predicted four hour sweat rate index (P4SR) which incorporates ambient temperature, wet-bulb temperature, wind speed and correction for the particular clothing. With our equation, the predicted sweat loss was within the ±20% range for 29 out of the 30 experimental conditions that were evaluated with only one condition (37°C, 80% rh, walking in a sweat suit) greater than 20% from the measured value ($r=0.95$). Lustinec's equation showed eight conditions out of the

30 beyond the $\pm 20\%$ range (four additional conditions were beyond the equation's range) while for Givoni and Berner-Nir's equation 14 conditions were out of the $\pm 20\%$ range, and for the P4SR method 12 conditions were beyond the $\pm 20\%$ range (2 additional conditions were beyond this nomogram's range). Thus, the present formula was seen to predict steady-state sweat loss more accurately than other methods especially for extreme climatic conditions [17]. However, these same authors state that the present prediction equation may have some limitations at very high sweat rates.

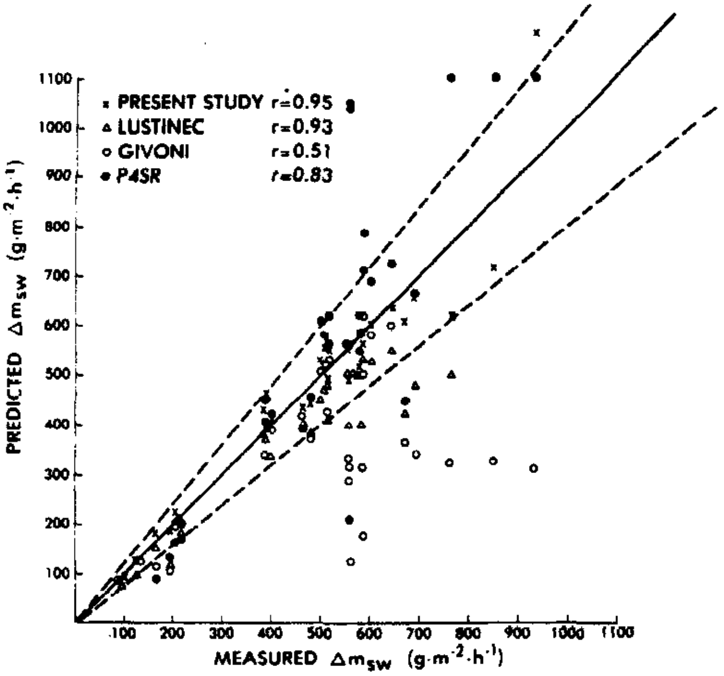


FIGURE 5. Comparison of four methods of predicting sweat loss using data from our Division derived from 34 soldiers as published by Shapiro et al. [17]. The solid line represents the line of identity while the dashed lines represent the $\pm 20\%$ range from the line of identity.

Heart Rate Prediction

The general formulas for predicting the final equilibrium heart rate (HR_f) as proposed by Givoni and Goldman [6] for heat acclimated individuals are

$$HR_f(\text{beats} \cdot \text{min}^{-1}) = 65 + 0.35(I_{HR} - 25) \text{ for } 25 \le I_{HR} \le 225 \quad [8]$$

$$HR_f(\text{beats} \cdot \text{min}^{-1}) = 135 + 45[1 - e^{(0.01[I_{HR} - 225])}] \text{ for } I_{HR} \ge 225 \quad [9]$$

where $I_{HR} = 100(T_{ref} - 36.75) + 0.4\ W_{ex}$

The time patterns for heart rate responses of heat acclimated individuals necessary to predict work-rest cycles have been described for work and rest at any time t as working HR_t and resting HR_t, respectively [6]. In addition, these same authors have presented a formula to predict the time pattern for heart rate recovery from the cessation of physical work towards the appropriate equilibrium resting level as recovery HR_t [6]. Further, Givoni and Goldman [7] published a predictive equation to describe the equilibrium heart rate responses expected for non- and partially- acclimated individuals. The computational adjustments necessary to predict the time patterns of heart rate during rest, work and recovery from work for non- and partially-acclimated persons are also displayed in this same reference.

A comparison between predicted and measured final equilibrium heart rate responses from our own investigations (n=33) and the investigations [9,18] of others (n=75) as originally presented by Givoni and Goldman [6] is shown in Figure 6. For both our own observations and the observations of others, the agreement between the measured and the predicted heart rate responses is excellent as shown in the figure.

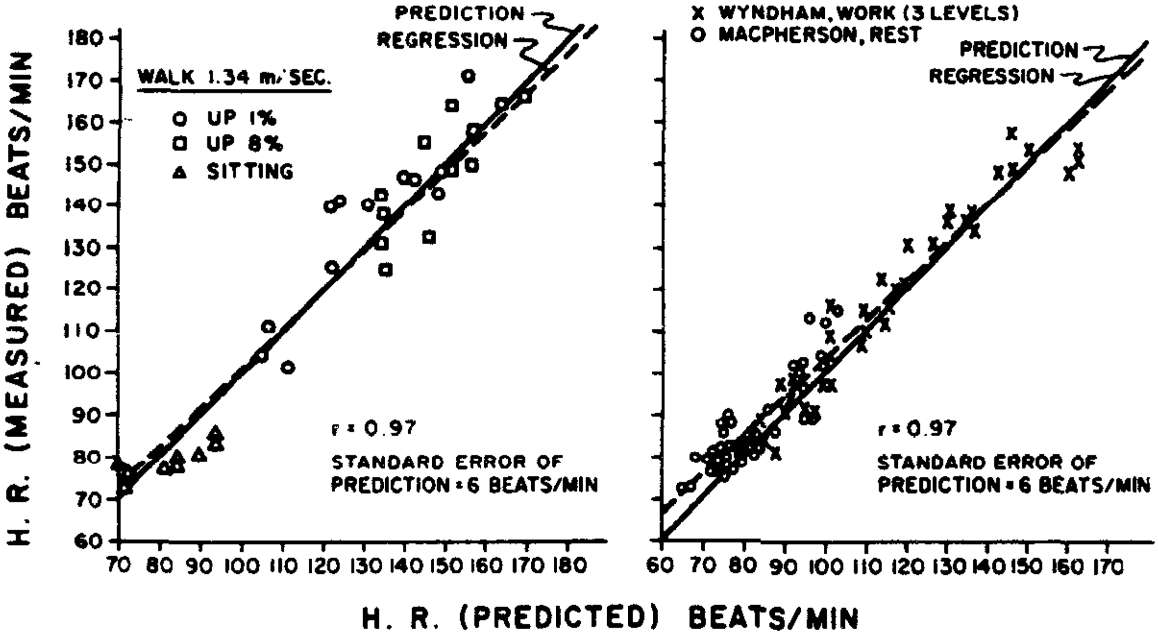


FIGURE 6. Relationship between predicted and measured final heart rate responses: left, from our own Division studies (n=33); right, from observations (n=40) by Macpherson [9] and observations (n=35) by Wyndham et al. [18] as published by Givoni and Goldman [6].

CURRENT HEAT STRESS PREDICTION MODEL

As stated earlier, the current version of our heat stress prediction model is programmed on a standard Hewlett Packard (HP) 41CV hand-held calculator. The only major modifications to the standard HP 41CV involve (a) the addition of a specially designed portable eprom (Hand Held Products, Inc.) for 32K added memory and (b) a redesigned touch pad. With the 32K of added memory, the HP 41CV presents 36K of memory of which 8K is currently programmed. The redesign of the touch pad for the HP 41CV to incorporate our heat stress prediction modeling needs is shown in Figure 7.

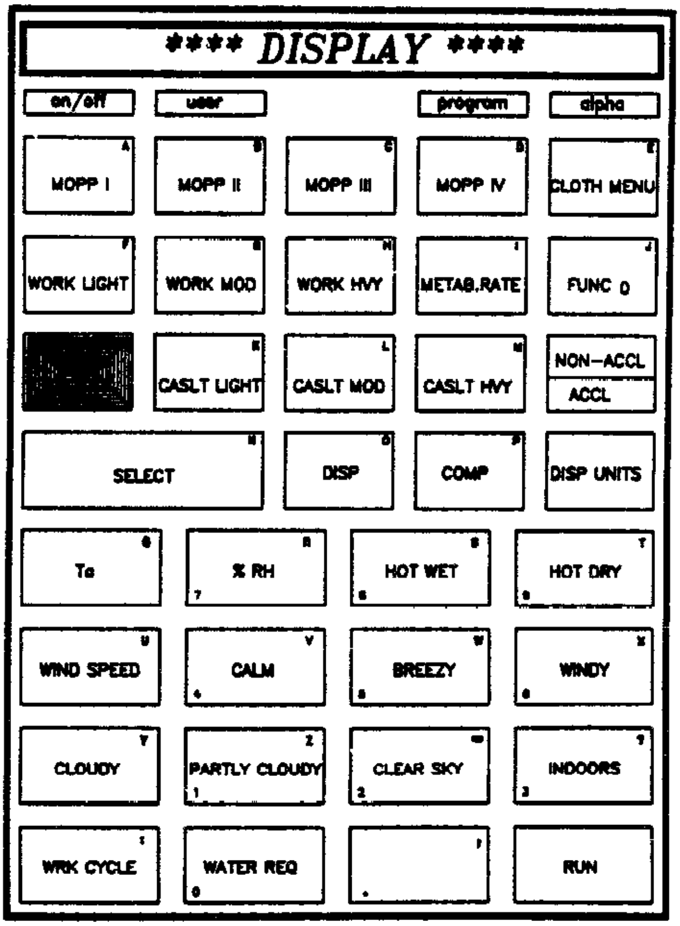


FIGURE 7. The redesigned touch pad of the Hewlett Packard 41CV which encompasses the input parameters of our heat stress prediction model.

As seen in Figure 7, the prefix keys ("select","disp","comp"and "disp units") are located near the center of the touch pad. Above these prefix keys the user observes keys for parameters which describe the soldier. The three rows of keys immediately below the prefix keys describe the environment. The bottom row of keys are output or information keys.

The top row of keys are used to set the computer programming parameters for the soldier's clothing system. Separate keys are designated for Mission Oriented Protective Posture (MOPP) levels I-IV which are based on the protective clothing and equipment worn. The various levels of MOPP provide a flexible clothing system to protect soldiers against suspected chemical agents during chemical warfare which may help facilitate mission accomplishment. In addition, 21 other clothing systems are available in a clothing menu which is displayed in Table 1. This table shows the description of the particular clothing system and the display given on the HP 41CV. Each of the 25 clothing systems which are available to the user have individual coefficients which describe the thermophysical properties of the clothing as a function of the work rate and effective wind velocity. This concept has been presented in some detail elsewhere [5].

TABLE 1. CLOTHING MENU

	DISPLAY	DESCRIPTION
1.	:AVIAT	Aviators
2.	:AVIAT+ARM	Aviators + armor (mask+hood)
3.	:AV+OG+ARM	Aviators +OG+armor (MOPP IV)
4.	:AV+UK+UNDW	Aviator+UK underwear (MOPP IV)
5.	:BDO+RAIN	BDO + rainsuit
6.	:BDU	BDU
7.	:BDU+ARMOR	BDU + armor
8.	:BDU+RAIN	BDU + rainsuit
9.	:CVC	CVC
10.	:CVC+CBR, MI	CVC + CBR (MOPP I)
11.	:CVC+CBR, MIV	CVC + CBR (MOPP IV)
12.	:DESRT CAMOF	Desert camouflage
13.	:DESERT TAN	Desert tan
14.	:EOD+FATIGUE	EOD over fatigues
15.	:FIRE+FATIG	Firefighters over fatigues
16.	:FUEL HANDLR	Fuel handlers (TAP)
17.	:MOPP I	MOPP I
18.	:MOPP II	MOPP II
19.	:MOPP III	MOPP III
20.	:MOPP IV	MOPP IV
21.	:PONCH+FATIG	Poncho over fatigues
22.	:TROP CAMOFL	Tropical camouflage
23.	:TROP FATIG	Tropical fatigues
24.	:TROP FA+ARM	Tropical fatigues + armor
25	:UTIL FATIG	Utility fatigues

The second row of keys from the top sets the internal parameters for the soldier's metabolic work rate. Individual keys are available to describe light, moderate and heavy physical work which are categorized as 250, 425 and 600 W, respectively. If another known metabolic rate is desired, it can be entered using the "metab rate" key. While the preferred units for metabolic rate are watt, values can be entered in kcal$\bullet$hr^{-1}, BTU$\bullet$hr^{-1}, or METS. This same key can be used to input the components necessary to compute the metabolic rate where body weight (kg), external load (kg), walking speed (m$\bullet$s^{-1}), grade (%), and a terrain coefficient are necessary [see 11]. The multiplication factors necessary to compute metabolic rate as a function of terrain are presented in Table 2. Finally, an additional key ("func") is available, but yet unprogrammed, to possibly compute metabolic rate for other modes of locomotion than walking such as running, lifting, etc..

TABLE 2. MULTIPLICATION FACTORS FOR ENERGY COST AS A FUNCTION OF TERRAIN

TERRAIN	η
BLACKTOP SURFACE	1.0
DIRT ROAD	1.1
LIGHT BRUSH	1.2
HARD PACKED SNOW	1.3
HEAVY BRUSH	1.5
SWAMPY BOG	1.8
LOOSE SAND	2.1
SOFT SNOW	1.3+0.08 (CMS. OF SNOW PRINT DEPTH LEFT BY FOOT)

The third row of keys from the top are to individually categorize casualties ("caslt") as light, moderate or heavy, and to describe state of heat acclimation as either non-acclimated ("non-accl") or fully acclimated ("accl"). Light, moderate and heavy casualties are described as less than 5% casualties, about 20% casualties, and greater than 50% casualties, respectively. These casualty categories are also based on individual upper limits for deep body temperature which were developed from information by Goldman [3] and scientific results provided by Israel Defence Forces Technical Reports.

The first row of keys below the prefix keys address the ambient air temperature (T_a) and relative humidity (%rh). The T_a can be entered in either °C or °F while relative humidity can be evaluated as per cent relative humidity, wet bulb temperature (°C or °F), dew point or vapor pressure. If this information is not available to the user, input keys for our standard hot-wet (35°C, 75% rh) or hot-dry (49°C, 20% rh) climatic conditions are available.

The second row of keys below the prefix keys allow the user to provide input concerning the wind speed. While the preferred units are $m \cdot s^{-1}$, the expected wind speed ("wnd spd") can be entered in units of mph, $km \cdot hr^{-1}$, $ft \cdot sec^{-1}$ or knots. Calm, breezy or windy conditions are categorized as 0.5, 2.0 and 4.0 $m \cdot s^{-1}$, respectively.

The third row of keys below the prefix keys address the impact of the solar heat load. The internal parameters used in considering solar heat load were developed from the concept of mean radiant temperature [2] as applied to clothing heat exchange by Breckenridge and Goldman [1]. Categorizations are cloudy, partly cloudy ("prt cloudy") or clear sky with allowance for the indoors where there is no appreciable solar load.

The output keys are at the bottom of the calculator. The "wrk cycle" key provides output for the calculated work-rest cycle, and the one time only maximum work period with time periods in minutes. The "water req" key allows the user to compute the water requirements during work, rest and combined in canteens per hour or quarts per hour. One of the output keys remains uncommitted and remains available for future use.

TABLE 3. PREDICTED PHYSICAL WORK-REST CYCLES AND WATER REQUIREMENTS ASSOCIATED WITH FOUR DIFFERENT MILITARY SCENARIOS

	SCENARIO 1	SCENARIO 2	SCENARIO 3	SCENARIO 4
INPUTS:	MOPP 1	MOPP 1	MOPP 4	MOPP 4
	HVY.WRK.	HVY.WRK.	HVY.WRK.	MOD.WRK.
	HVY.CASLT.	HVY.CASLT.	HVY.CASLT.	HVY.CASLT.
	ACCL.	ACCL.	ACCL.	ACCL.
	HOT DRY	HOT DRY	HOT DRY	HOT DRY
	WINDY	WINDY	WINDY	WINDY
	CLOUDY	CLEAR SKY	CLEAR SKY	CLEAR SKY
RESULTS:	Time W:R:M=33*27*84	Time W:R:M=28*32*74	Time W:R:M=14*46*52	Time W:R:M=24*36*87
	Water W:R:C=2.3*0.9*1.7	Water W:R:C=2.4*1.1*1.7	Water W:R:C=2.4*1.1*1.4	Water W:R:C=2.2*1.1*1.6

W:R:M=work:rest:maximum work [time periods (minutes)]
W:R:C=work:rest:combined [water requirements (canteens per hour)]

Table 3 illustrates the predicted physical work-rest cycles and associated water requirements for four different military scenarios. The required inputs are the clothing worn, physical work rate, casualty level, acclimation state, environmental conditions, wind speed and solar heat load. The expected outputs are the physical work-rest cycle (min), one time only maximum work period (min), and the associated water requirements (canteens per hour). Compared to Scenario 1, the results of Scenario 2 illustrate the importance of the solar load in reducing both the physical work-rest cycle and one time only maximum work period while increasing the associated water requirements. Results from Scenario 3 display the dramatic reduction in the work component of the work-rest cycle and the associated reduction in the one time only maximal work period while wearing MOPP IV. The results from Scenario 4 show the benefits of reducing the metabolic work rate from heavy to moderate in terms of improvement in the work component and the work-rest cycle and enhancement of the one-time only maximum work period. Hopefully, the military user can employ this calculator to help avoid unnecessary casualties associated in the environmental heat extremes, and by predicting appropriate work-rest cycles and water requirements facilitate the achievement of mission objectives.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the many significant contributions of Mr. John R. Breckenridge and Drs. Baruch Givoni and Ralph F. Goldman in the development of this project, and the technical assistance of Mrs. Edna R. Safran in preparing the manuscript.

This report was published in large part by Pandolf, K.B., L.A. Stroschein, L.L. Drolet, R.R. Gonzalez and M.N. Sawka. Prediction modeling of physiological responses and human performance in the heat. Comput. Biol. Med. 16, 319 (1986).

REFERENCES

1. Breckenridge, J.R. and R.F. Goldman. Solar heat load in man. J.Appl.Physiol. 31,659 (1971).
2. Gagge, A.P. Effective radiant flux, an independent variable that describes thermal radiation on man. In: Physiological and behavioral temperature regulation. J.D. Hardy, A.P. Gagge, and J.A.J. Stolwijk (eds), Charles C. Thomas, Springfield, pp. 34-45 (1970).
3. Goldman, R.F. CW protective clothing: the nature of it performance degradation and some partial solutions. Handbook of the thirteenth commwealth defence conference on operational clothing and combat equipment. pp.66-68 (1981).
4. Givoni, B. and E. Berner-Nir. Expected sweat rate as a function of metabolism, environment, and clothing. Res.Rep.UNESCO. Israel Institute of Technology, Haifa (1967).
5. Givoni, B. and R.F. Goldman. Predicting rectal temperature response to work, environment, and clothing. J. Appl. Physiol. 32,812 (1972).
6. Givoni, B. and R.F. Goldman. Predicting heart rate response to work, environment, and clothing. J. Appl. Physiol. 34,201 (1973a).
7. Givoni, B. and R.F. Goldman. Predicting effects of heat acclimatization on heart rate and rectal temperature. J. Appl. Physiol. 35,875 (1973b).
8. Lustinec, K. Sweat rate, its prediction and interpretation. Arch. Sci. Physiol. 27,A127 (1973).
9. Macpherson, R.K. Physiological responses to hot environments. Medical research council special report series no. 298. Her Majesty's Stationary Office, London, pp.219, 294-299 (1960).
10. Pandolf, K.B., R.L. Burse, and R.F. Goldman. Role of physical fitness in heat acclimatisation, decay and reinduction. Ergonomics 20,399 (1977a).

11. Pandolf, K.B., B. Givoni, and R.F. Goldman. Predicting energy expenditure with loads while standing or walking very slowly. J. Appl. Physiol. 43,577 (1977b).
12. Sawka, M.N., M.M. Toner, R.P. Francesconi, and K.B. Pandolf. Hypohydration and exercise: effects of heat acclimation, gender and environment. J. Appl. Physiol. 55,1147 (1983).
13. Sawka, M.N., R.P. Francesconi, N.A. Pimental, and K.B. Pandolf. Hydration and vascular fluid shifts during exercise in the heat. J. Appl. Physiol. 56,91 (1984).
14. Shapiro, Y., K.B. Pandolf, and R.F. Goldman. Sex differences in acclimation to a hot-dry environment. Ergonomics 23,635 (1980a).
15. Shapiro, Y., K.B. Pandolf, B.A. Avellini, N.A. Pimental, and R.F. Goldman. Physiological responses of men and women to humid and dry heat. J. Appl. Physiol. 49,1 (1980b).
16. Shapiro, Y., K.B. Pandolf, B.A. Avellini, N.A. Pimental, and R.F. Goldman. Heat balance and transfer in men and women exercising in hot-dry and hot-wet conditions. Ergonomics 24,375 (1981).
17. Shapiro, Y., K.B. Pandolf, and R.F. Goldman. Predicting sweat loss response to exercise, environment and clothing. Eur. J. Appl. Physiol. 48,83 (1982).
18. Wyndham, C.H., N.B. Strydom, J.F. Morrison, F.D. Dutoit, and J.G. Kraan. Responses of unacclimatized men under stress of heat and work. J. Appl. Physiol. 6,681 (1954).

USE OF EXPERT SYSTEMS TO PREDICT SOLDIER PERFORMANCE

Philip H. Warren, III, Dr.
U.S. Army Natick Research, Development and Engineering Center
Natick, MA 01760-5020
(508)-651-4381
AV 256-4381

Heather L. Dragsbaek
GEO-CENTERS, Inc.
Newton Centre, MA 02159
(508)-651-4631

Introduction

Predicting the outcome of a mission in the modern battlefield is complicated by the difficulty in predicting the performance of individual soldiers. There are many models for predicting human performance in limited conditions of temperature, load, terrain, etc. (see References 1-3). However, there are few mathematical models that provide good predictions of human performance under a wide variety of conditions (see References 4-5). These models can be used to assist military commanders as they make decisions to establish battlefield objectives and to allocate resources. With the use of computers, and especially the techniques of artificial intelligence (AI), there is now the potential to automate these models and make them available to more users under numerous conditions.

The combination of a broadly based model of human performance and the technology of expert systems can be used to solve important problems which have resisted solution by more traditional means. Such a problem is that of the soldier's load. Historically, infantry soldiers have marched into combat carrying loads that are too heavy (see Reference 6). This load, combined with high heat, humidity, or a requirement to move quickly contributes to a reduction in combat effectiveness of the soldier, increases the chances that he will become a casualty, and, ultimately, jeopardizes his mission. Figure 1 shows that with the increase in sophistication of man's ability to conduct warfare, he has gradually increased the load the soldier must bear in combat.

Research in human factors engineering, physiology and biomechanics has yielded limits for the maximum load a soldier can carry and the maximum amount of work he can perform. Beyond these limits, the soldier is at risk of becoming a casualty prior to engaging the enemy and, therefore, of failing to complete his mission. As a result, the Army has adopted the doctrine that a soldier's marching/approach load should not exceed 45% of his body weight, and his combat load should not exceed 30% of his body weight (MIL-STD 1472C, see Reference 7). For the typical soldier of 160 lbs (72 kg), these percentages are equivalent to 72 lbs (32.4 kg) and 48 lbs (21.6 kg), respectively. Unfortunately, soldiers often carry a great deal more than this; in some cases, weights excede 150 lbs (67.5 kg). For example, the total load for the assistant Dragon gunner, in a temperate climate, is 167 lbs (75.2 kg). For the assistant machine gunner, it is 152 lbs (68.4 kg). In fact, the total load for each position in a rifle platoon exceeds 110 lbs

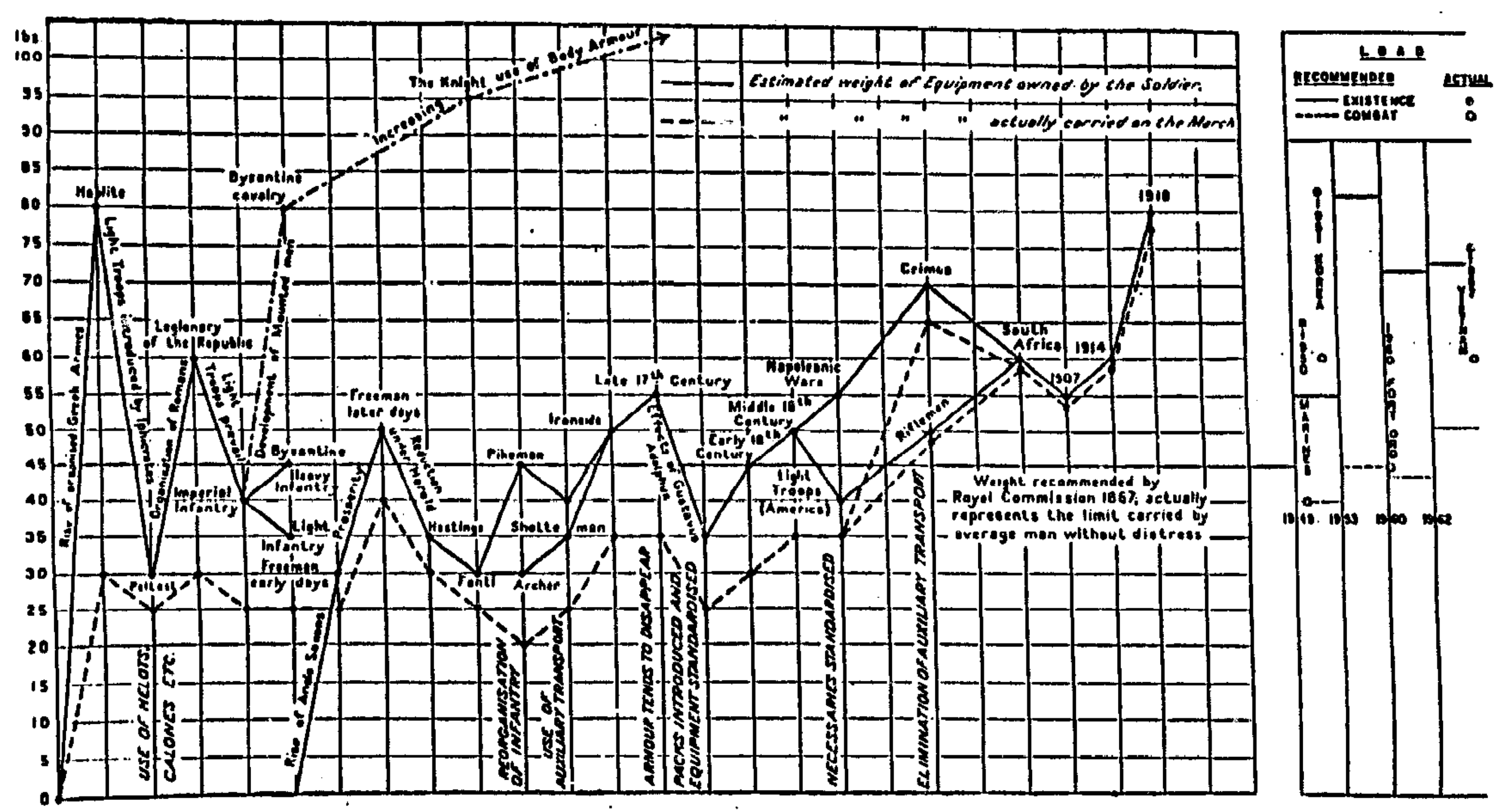


Figure 1. This figure illustrates how long the soldier's load has been a problem. Since the rise of the organized armies of the Greeks, soldiers have carried significant loads into combat. The dashed line indicates the estimated weights of equipment soldiers actually carried on a march. From the time of the Greeks to the early 18th century, soldier loads were fairly constant. However, for the next 150 years or so, soldier loads increased dramatically to the level where they are today. For example, it is indicated on the right portion of the figure that the combat load for the soldier involved in the Vietnam conflict was 60 lbs (27 kg). That is 33% over the recommended weight for optimal performance. (Figure 1 from Reference 9.)

(49.5 kg) (see Reference 8). If loads are tailored, that is, if decisions are made about what to carry and what not to carry, load weights can be reduced significantly, but they still exceed the doctrinal limits by substantial amounts. Therefore, soldiers are still at risk of becoming casualties.

Traditional Approach: The traditional approach to this persistent problem has been to develop lighter materials for clothing, ballistic protection, food, weapons, communication equipment, etc. There have been remarkable achievements in this arena, and goals for future weight reductions range from 12 to 50% (see Reference 8). While making lighter equipment will remain critical, doing so does not solve the soldier load problem. The reason is that commanders make decisions about what to carry. The result is, that for every reduction in weight there is a concomitant increase in capability for fire power, communication, and detection of the enemy, each one of which requires equipment that adds weight to the soldier's load (see Reference 10). Thus, real-life decisions about what to carry involve complex cost-benefit analyses, often based on minimal hard information when there is little time to consider alternatives and their consequences. Under these conditions, it is easy to understand how the decision-maker would tend to err on the side of carrying too much with him: "We don't know what's ahead, so let's make sure we are prepared for every contingency" (see Reference 11).

New Approach - Decision Support: What is needed is a novel approach to solving the soldier load problem. Decision support technology is such an approach. It provides a means for assisting commanders to make decisions about tailoring soldier loads based on mission information. Decision support technology takes the form of expert systems which are computer programs that process information in ways similar to the ways human experts do. Expert systems have been developed for use in a wide variety of situations, such as diagnosis of medical conditions (see Reference 12), location of mineral deposits (see Reference 13), determination of the structure of organic compounds (see Reference 14), and others (see References 15-17). They use many elements of information to make deductions, inferences, and recommendations. A significant advantage of these systems is that deductions and inferences are made without being degraded by the kinds of stressors to which humans are susceptible. The deductions are derived from hard rules that constrain the output in ways similar to the way a database program does. In contrast, inferences are made using rules, or heuristics, which are true only some of the time: e.g., Boys like to play baseball (that may be true only 85% of the time). On the basis of these deductions and inferences, the expert system makes recommendations which aid the user in making many decisions rapidly.

A soldier Load Expert System (LES) is under development at Natick. LES will accept input about the soldier and his mission and provide output in the form of recommendations about what to carry and possible changes to the mission scenario. The program will also provide dialogue and "what if" facilities. In the dialogue mode, the user can respond to recommendations or questions from the program. This facility might lead the user/decision-maker to change his mind about what to carry or some other aspect of performing his mission. This leads directly to the "what if" capability. Here the decision-maker can change his input to determine if it improves the assessment of the mission by the expert system.

The initial user of LES is projected to be company and field grade commanders in training at the Infantry School, Fort Benning. In their current training they have a block of instruction on load configuration and the importance of not overloading the soldier. With LES we can go beyond telling the commander, "The decisions you make about load can mean the success or failure of a mission." LES will allow the commander to receive feedback on his decisions faster than he would in realtime. He can then modify parts of the input and test the effects. Thus, he can learn by following the consequences of his decisions.

Development of the System

The first step in the development process of a system such as this is to identify an important problem that is amenable to solution using decision support technology. This means that problems in this domain do get solved. Also, if there are some individuals who are significantly better than others (i.e., experts) at solving this type of problem, then this expertise can be captured in the expert system and shared with those who have less expertise.

Knowledge Engineering: Once the problem has been defined and expert system technology has been chosen as the way to solve the problem, then the knowledge engineering phase begins (see Figure 2). In this phase the knowledge or expertise that can be used to solve the problem must be identified, collected and organized into a form that a programmer can put into code. Two sources of knowledge have been identified for LES. The first is a body of scientific information related to the heat stress model developed at U.S. Army Research Institute for Environmental Medicine (USARIEM). The details of this model have been published in the scientific literature over the last two decades (see References 5 and 18-21). The model provides excellent predictions of rectal temperature of a soldier under a wide variety of load, heat, humidity, acclimatization and other mission related conditions. This information can then be used to determine how long a soldier can work at a given rate, compute work/rest cycles, determine water requirements and the chances of the soldier becoming a heat casualty. While other models of predicting casualties are possible, and some are available (see Reference 4), we believe the USARIEM model provides the most robust and accurate predictions over the widest range of conditions the soldier is likely to encounter.

The second phase of knowledge engineering for this project is capturing the knowledge of commanders who have experience making decisions about load configuration. This process requires a different approach than that used above. In this case commanders are interviewed about how they make load configuration decisions. They are provided with mission scenarios with which to work, and, with their help, test cases are developed to evaluate the expert system during development. One of the most difficult challenges of this phase is converting the expertise of the commanders into clear rules that can be coded into the expert system.

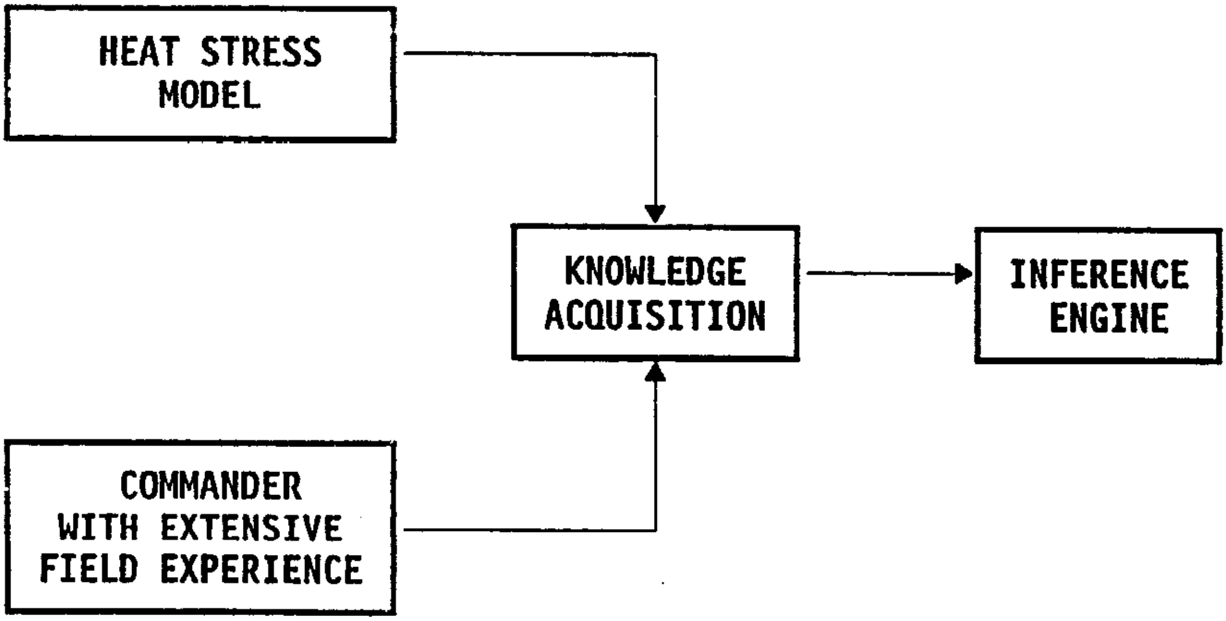


Figure 2. This figure illustrates the knowledge engineering process for the LES project. During knowledge acquisition, expertise is captured and put into a form that a computer programmer can code into the computer. Expertise about the effects of load and other mission variables comes from two different sources. The heat stress model developed at USARIEM provides formulae for computation of the effects of mission variables that affect the soldier's ability to complete his mission. The second source of expertise is the military commander with extensive experience making the kinds of decisions with which LES is designed to help. Through detailed interviews knowledge engineers derive rules that the expert uses to make decisions in real-life situations. Validated rules are then collected together to form the basis of the inference engine.

<u>Implementation</u> - <u>Software</u>: The present prototype of the system uses two computer languages; Turbo-C and Turbo-Prolog (both products of Borland). C is used for the mathematically intense portions of the program in which several equations are used to calculate heat exchange, rectal temperature,

water requirements, etc. Prolog is used for all other portions of the system. Prolog is a computer language specifically designed for use in artificial intelligence (AI) applications. It is very good for supporting user/system dialogue and making inferences based on a set of rules. Turbo-Prolog is particularly strong in supporting screen control and comes with many useful tools for software development. The program modules developed in these two languages can be compiled, thus allowing them to be run on any MS-DOS compatible computer.

Hardware: Our objective is to develop a system that is accessible to the widest possible range of users. Therefore, the minimum hardware requirements to support this system are a PC which runs MS-DOS, has 640k of core memory, and a hard disk. A color monitor is not be required, but is recommended.

System Description: LES will accept user input in several domains and provide output in the form of summary information and recommendations derived from the input (see Figure 3). Input will include information related to the soldier, such as height, weight, the uniform he will wear on the mission, and days he has acclimated to the heat. The user will select items of equipment for the soldier to carry from various menus. The weights of the items will be stored in a database. Finally, there will be input related to the mission such as distance to be traveled, time available, and information about the environment (i.e., nature of the terrain, grade, temperature, humidity, wind speed, cloud cover). All of these variables will have default values (the average or most common value) to allow the user to utilize the system without entering information for every variable. Output includes results based directly on data about the soldier, load, and mission. Derived results will include a one-time maximum duration that the soldier can work and a recommended work/rest cycle. Also included in the output will be recommendations about how much food and water the soldier should consume in order to meet his energy needs and to avoid dehydration. Finally, if the system detects a problem, for example not enough time has been allowed for covering the required distance, LES will automatically initiate a dialogue with the user and recommend that he consult the expert advisor portion of the system. The dialogue will begin with a statement of the problem and will include recommendations for solving it. For example, a soldier might be configured with a load that weighs 115 lbs (51.8 kg) and be required to march 15 miles over uneven terrain in 5 hrs in an environment of high temperature and high humidity. The expert system would state that this mission would result in a significant risk of the soldier becoming a heat casualty and would recommend either a lighter load (Can other means be found for transporting heavy equipment to the destination?), or a change in the conditions (Can another route be used? or Can the marching begin in the very early morning when it is cooler?). Commanders with field experience will be able to guide us toward options that are viable and away from those that are not.

The user will have available a "what-if" capability. Thus, he will be able to test certain hypotheses on the spot. "What happens if I leave my night scope (3.5 lbs, 1.6 kg) and sleeping bag (7 lbs, 3.2 kg) behind, and bring only my chemical protective mask, leaving behind the rest of my NBC

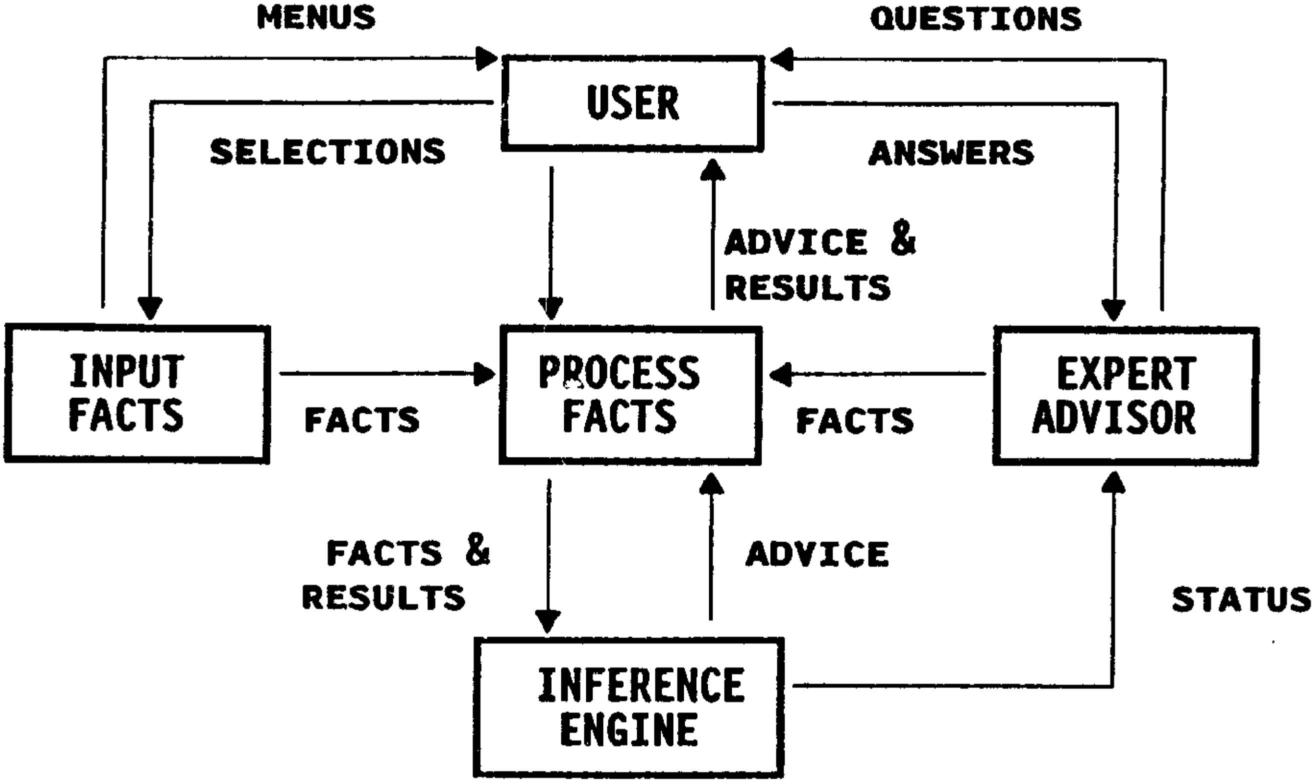


Figure 3. This diagram depicts the flow of information (facts, user initiated commands, expert advice) in the Load Expert System (LES). Default values exist for all variables used, therefore, the user can enter data for those variables for which he has information, and then process the data. When the results of the processing are returned, the user may consult the expert advisor. The advisor can recommend that changes be made in the soldier load or some other aspect of the mission. This cycle of user input - processing - results - questions/recommendations, will continue until a successful mission scenario is achieved or the user decides that he can not make further changes to the mission and, therefore, further processing is unnecessary.

gear (savings of 9.3 lbs, 4.2 kg). The expert system would then recompute to determine if those decisions were helpful enough. Other alternatives can be considered until the user finds a set of conditions with which he is satisfied.

The most direct benefit of LES in the classroom is that it will show the commander/user the consequences of his decisions. He can define different mission scenarios and get feedback about whether there is sufficient benefit derived from the costs associated with not having the soldier carry selected items of equipment with him. He will learn the relative impact of dropping 10 lbs (4.5 kg) of weight versus slowing the rate of marching (if time permits) versus altering the route (this could be used to improve the terrain, but it might increase the required time at a given marching speed). Finally, by becoming better at solving problems related to the soldier's load, the commander will be better able to deal with both logistical and operational problems.

System Evaluation: LES will go through several phases of evaluation. User-friendliness will be evaluated using two methods. The first method involves a new technology: automated analysis of screen layout (see Reference 22). These analyses will be most useful in assuring a system that is optimally efficient from the viewpoint of the decision-maker and, we hope, increase the likelihood of acceptance of the system. Secondly, of course, the user (commander in training at Ft. Benning) must give us his feedback. This system, like all true expert systems, is a dynamic system, and therefore, can grow with the needs of the user. During the next phase of development we will test LES with decision-makers at Ft. Benning and use the information we gain to improve the system.

Validation of the computations and recommendations of LES will depend on three processes. First, test cases will be developed which represent diverse but likely mission scenarios. These test cases will be developed with field experts and will be performed every time a change is made in LES. The second procedure involves continuous use of the system by instructors at the Infantry School as a normal part of the training process. This latter procedure will provide important feedback from several students and instructors who will test the system with their own mission scenarios. Finally, when the first version of LES is complete it should be tested in a "shoot-off" where the success of actual field missions is compared using LES against new commanders. LES will be a success if it helps commanders get more soldiers to their destinations and if those soldiers are more capable of carrying out their missions.

Conclusion

Throughout history the infantry soldier has had to conduct combat missions while carrying extreme loads. Military commanders recognize the significant effects this weight can have on a soldier's performance in combat. The traditional solution, lighter materials, has not solved the problem. This paper describes an application program which combines the USARIEM heat stress model with expert system technology to produce the soldier load expert system (LES). The field commander can use LES to assist him in making decisions about tailoring loads. The system however, goes beyond simply the load tailoring function. Mission information such as rate of marching, terrain footing, meteorological information, etc. are put into the system. This additional information allows for a much broader range of decision-support. The system's expert advisor can suggest that specific items be excluded from the load, that alternative routes be considered, or that a march begin at a certain time during the day. The system is interactive and dynamic. That is, the advice given depends on the input provided and will change as the user revises his information base about the mission under consideration.

LES is designed initially for the commander in training. During training the commander will be presented with facts related to a combat mission as a classroom exercise. He will then use the expert system to define the mission, to determine the outcomes of several alternative mission scenarios, and in planning operational and logistical support. Through this process, the commander will get immediate feedback about the consequences of his decisions. As the commander works with the system, he becomes more expert at predicting the consequences of his decisions. Ideally, he will progress to the point where he can make expert decisions without the requirement for LES. Just as the commander in training will become better at making load configuration and other mission related decisions, LES will also grow. As we acquire more knowledge about factors that impact on the success of missions of the infantry soldier, we can expand the capabilities of LES. Eventually, commanders may want to have LES with them in the field to aid them in real-life decision-making.

This gradual growth of expert systems is a natural process for AI applications. Instead of writing and rewriting new versions of old programs, as occurs with more conventional software like word processors, AI applications, which are heavily rule-based, can be added to in a gradual manner as new information becomes available. Also, individual rules can be changed as part of the refinement process. The other side of this feature of easy, gradual improvement, is that users get in the habit of making suggestions and expecting those changes to be implemented quickly.

The most recent models of human performance show an increasing capability for providing excellent estimates of human performance in an ever widening range of situations. Expert system technology is emerging as the most useful area in the AI arena. By combining the USARIEM heat stress model with the technology of expert systems, we have produced an application which could be of significant benefit to both foot soldiers and their commanders.

References

1. Aghazadeh, F. and Ayoub, M.M., "A Comparison of Dynamic- and Static-Strength Models for Prediction of Lifting Capacity.", Ergonomics, Vol. 28, No. 10, pp. 1409-1418, Oct., 1985.

2. Genaidy, A.M., Asfour, S.S., Mital, A., and Tritar, M., "Psychophysical Capacity Modeling in Frequent Manual Materials Handling Activities.", Human Factors, Vol. 30, No. 3, pp. 319-337, 1988.

3. Giambra, L.M. and Quilter, R.E., "A Two-Term Exponential Functional Description of the Time Course of Sustained Attention.", Human Factors, Vol. 29, No. 6, pp. 635-643, 1987.

4. Haslam, R.A. and Parsons, K.C. "A Comparison of Models for Predicting Human Response to Hot and Cold Environments." Ergonomics, Vol. 30, pp. 1599-1614, 1987.

5. Pandolf, K.B., Stroschein, L.A., Drolet, L.L., Gonzalez, R.R., and Sawka, M.N., "Prediction Modeling of Physiological Responses and Human Performance in the Heat.", Computers in Biology and Medicine, Vol. 16, pp. 319-329, 1986.

6. Marshall, S.L.A., The Soldier's Load and the Mobility of a Nation, Washington, D.C.: The Combat Forces Press, 1950.

7. Military Standard 1472C, pp. 209, 1981; Notice 1, 1983; Notice 2, 1984; Notice 3, 1987.

8. Sampson, J., Technology Demonstrations for Lightening the Soldier's Load, TR-88/027L, U.S. Army Natick Research, Development and Engineering Center, 1988.

9. Kennedy, S.J., Goldman, R., and Slauta, J., The Carrying of Loads Within an Infantry Company, Natick Technical Report No. 73-51-CE, C & PLSEL-108, U.S. Army Natick Research, Development, and Engineering Center, Natick, MA, 1973.

10. Nordeen, L.O., "Today's Infantryman.", National Defense, pp. 35-40, January, 1986.

11. Dubik, J.M. and Fullerton, T.D. "Soldier Overloading in Grenada.", Military Review, pp. 38-47, Jan., 1987.

12. Barnett, O., Director of Laboratory of Computer Science, Project DXplain, Mass. General Hospital, 1987.

13. Campbell, A.N., Hollister, V.F., Duda, R.O., and Hart, P. "Recognition of a Hidden Mineral Deposit by an Artificial Intelligence Program.", Science, Vol. 217, pp. 927-929, Sept. 1982

14. Buchanan, B.G. and Feigenbaum, E.A. "Dendral and Meta-Dendral.", Artificial Intelligence, Vol. 11, pp. 5-24, 1978

15. Gevarter, W.B., "Expert Systems: Limited but Powerful", in S.J. Andriole (ed.), Applications in Artificial Intelligence, pp. 125-139, Princeton, NJ: Petrocelli Books, Inc., 1985.

16. Lechner, P.E, and Barth, S.W., "Expert Systems on Microcomputers", in S.J. Andriole (ed.) Applications in Artificial Intelligence, pp. 109-124, Princeton, NJ: Petrocelli Books, Inc., 1985.

17. Siegel, S.F., and Warren, P.H., Designer Compacts: Artificial Intelligence Strategies, Procedings of the Research and Development Associates, Natick R,D and E Center, Natick, MA, 39(1) pp. 183-190, October, 1987.

18. Berlin, H.M., Stroschein, L., and Goldman, R.F., A Computer Program to Predict Energy Cost, Rectal Temperature, and Heart Rate Response to Work, Clothing, and Environment, ED-SP 75011, Department of the Army, Edgewood Arsenal, MD, 1975.

19. Givoni, B., and Goldman, R.F., "Predicting Rectal Temperature Response to Work, Environment, and Clothing.", Journal of Applied Physiology, Vol. 32, pp. 812-822, 1972.

20. Givoni, B. and Goldman, R.F., "Predicting Effects of Heat Acclimatization on Heart Rate and Rectal Temperature.", Journal of Applied Physiology, Vol. 35, pp. 875-879, 1973.

21. Soule, R.G., Pandolf, K.B., and Goldman, R.F. "Energy Expenditure of Heavy Load Carriage.", Ergonomics, Vol. 21, no.5, pp. 373-381, 1978.

22. Tullis, T.S. "Predicting the Usability of Alphanumeric Displays.", Doctoral Thesis, Rice University, 1983.

THE USE OF COMPETITIVE MARKSMANSHIP AS A STRESSOR IN SOLDIER/EQUIPMENT PERFORMANCE TESTING

Fatkin, L.T.[1], Hudgens, G.A.[1], King, J.M.[1], Torre, J.P., Jr.[1], Wansack, S.[1],
Mazurczak, J.[1], Myers, J.S.[1], Slager, S.E.[2], and Chatterton, R.T., Jr.[2]

[1]Behavioral Research Division, U.S. Army Human Engineering Laboratory,
Aberdeen Proving Ground, MD 21005-5001; [2]Northwestern University Medical
School, Chicago, IL 60612

The development of a research data base on human behavior and performance is essential to the formulation of valid combat models. If the area of focus is directed towards accounting for the human factor in battle outcomes, then there is a need to research aspects of behavior that are affected by the manipulation of variables which would require soldiers to perform under stressful conditions. However, stress researchers are no longer acting under the assumption that certain environmental conditions always induce predictable physiological or psychological effects on performance (1). Different individuals respond to the same conditions in different ways. Some show almost immediate decrements in performance in response to the stress, others show increased alertness and apparently improved performance, and still others appear to be "immune" to the qualities of the stress conditions (2). We can no longer predict an individual's response to stress by simply identifying the stressor. The type of stress, the intensity of the stressor, and time of measurement must be considered, along with those personal factors that might account for the individual variability in response to stress.

This paper describes the results of a recent field study (Salvo Stress Study) investigating the use of competition as a stressor in soldier/equipment performance testing. Performance data were obtained with respect to the soldier's use of the M16A2 rifle system in support of the projected Advanced Combat Rifle field evaluation. The Salvo Stress Study is part of an innovative, comprehensive stress research program (3) initiated at the U.S. Army Human Engineering Laboratory (HEL) with four specific goals in mind: (a) to verify the notion that different kinds and levels of stress, interacting with personal factors, yield unique physiological and psychological response profiles; (b) to create, in effect, a collection of such profiles against which other stressors can be evaluated; (c) to determine which combination of physiological and psychological indices might be most efficiently and effectively used to measure stress experienced by subjects; and (d) to develop a data base which would support modeling of combat stress by quantifying data using combat-like stressors. This would include integrating the effects of multiple and chronic stressors.

A fundamental contribution of this program will be the development of SOP's for testing new soldier/machine systems under combat-like conditions, thereby obtaining data relevant for systems analysts and operations researchers. The Salvo Stress Study represents a pilot effort to test the notion that competition could be used in this way.

Methods

The Salvo Stress Study was conducted at a highly instrumented small arms range located at Aberdeen Proving Ground, MD. The subjects were 60 volunteer infantrymen, 40 from the 82nd Airborne Division and 20 from the 101st Airborne Division (Air Assault).

Firing took place over a three-week period of time. The first and third weeks were competition weeks during which the soldiers from each unit competed for a highly coveted trophy. The intensity of the competition was enhanced through peer observations. Performance scores were displayed on a scoreboard for all competitors to see. All firing was conducted under the scrutiny of a video taping system placed behind the firing point. The second week was a control week during which no formal competition occurred. The video camera, recorder, and scoreboard were removed from the range for this week, and scores were not publicly announced.

Performance measures were obtained using modified M16A2 rifles. Soldiers fired for record in both the semiautomatic and three-round burst modes. The total number of targets (possible hits) was 72 for each mode.

Demographic Information

A General Information Questionnaire was administered to all subjects to obtain pertinent demographic and medical information, such as age, pay grade, length of service, education level, physical profile status, current use of prescription or non-prescription drugs, current weapons qualifications (number of weapons), last weapons qualification achieved (sharpshooter, marksman, expert), total rounds fired, and specialized training (sniper school, formal small arms training, etc.).

Physiological Measures

Hormonal data were obtained on a familiarization firing day prior to weapon firing (Baseline Day) and on the subsequent Record-Fire Day. A total of ten blood samples was obtained from each of the 60 subjects by way of an indwelling, intravenous catheter. Blood samples were staggered relative to each subject's 15-min anticipated firing interval as indicated in Figure 1. Hormonal data were collected on the catecholamines, cortisol, and other hormones more recently shown to be responsive to various stressors (growth hormone, luteinizing hormone, testosterone, and prolactin).

Psychological Measures

Psychological measures were obtained using a battery of standardized questionnaires designed to assess the subjects' personality traits, coping strategies, and perceptions of stress. Personality measures were obtained two days before the competitive weapons firing was scheduled. The stress perception and coping measures were completed by the subjects 10 minutes before and after the firing interval, as indicated in Figure 1. The measures given just prior to the firing interval assessed how the subjects were feeling "right now," while the post measures instructed them to rate how they felt "during" the firing event/interval. Each of the measures used is designed to be self-administering, relatively brief, and easily given individually or to groups.

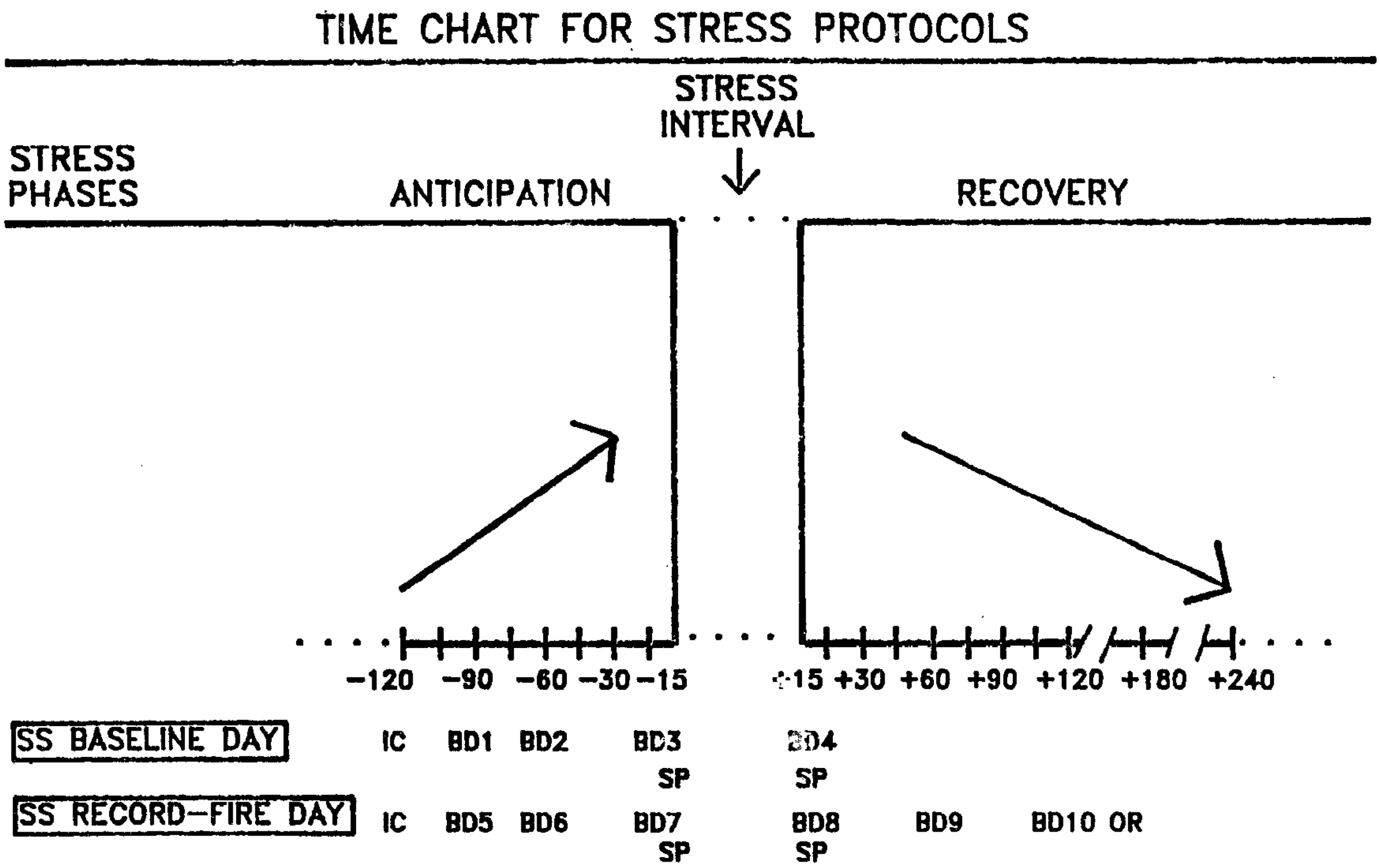


Figure 1. Time chart for stress protocols indicating insertion of catheter (IC), time of blood draws (BD1-BD10), stress perception measures (SP), and overall rating of events (OR).

Perceptions on Unit Cohesion and Morale. Measures of unit perceptions included the following questionnaires:

1. The Unit Cohesion and Morale Questionnaire (4) required individuals to rate their perception of their unit's level of morale, readiness for combat, and their confidence in their leaders and weapon systems, using a five-point scale.
2. The Squad/Platoon Perceptions Questionnaire (4) asked individuals to indicate (on a five-point scale) the extent to which they agreed or disagreed with statements concerning squad/platoon members and their leaders.

Trait measures. The following trait measures were used:

1. The State-Trait Anxiety Inventory (STAI) Form Y-2 (5), which consists of 20 statements that assess how the respondents "generally" feel.
2. The Multiple Affect Adjective Check List - Revised (MAACL-R), General form (6). This General or Trait form consists of five primary subscales (Anxiety, Depression, Hostility, Positive Affect, and Sensation Seeking) derived from a one-page list of 132 adjectives. The respondents were instructed to check all the words which described how they "generally" feel.
3. The Sensation Seeking Scale (SS), Form V (7) which contains four subscales (Thrill and Adventure Seeking, Experience Seeking, Disinhibition, and Boredom Susceptibility). Respondents were presented with a 40-item, forced-choice questionnaire that was titled, "Interest and Preference Survey." A "Total" score was used which was based on the sum of the four subscale scores.
4. Rotter's Internal-External Scale (8) was used as a measure of Locus of Control. Respondents were asked to complete 29 forced-choice items (including six "filler" statements) relating to their locus of control beliefs.
5. The Eysenck Personality Questionnaire (EPQ), which recognizes three distinct dimensions of personality: Extraversion-Introversion, Neuroticism, and Psychoticism (9).

Coping measures. Measures of coping included:

1. The Revised Ways of Coping Checklist (RWCCL) (10, 11), which identified five individual coping efforts: problem-focused thoughts or behaviors, seeking social support, wishful thinking, blaming self, and avoidance.
2. A Coping Efficacy scale which asked respondents to rate (from 0 to 10) their level of confidence in their ability to do well. This scale was adapted from a self-efficacy scale developed by Bandura (12) for investigating the predictive power of efficacy expectations on behavior or performance.

State measures. A five-minute battery of state measures was given immediately before and after the competitive firing on Record-Fire Day and before and after a comparative interval on the previous Baseline Day. The battery included:

1. Form Y-1 (State Form) of the STAI (5).
2. The Today Form of the MAACL (6).
3. The Subjective Stress Scale (SSS) which was developed by Kerle and Bialek (13) to detect significant affective changes in stressful situations.
4. The Specific Rating of Events scale (SRE) was a post measure designed for this program. The subjects rated (on a scale of 0-100) how stressful the event was to them.

Results and Discussion

Multivariate Analyses of Variance (MANOVAs) indicated that there were significant group differences on both the physiological and psychological measures. Figures 2 through 11 present mean responses (± standard error of the mean, SEM) for these variables. Unless otherwise stated, post-hoc tests of significance relating to data presented in these figures are based on the following criteria: for $p<0.05$, the difference between two means must exceed 1.96 SEM; and for $p<0.01$, the difference must exceed 2.58 SEM (14).

Group Differences

Demographic Data

There were no significant differences between the Control and Competition Groups on demographic factors such as age and rank; nor were differences found on variables reflecting type and amount of military experience, such as length of service and current weapons qualifications.

Physiological Data

Because of the extensive amount of data collected, the results reported will be limited to cortisol and testosterone, as traditional stress response hormones that are differentially effected by stressors. A complete report of the physiological data can be found in the HEL report presently in preparation (15).

Cortisol. Preliminary results indicated that cortisol reflects a stress-response pattern. Figure 2 shows the blood cortisol levels of the competition and control groups observed on Baseline Day and Record-Fire Day. For Baseline Day, the MANOVA yielded no significant differences between the groups. On Record-Fire Day, there was a significant Groups X Time Point interaction ($p<0.003$). The competition group showed a large positive increase in cortisol levels at 15 minutes following firing, followed by a large rebound below baseline levels. This pattern is characteristic of a stress response. The control group showed a much smaller change in cortisol in response to firing, and no rebound was observed.

Testosterone. The groups did not differ significantly in testosterone levels at any of the 10 time points, nor did their patterns of response differ significantly over those time points.

Because of the exploratory nature of this study, the hormone data were also analyzed using change values from Baseline Day to Record-Fire Day, at the four common time points. With each subject as his own control, change values were computed for each hormone by subtracting Baseline Day values from Record-Fire Day values.

The mean group changes in testosterone are shown in Figure 3. MANOVA of the testosterone change data yielded a significant Groups main effect ($p<0.02$), reflecting the overall decrease in testosterone for the Competition group from Baseline to Record-Fire Day as compared with Controls. This decrease reflects the typical stress response for testosterone. Only the Competition group's change in testosterone from Baseline to Record-Fire Day

Figure 2

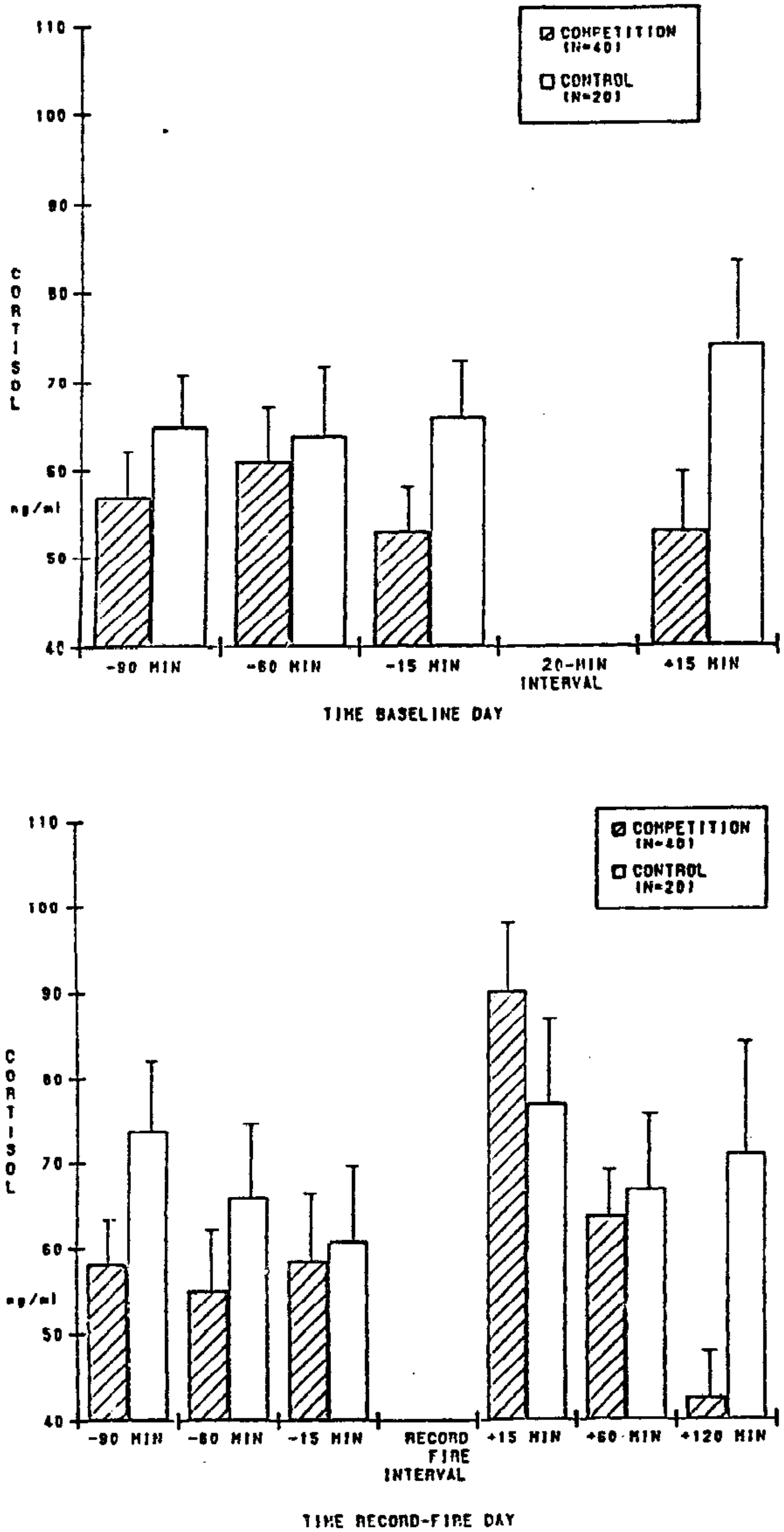
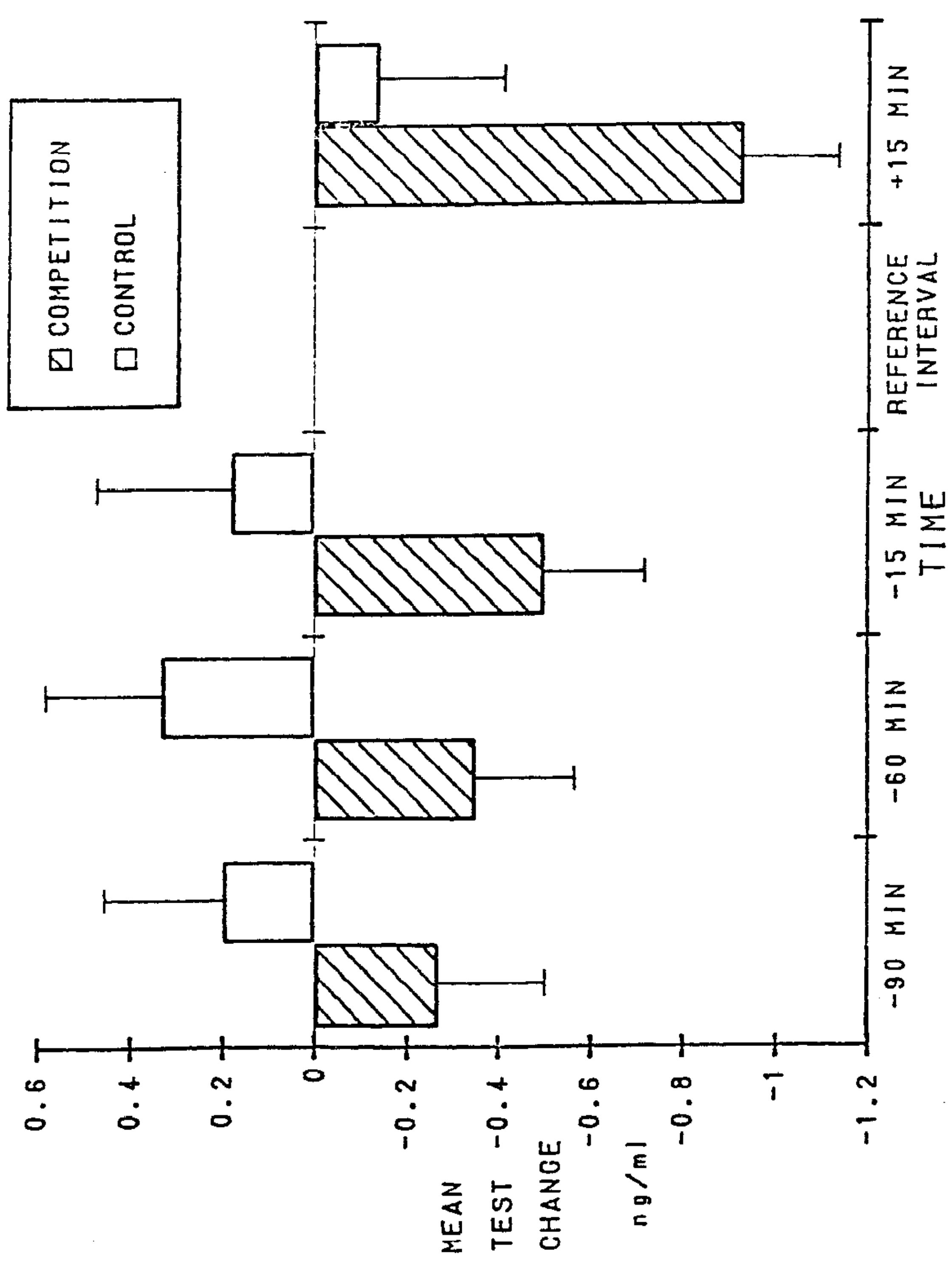

110
100
90
80
70
60
50
40
CORTISOL
ng/ml
▨ COMPETITION (N=40)
☐ CONTROL (N=20)
-90 MIN
-60 MIN
-15 MIN
20-MIN INTERVAL
+15 MIN
TIME BASELINE DAY
110
100
90
80
70
60
50
40
CORTISOL
ng/ml
▨ COMPETITION (N=40)
☐ CONTROL (N=20)
-90 MIN
-60 MIN
-15 MIN
RECORD FIRE INTERVAL
+15 MIN
+60 MIN
+120 MIN
TIME RECORD-FIRE DAY

Figure 3


COMPETITION
CONTROL
MEAN TEST CHANGE ng/ml
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-90 MIN
-60 MIN
-15 MIN
REFERENCE INTERVAL
+15 MIN
TIME

deviated significantly from no change ($p<0.05$ at -15 min and $p<0.01$ at +15 min).

Comparative measures. In Figure 4, the Salvo Stress Study hormone values at +15 min are compared with values obtained through an ongoing contractual effort at Northwestern University (16). The first three groups consist of spouses of patients undergoing abdominal surgery (under general anesthesia) for infertility, a group of medical students taking an important written examination, and a combined control group. The last two groups shown are the Control group and the Competition group from the Salvo Stress Study.

The Competition group mean cortisol response was significantly greater than that for the Combined Control and the Moderate Surgery groups, but not significantly different from the Written Exam group. The SS Control group's mean cortisol response, however, was significantly less than that for the Written Exam group, but not significantly different from the Moderate Surgery or Combined Control groups.

The Competition and SS Control groups' mean testosterone responses both were significantly greater than those for the Moderate Surgery and Combined Control groups, but were not significantly different from the mean responses for the Written Exam group or each other.

Psychological Data

Unit Cohesion and Morale. There were no significant differences in unit perceptions concerning leadership qualities, level of morale, readiness for combat, or level of unit confidence between the 101st and the 82nd Airborne Divisions.

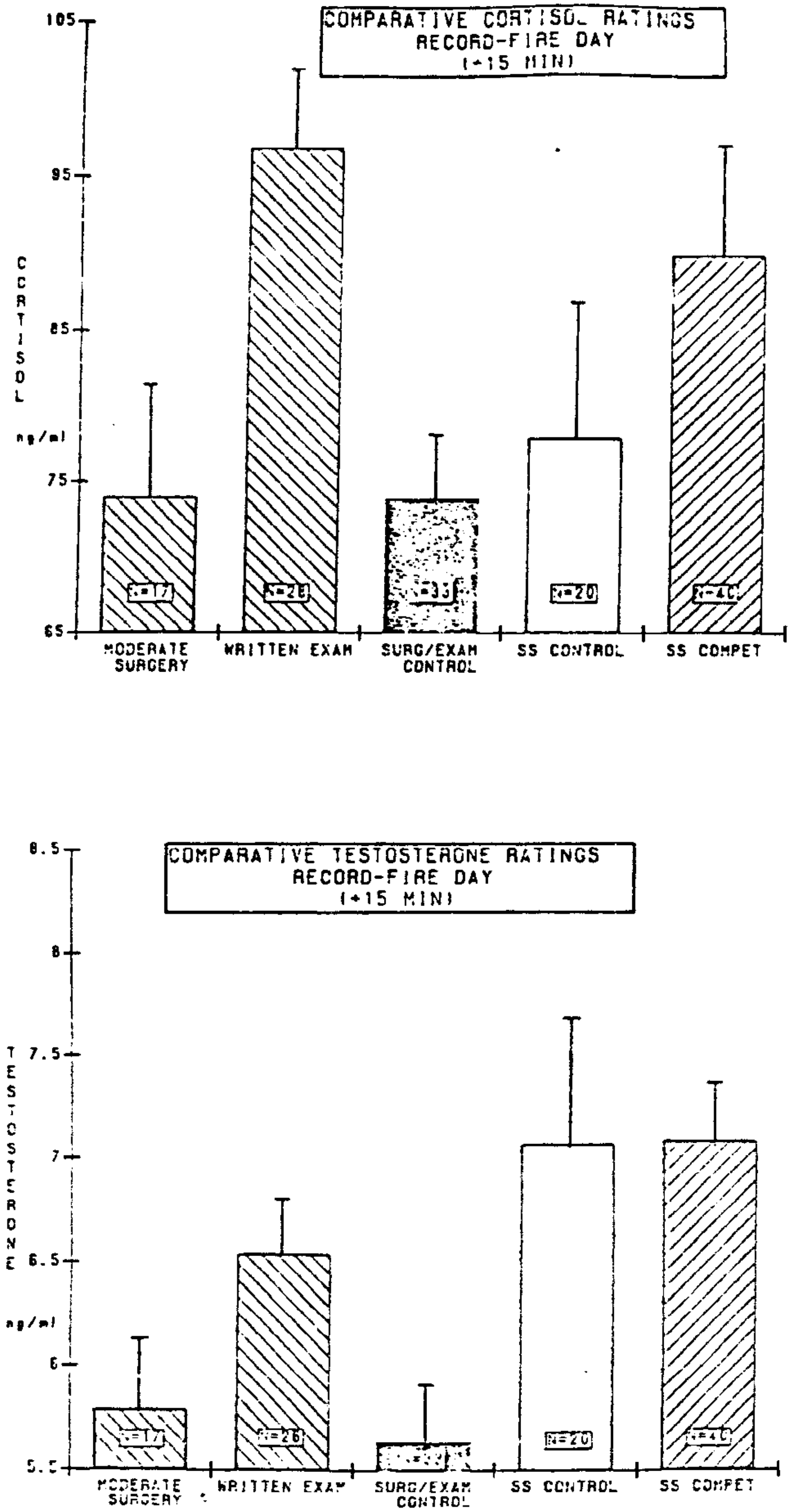
Trait measures. A MANOVA indicated that there were no significant differences between the Competition group and the Control group on the trait measures used.

State measures. A MANOVA indicated no significant differences in stress perception between the two groups on Baseline Day. However, on Record-Fire Day, there were significant differences between the groups on several of the measures. Once again, because of the extensive amount of data obtained from the various measures, results will be presented on two primary subscales of the MAACL, Anxiety and Hostility. Other measures of anxiety from the STAI, SSS, and SRE reflected parallel results and patterns.

Figure 5 demonstrates the Anxiety and Hostility ratings of both groups for Baseline Day and for Record-Fire Day. On Record-Fire Day, the Competition group reported significantly higher Anxiety than the Control group in both the pre (10 min before firing) and post (10 min after firing) time periods ($p<0.001$).

On Record-Fire Day both groups reported significantly higher hostility after firing compared with other times ($p<0.05$). Furthermore, this effect was significantly greater for the Competition group than the Control group ($p<0.001$). This finding is particularly interesting, since the increased

Figure 4


COMPARATIVE CORTISOL RATINGS
RECORD-FIRE DAY
(+15 MIN)
CORTISOL
ng/ml
105
95
85
75
65
N=17
N=26
N=33
N=20
N=40
MODERATE SURGERY
WRITTEN EXAM
SURG/EXAM CONTROL
SS CONTROL
SS COMPET
COMPARATIVE TESTOSTERONE RATINGS
RECORD-FIRE DAY
(+15 MIN)
TESTOSTERONE
ng/ml
8.5
8
7.5
7
6.5
6
5.5
N=17
N=26
N=20
N=40
MODERATE SURGERY
WRITTEN EXAM
SURG/EXAM CONTROL
SS CONTROL
SS COMPET

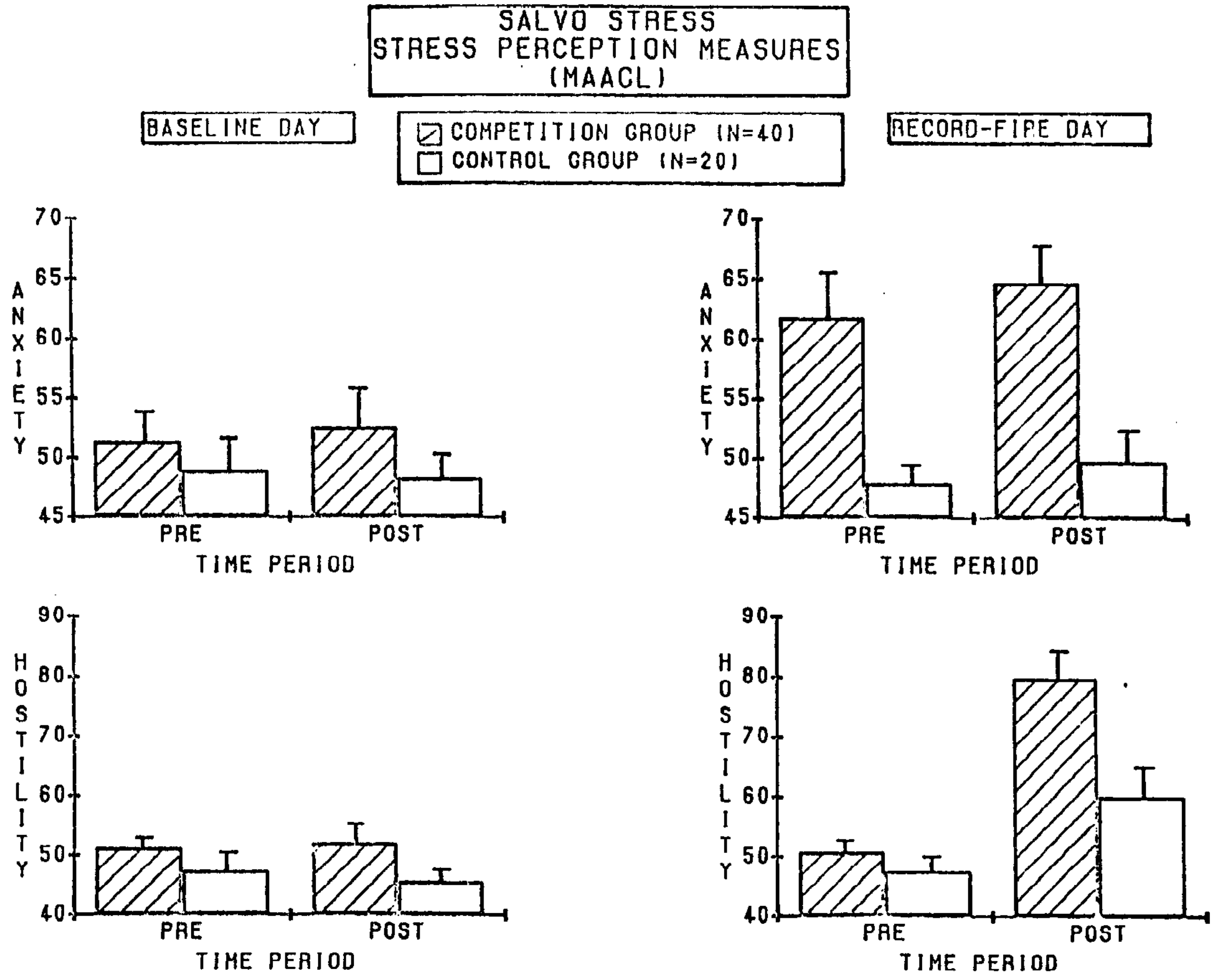

SALVO STRESS
STRESS PERCEPTION MEASURES
(MAACL)
BASELINE DAY
▨ COMPETITION GROUP (N=40)
☐ CONTROL GROUP (N=20)
RECORD-FIRE DAY
ANXIETY
70
65
60
55
50
45
PRE
POST
TIME PERIOD
HOSTILITY
90
80
70
60
50
40

Figure 5

hostility experienced during the record weapon firing may have interfered with performance.

Comparative measures. Figure 6 illustrates how the anxiety experienced by the soldiers in the Salvo Stress Study compares with results obtained on the same measure for the conditions studied at Northwestern University. These data indicate that the Competition group had a level of anxiety associated with a moderate level of stress (about comparable to taking a written exam).

Figure 7 shows a similar comparison for Hostility ratings. Both groups from the Salvo Stress Study reported significantly higher ratings of Hostility than all other groups, with the Competition group reporting an unusually high rating. The Salvo Stress hostility ratings are similar to those expressed by medical students taking an important written exam. While increases in hostility do not seem to be part of the anticipatory response, they do occur immediately afterward, possibly in response to negatively perceived aspects of the outcome.

Correlations with Performance

Pearson correlation coefficients were computed for the Salvo Stress groups between the performance measures (number of targets hit in the Semi-automatic and Burst modes) and the military experience, physiological, psychological, and coping measures. Performance was measured by the number of targets hit in each mode. The total number of possible hits was 72 per mode.

Military Experience

Since no significant differences between groups were found for the military experience variables, Pearson correlation coefficients were computed between these variables and performance for the combined Salvo Stress groups. As shown in Table 1, both length of service and current weapons qualifications correlated significantly and positively with performance. While a longer length of service was associated with better Burst mode performance, the number of weapons for which the soldier was currently qualified was positively correlated with Semi-automatic mode performance.

Performance and Physiological Responses

Separate Pearson correlation coefficients were computed for the Competition and Control Groups. Correlations were obtained between Record-Fire Day performance scores and the various hormone values for all 10 time points over both the Baseline and Record-Fire Days.

For the Competition group, it appears that performance is most closely associated with levels of testosterone. Better performance within the Competition group in the Burst mode was predicted by relatively lower levels of testosterone on Baseline Day at all time points and on Record-Fire Day in the early morning, 90 min before firing. As illustrated by the mean change values in Figure 3, there was an overall decrease in testosterone for the Competition group from Baseline to Record-Fire Day as compared with Controls. Better performance in the Semi-automatic mode was predicted by this relatively greater decrease in testosterone at -90 min (see Table 1).

Figure 6

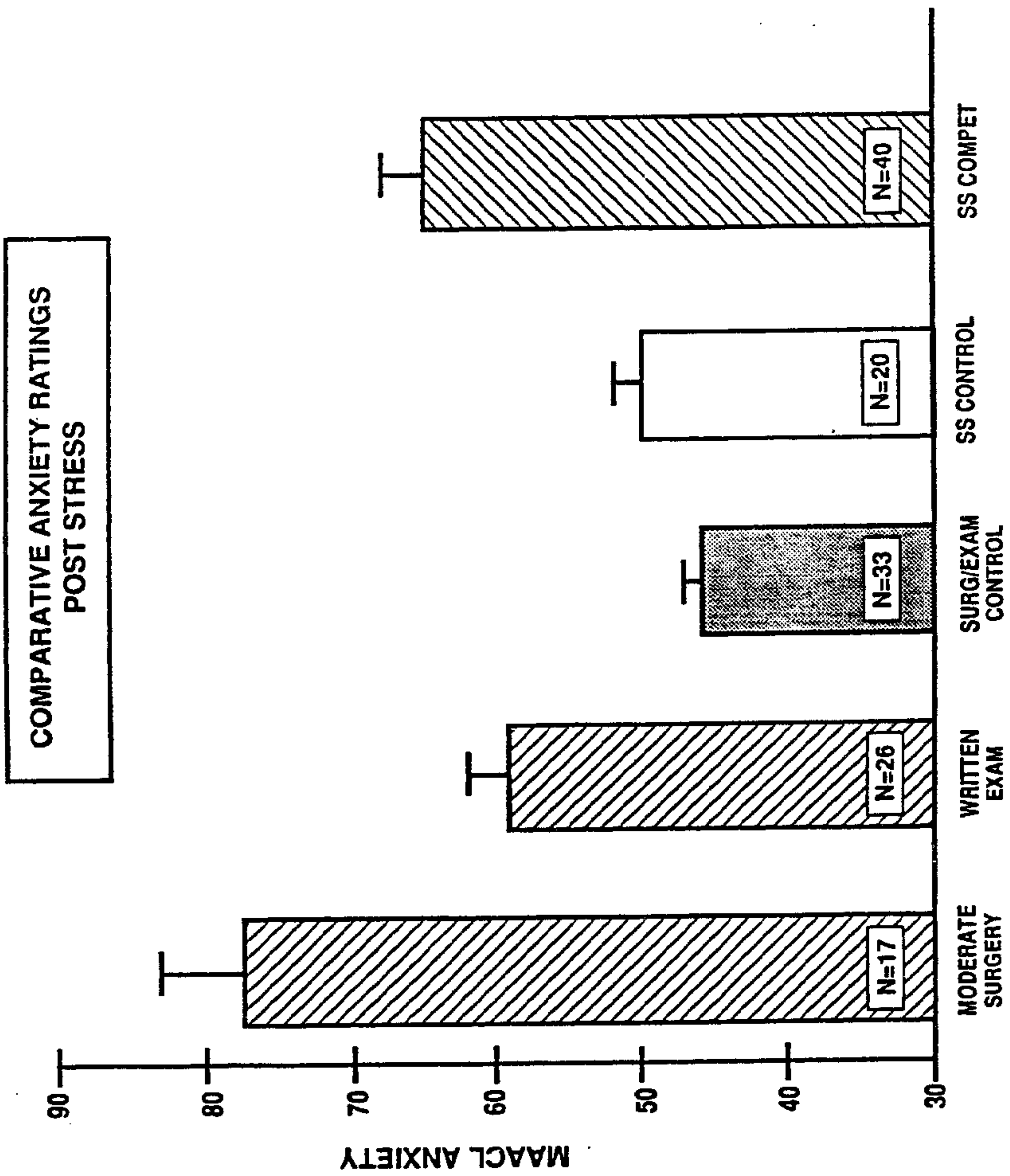

COMPARATIVE ANXIETY RATINGS
POST STRESS
MAACL ANXIETY
90
80
70
60
50
40
30
N=17
MODERATE SURGERY
N=26
WRITTEN EXAM
N=33
SURG/EXAM CONTROL
N=20
SS CONTROL
N=40
SS COMPET

Figure 7

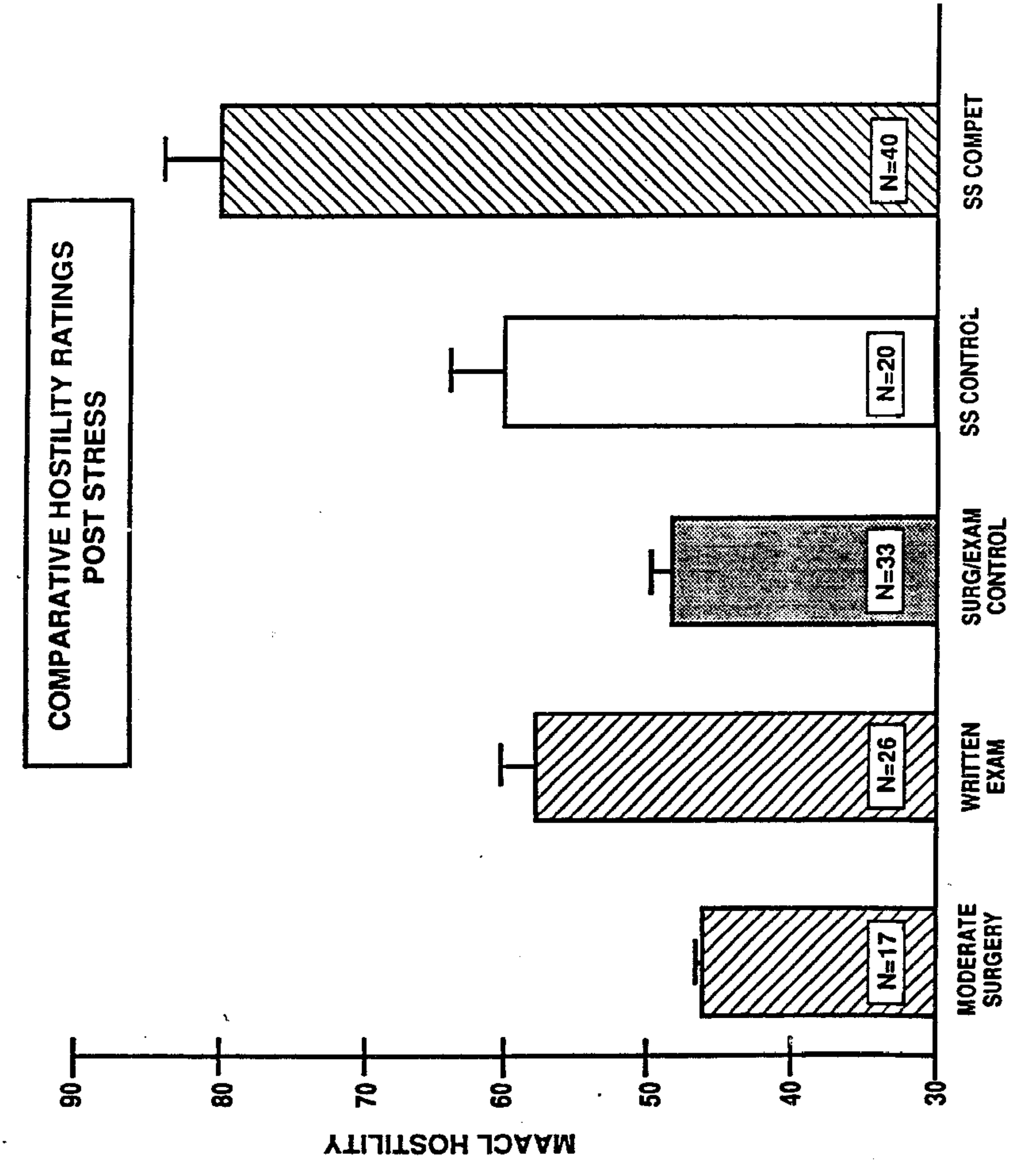

COMPARATIVE HOSTILITY RATINGS
POST STRESS
MAACL HOSTILITY
90
80
70
60
50
40
30
N=17
MODERATE SURGERY
N=26
WRITTEN EXAM
N=33
SURG/EXAM CONTROL
N=20
SS CONTROL
N=40
SS COMPET

TABLE 1

Significant Correlations with Performance

MEASURE	GROUP	CORRELATIONS WITH PERFORMANCE (SM[a], BM[b])
MILITARY EXPERIENCE	ALL (N=60, df=58)	Length of Service/BM, r = 0.33* Current Weapons Qualif/SM, r = 0.26*
PHYSIOLOGICAL MEASURES Blood Sampling Time -90 Min	COMPETITION (N-40, df=38)	T[c]/SM, r = -0.38*
+15 Min	COMPETITION (n=40, df=38)	T/BM, r = 0.39*
	CONTROL (n=20, df=18)	T/BM, r = 0.55*
TRAIT MEASURES	ALL (N=60, df=58)	MAACL Depression/SM, r = -0.31* MAACL Hostility/SM, r = -0.29*
COPING MEASURES	CONTROL (N=20, df=18)	AVOID/BM, r = -0.46*
STATE MEASURES	COMPETITION (N=40, df=38)	MAACL Sensation Seeking/SM, (post) r = 0.55** MAACL Sensation Seeking/BM, (post) r = 0.41**
	CONTROL (N=20, df=18)	MAACL Hostility/BM, (post) r = -0.46*

---

[a] SM = Targets Hit in Semi-automatic Mode
[b] BM = Targets Hit in Burst Mode
[c] T = Testosterone

* = $p < .05$
** = $p < .01$

For the Control group, it appears that better performance in the Semi-automatic mode is associated with increased testosterone from Baseline to Record-Fire Day. This was most pronounced for performance in the Burst mode at +15 min after firing (see Table 1).

Performance and Psychological Responses

Trait measures. Pearson correlation coefficients were also computed for the combined Salvo Stress groups between the personality measures and Record-Fire Day performance scores. As indicated in Table 1, trait Depression and trait Hostility were both significantly and negatively correlated with performance in the Semi-automatic mode of fire. In other words, individuals who scored higher on the trait Depression and Hostility subscales did not perform as well in the Semi-automatic mode.

Stepwise regression analyses using Semi-automatic mode (SM) and Burst mode (BM) performance as separate outcome variables and trait measures as the common predictor variables were performed for all subjects. The regression on SM performance yielded a multiple R of 0.384, accounting for 14.8% of the variance ($F(2,57)=4.931$, $p=0.011$). The final model, which included trait Depression (TDEPR) and the Boredom Seeking subscale of the Sensation Seeking Scale (BS) as the dominant predictors of SM performance, is given below.

$$\text{SM PERF} = 28.298 - 0.362*\text{TDEPR} - 0.233*\text{BS}$$

The final model used in predicting BM performance was not statistically significant.

Coping measures. Separate Pearson correlation coefficients were computed for the Competition and Control groups between the coping measures and performance scores. As indicated in Table 1, a significant negative correlation was found between the Avoidance subscale and Burst mode performance for the Control group only. Those individuals who reported sleeping whenever they could and generally avoiding others did not perform as well as those who were more situationally oriented.

Separate stepwise regression analyses were conducted for the two groups, again using Semi-automatic mode (SM) and Burst mode (BM) performance as separate outcome variables with the coping measures as the common predictor variables.

For the Competition group, the regression analysis on BM performance yielded a multiple R of 0.510, accounting for 26% of the variance ($F(3,36)=4.210$, $p=0.012$). The final model, which included Coping Efficacy (COPEFF), Problem-focused coping (PFOC), and Blaming-self (BLAME) as the dominant predictors of BM performance, is given below.

$$\text{BM PERF} = 20.446 - 0.529*\text{PFOC} - 0.453*\text{BLAME} + 0.450*\text{COPEFF}$$

Since the Coping Efficacy measure was completed by the soldiers prior to their record weapon firing, it can be a useful predictor of performance. A high degree of self-efficacy indicates confidence in one's ability to produce a desired effect and is presumed to influence performance level by enhancing the intensity and persistence of effort (12, 17, 18, 19).

For the Control group, the regression on BM performance yielded a multiple $\underline{R}$ of 0.457, accounting for 20.9% of the variance ($\underline{F}(1,18)=4.745$, $\underline{p}=0.043$). The final model, which included Avoidance coping (AVOID) as the dominant predictor of BM performance, is given below.

$$\text{BM PERF} = 24.592 - 0.457*\text{AVOID}$$

The final model used in predicting SM performance was not statistically significant for either group.

State measures. Separate correlation analyses were also performed for the Competition and Control groups for the state measures. As indicated in Table 1, a significant negative correlation was found between post Hostility ratings and Burst mode performance for the Control group. For the Competition group, the post Sensation Seeking ratings were positively correlated with both Semi-automatic and Burst mode performance.

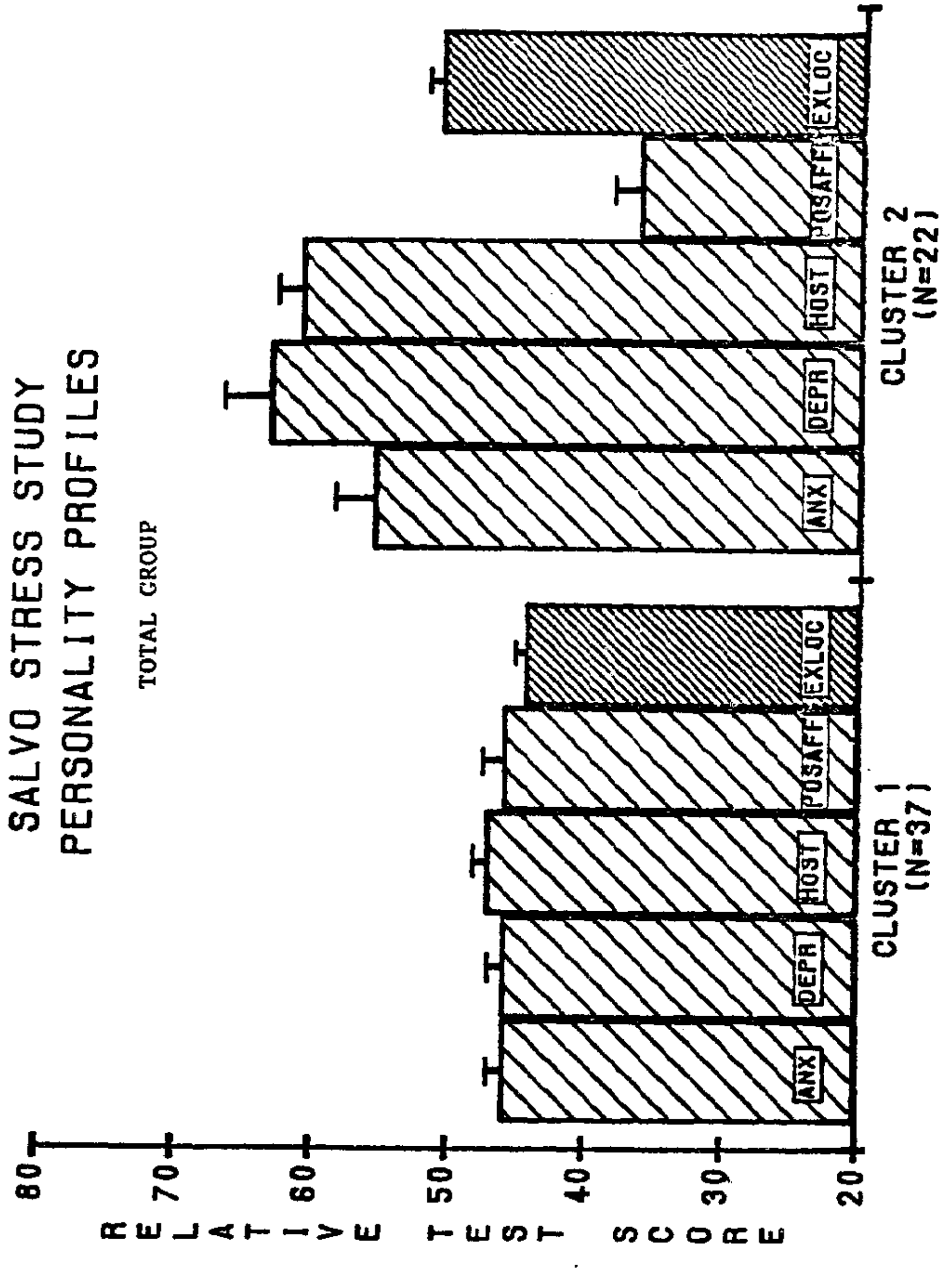
Individual Differences in Psychological Responses

Trait measures. In order to address the effect of individual variability in stress response, cluster analysis was performed using Version 3 of the statistics software package, SYSTAT (20). Cluster analysis is a method of statistically grouping subjects based on the dependent measures (e.g., evaluating whether the subjects tend to fall into groups having similar personality characteristics). It minimizes the variance for each cluster across the measures so that the result is groups or clusters of individuals that are most alike.

When cluster analysis was performed on the personality measures of the entire Salvo Stress group, two distinct clusters of individuals emerged: one with a low-stress profile, and the other with a relatively high-stress profile (Figure 8). Individuals with a trait profile of lower scores on Anxiety, Depression, and Hostility subscales, less External Locus of Control, and higher scores on the Positive Affect subscale, performed slightly better in Semi-automatic mode than those with a profile of higher Anxiety, Depression, Hostility, External Locus of Control, and lower Positive Affect. Individuals in the low-stress profile also reported using Problem-focused coping on Record-Fire Day significantly more often than those in the high-stress profile. It is important to keep in mind that this is a preliminary study, and that there is a need for further assessment of the use of personality profiles for the prediction of performance.

State measures. When cluster analysis was used to assess individual variability in stress perceptions, it was found that within the Competition group, there was a subgroup of individuals (Cluster 1) who did not view the situation as stressful as the other subjects in that group (Clusters 2 and 3, Figure 9). Their lower state scores on Anxiety, Depression, and Hostility measures, and higher Positive Affect scores contrast with the other subjects in that group. It is interesting to note that the profile of this "nonstressed" competition subgroup is similar to that of the Control group (Figure 10). It appears that this competition subgroup did not feel any more

Figure 8


SALVO STRESS STUDY
PERSONALITY PROFILES

TOTAL GROUP

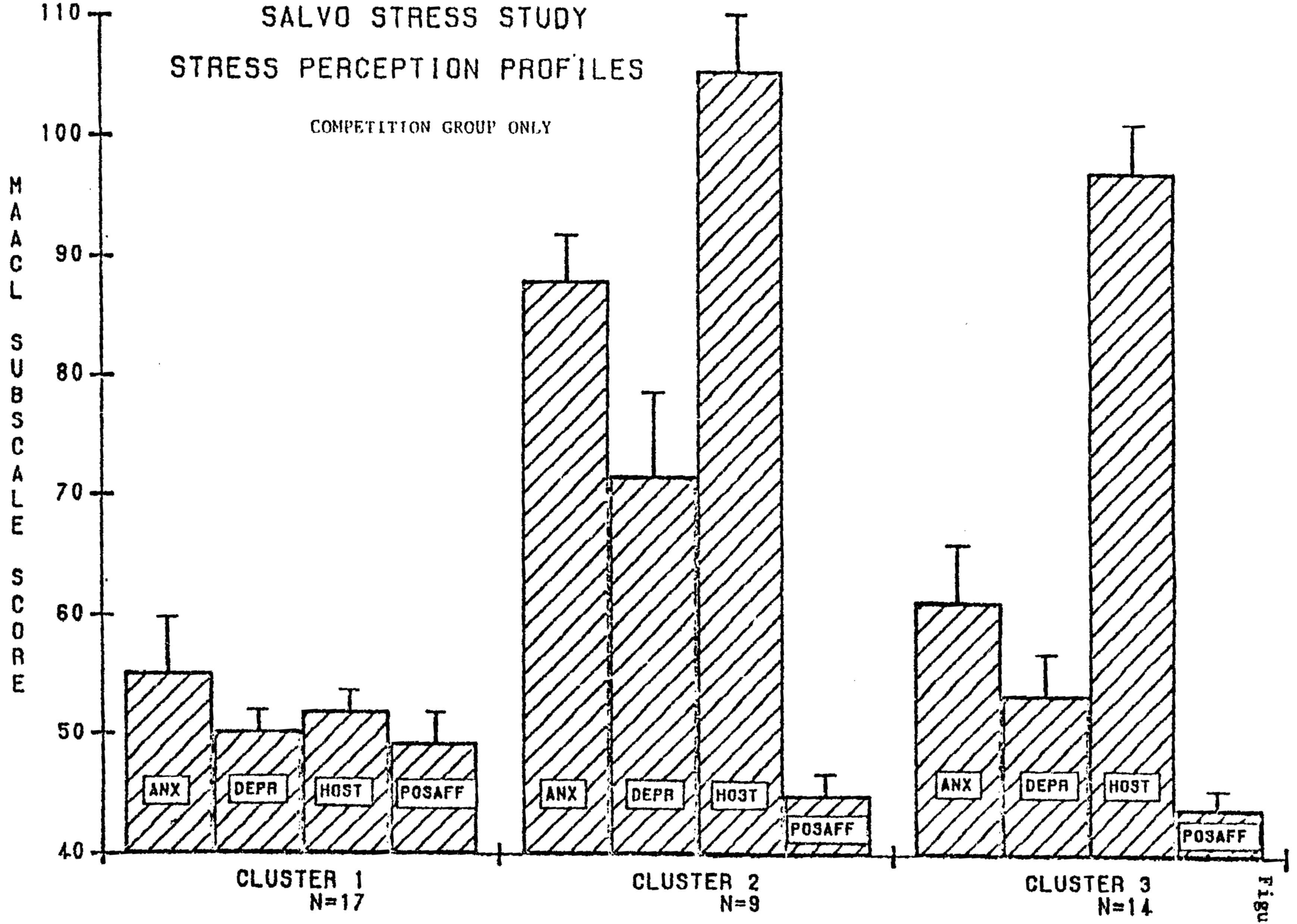
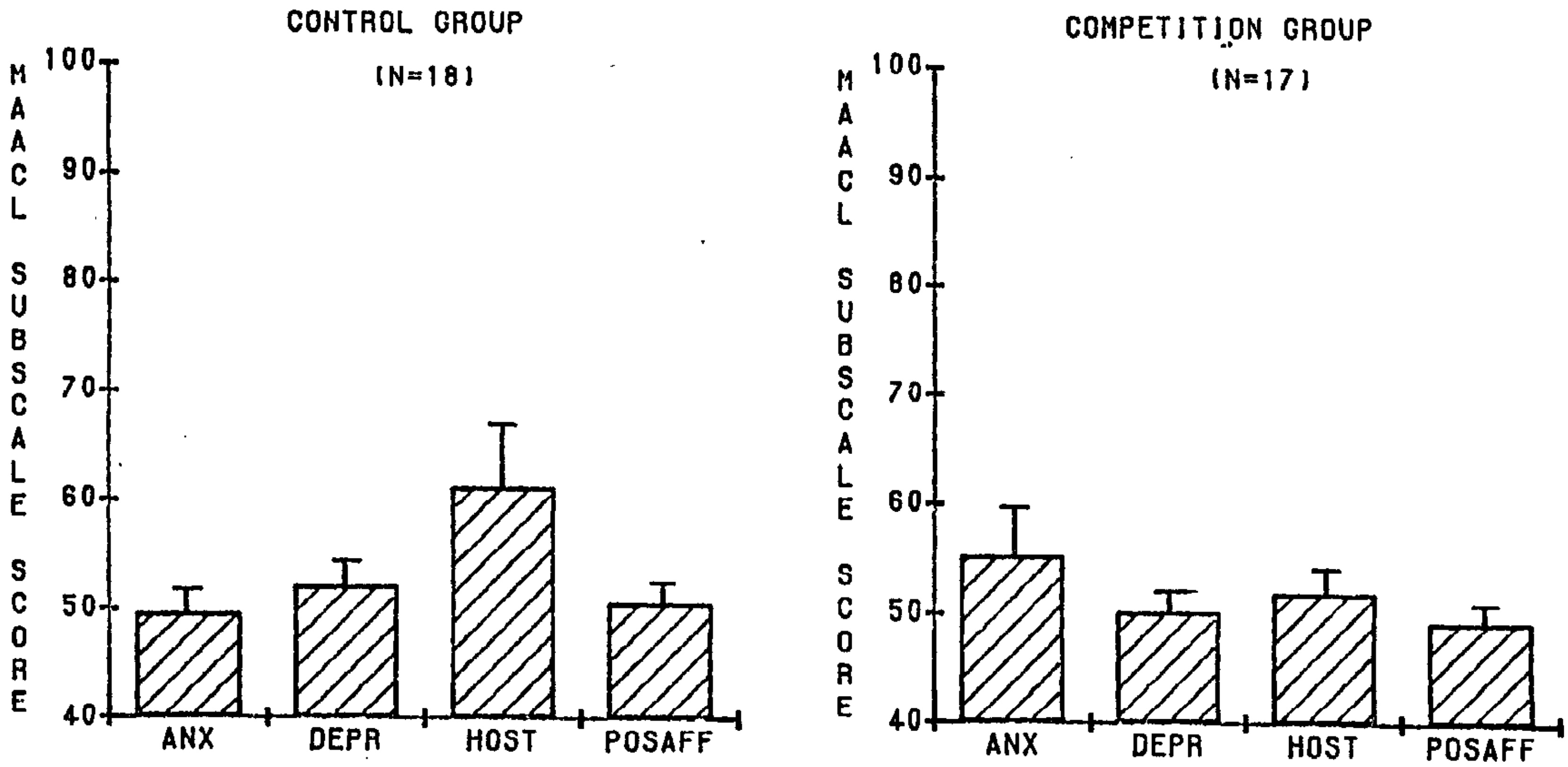

SALVO STRESS STUDY
STRESS PERCEPTION PROFILES
COMPETITION GROUP ONLY
MAACL SUBSCALE SCORE
110
100
90
80
70
60
50
40
ANX
DEPR
HOST
POSAFF
CLUSTER 1
N=17
ANX
DEPR
HOST
POSAFF
CLUSTER 2
N=9
ANX
DEPR
HOST
POSAFF
CLUSTER 3
N=14

Figure 9

Figure 10

SALVO STRESS STUDY

STRESS PERCEPTION PROFILES



CONTROL CLUSTER

COMPETITION CLUSTER I
(low-stress subgroup)

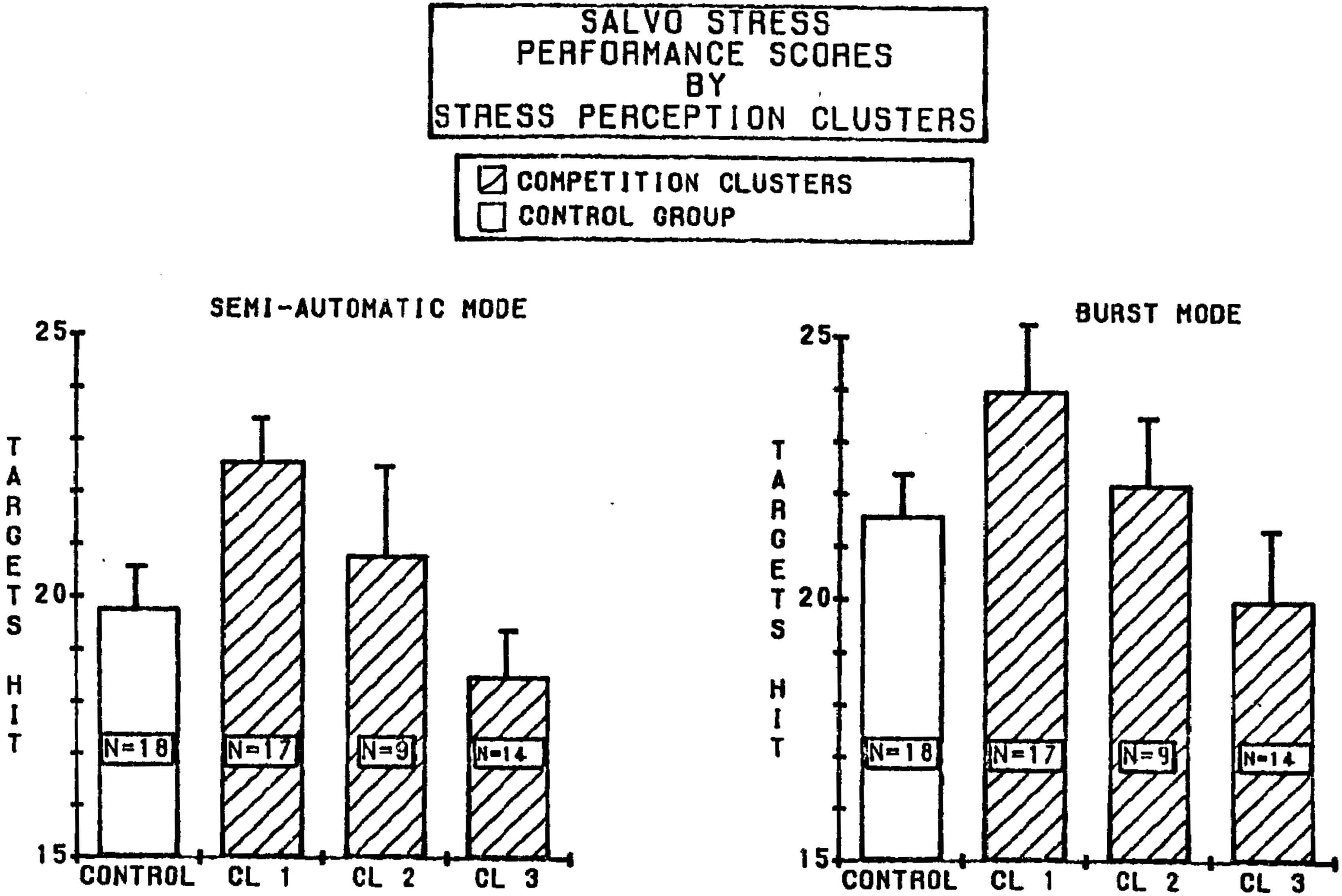

SALVO STRESS
PERFORMANCE SCORES
BY
STRESS PERCEPTION CLUSTERS
▨ COMPETITION CLUSTERS
☐ CONTROL GROUP
SEMI-AUTOMATIC MODE
BURST MODE
TARGETS HIT
25
20
15
CONTROL
CL 1
CL 2
CL 3
N=18
N=17
N=9
N=14

Figure 11

anxious under competitive firing than the control group felt during their comparison firing. As shown in Figure 11, this competition subgroup (Cluster 1) performed significantly better in Semi-automatic mode than the Control group.

The analysis of the performance, physiological, and psychological data from the Salvo Stress Study is still in progress. We have demonstrated that competition can be used to reliably produce a moderate level of stress in soldiers. When these results are considered along with findings from a number of similar studies, we see the emergence of a reliable profile of psychological and hormonal responses to stress which can be related to soldiers performing combat-relevant tasks.

We are continuing to refine our methodology for stressing subjects under combat-like conditions in the evaluation of new weapon systems. These data are essential for realistic incorporation of the role of the individual at multiple organization levels of combat modeling.

References

1. Monat, A., & Lazarus, R.S. (Eds.) (1977). Stress and coping: An anthology. New York: Columbia University Press.
2. Appley, M.H., & Trumbull, R. (1977). On the concept of psychological stress. In A. Monat & R.S. Lazarus (Eds.), Stress and coping: An anthology (pp. 58-66). New York: Columbia University Press.
3. Hudgens, G.A., Torre, J.R., Jr., Chatterton, R.J., Jr., Wansack, S., Fatkin, L.T., & DeLeon-Jones, F.A. (1986). Problems in modeling combat stress: A program to meet the challenge. In G.E. Lee (Ed.), Proceedings of the Tenth Annual Symposium on Psychology in the Department of Defense (pp. 422-425A). Colorado Springs, CO: U.S. Air Force Academy.
4. Marlow, D.H., Furikawa, T.P., Griffith, J.E., Ingraham, L.H., Kirkland, F.R., Martin, J.A., Schneider, R.J., & Teitelbaum, J.M. (1 November 1985). New Manning System Field Evaluation (Technical Report No. 1). Washington, DC: Walter Reed Army Institute of Research.
5. Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R., & Jacobs, G.A. (1983). Manual for the State-Trait Anxiety Inventory (Form Y). Palo Alto, CA: Consulting Psychologists Press.
6. Zuckerman, M., & Lubin, B. (1985). Manual for the Multiple Affect Adjective Check List--Revised. San Diego, CA: Educational and Industrial Testing Service.
7. Zuckerman, M. (1979). Sensation seeking: Beyond the optimal level of arousal. Hillsdale, NJ: Erlbaum.
8. Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 80, Whole No. 609.

9. Eysenck, H.J., & Eysenck, S.B.C. (1975). Manual for the Eysenck Personality Questionnaire. San Diego, CA: Educational and Industrial Testing Service.

10. Vitaliano, P.P., Russo, J., Carr, J.E., Maiuro, R.D., & Becker, J. (1985). The Ways of Coping Checklist: Revision and psychometric properties. Multivariate Behavioral Research, 20, 3-26.

11. Vitaliano, P.P., Maiuro, R.D., Russo, J., & Becker, J. (1987). Raw versus relative scores in the assessment of coping strategies. Journal of Behavioral Medicine, 10, 1-18.

12. Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84, 191-215.

13. Kerle, R.H., & Bialek, H.M. (1958). The construction, validation, and application of a Subjective Stress Scale. Staff Memorandum Fighter IV, Study 23. Presidio of Monterey, CA: US Army Leadership, Human Research Unit.

14. McNemar, Q. (1959). Psychological statistics. New York: John Wiley & Sons.

15. Torre, J.P., Jr., Wansack, S., Hudgens, G., King, J.M., Fatkin, L.T., Mazurczak, J., & Myers, J. (in preparation). Effects of competition and mode of fire on physiological responses, psychological reactions, and shooting performance. Aberdeen Proving Ground, MD: Human Engineering Laboratory.

16. Hudgens, G.A., Chatterton, R.T., Jr., Torre, J., Jr., Slager, S.E., Fatkin, L.T., Keith, L.G., Rebar, R.W., DeLeon-Jones, F.A., & King, J.M. (1989). Hormonal and physiological profiles in response to a written examination. In S. Breznitz & O. Zinder (Eds.), Molecular biology of stress: Vol. 97. UCLA Symposia on Molecular and Cellular Biology (pp. 265-275). NY: Alan R. Liss.

17. Bandura, A., Taylor, C.B., Williams, S.L., Mefford, I.N., & Barchas, J.D. (1985). Catecholamine secretion as a function of perceived coping self-efficacy. Journal of Consulting and Clinical Psychology, 53, 406-414.

18. Levine, S., Weinberg, J., & Ursin, H. (1978). Definition of the coping process and statement of the problem. In H. Ursin, E. Baade, & S. Levine (Eds.), Psychobiology of stress: A study of coping men. (pp. 3-21). NY: Academic Press.

19. Wolff, C.T., Friedman, S.B., Hofer, M.A., & Mason, J.W. (1964). Relationship between psychological defenses and mean urinary 17-hydroxycorticosteroid excretion rates. I. A predictive study of parents of fatally ill children. Psychosomatic Medicine, 26, 576-591.

20. Wilkinson, L. (1987). SYSTAT: The system for statistics. Evanston, IL: SYSTAT.

STRESS PERCEPTIONS AMONG THE YELLOWSTONE ARMY FIRE FIGHTERS

MAJ James M. King, Ph.D., Linda T. Fatkin, M.A.,
& Gerald A. Hudgens, Ph.D.

Stress and Performance Team
Behavioral Research Division (SLCHE-BR)
Human Engineering Laboratory
Aberdeen Proving Ground, MD 21005-5001

Abstract

In September 1988, an effort to evaluate the stress experienced by the soldiers who had fought fires at Yellowstone National Park was conducted. It was hoped that this data collection effort would serve as a pilot effort for future evaluations. Our tentative conclusion is that the soldiers fighting the Yellowstone fires experienced a stress level on the high end of the moderate range. The measures of stress used were shown to be sensitive to variations in stress levels in this field situation, and to naturally and experimentally induced stress. The soldiers perceived the situation differently from their leaders. Forty percent of the subjects were unwilling or unable to compare their Yellowstone experience to their experiences in or their ideas of combat. The sixty percent willing to make that comparison noted several factors in common with combat. These findings suggest that it is advisable to maintain a team prepared to collect data from soldiers in situations analogous to the Yellowstone operation.

Introduction

In September 1988, the Concepts Analysis Agency (CAA) requested interagency participation in an effort to evaluate the level of stress experienced by the soldiers who had fought the fires at Yellowstone National Park and to begin to assess its impact on performance. The ultimate goal of this effort is to provide improved human factors data for use in combat models by assessing the degree to which the stress of fighting a fire resembles the stress of fighting a human enemy. The evaluation team, which included two personnel from CAA, two from the Human Engineering Laboratory (HEL), one from the Walter Reed Army Institute of Research (WRAIR), and one from the Army Research Institute (ARI), felt that these fires afforded an outstanding opportunity to study stress reactions and to collect human factors data in a real operational setting which shares with combat the elements of personal danger and uncertainty. It was hoped that this data collection effort would serve as a pilot effort for future evaluations in which performance will be more firmly tied to the level of stress experienced by the soldiers involved.

The effects of stress on human performance have received extensive attention (6). Stress-induced performance decrements have been demonstrated in soldiers performing combat-relevant tasks (5). The present HEL Stress Research Program (3), a combination of in-house and contract efforts, is presently studying the links between psychological and physiological stress reactions and performance in a variety of settings.

This program is well along in its effort to develop a psychological and physiological metric of stress which can be used to compare stress levels across situations. To date, studies conducted under this program have included a variety of situations, including viewing movies depicting stressful material, waiting while one's spouse has outpatient surgery, waiting while one's spouse has major surgery, taking an important medical school oral examination, taking a major written examination in medical school (4), and firing in an interunit competitive marksmanship situation while being observed by one's fellow soldiers (6). We were, of course, very interested in the opportunity to study a new, and potentially highly stressful situation.

The HEL Stress Program assembled a two-person team to participate directly in this evaluation of the Yellowstone fire fighting experience. They developed a questionnaire which was given to the soldiers, along with two standard psychological measures, to evaluate their stress levels during the fire fighting.

Method

Subjects: The subjects were 1100 soldiers, noncommissioned officers, warrant officers, and officers of the 9th Infantry Division Motorized and supporting units who participated in or supported the 1988 fire fighting operation at Yellowstone National Park. All subjects participated in the study voluntarily.

Survey: The survey used for this study contained adjective checklists (the Multiple Affect Adjective Check List-Revised or MAACL-R, 7) to describe how they felt when filling out the survey and when actually fighting the fire, a rating of the stressfulness of the fire fighting (the Rating of Events Scale), ratings of the risk of their duties, a subjective performance assessment, and items addressing coping with the situation. Although the survey was customized for this application, nearly all of the scales had been used in other studies which make up the HEL Stress Program. This was done in order to permit us to use the psychological portions of our stress metric which is in the early stages of development to assess, conditionally, the stressfulness of the Yellowstone fire fighting experience.

Procedures: The surveys were administered to soldiers in groups which nominally consisted of either company or battalion groupings. Soldiers were surveyed at Yellowstone National Park, Bozeman, Montana, and Fort Lewis, Washington after they had completed their fire fighting duties. The soldiers were provided with the survey, and a pencil, and were briefed on the purpose and content of the instrument. They were instructed to read the Volunteer Agreement Affidavit, and, if they agreed to participate in the study, proceed to fill out the rest of the questionnaire. Great care was taken to emphasize the voluntary nature of their participation in the study. Members of the evaluation team solicited comments from individuals who wished to elaborate on their responses or to address issues not covered in the surveys. The soldiers and their leaders were extremely cooperative throughout the data collection process.

Results and Discussion

Analysis of the data collected is ongoing, and a detailed technical report is in preparation. The present paper will present data on 1100 soldiers in the context of data obtained in the other HEL Stress Program studies. Thus, we will tie the level of stress experienced by these soldiers to that experienced by the subjects in the other studies. Data are also presented by Task Force and Rank. The data presented in the figures which follow are displayed in a mean (the bar) plus one standard error (the capped vertical bar) format. On each of the figures shown, the overall F ratio for group differences is significant at $p < 0.004$ or better.

The Ratings of Events for the situations studied to date are depicted in Figure 1. The groups are spouses of patients undergoing abdominal surgery, medical students taking a major written examination, combined surgery and examination control groups, soldiers firing in a noncompetitive marksmanship setting, soldiers firing in a highly competitive marksmanship setting, and the Yellowstone fire fighters. They had been asked to rate the stressfulness of their experiences on a scale of 0 to 100. Our tentative conclusion is that the soldiers fighting the Yellowstone fires experienced a stress level on the high end of the moderate range, comparable to that experienced by spouses of patients undergoing major abdominal surgery, for instance. This abdominal surgery group contains some of the most stressed individuals studied in our other efforts.

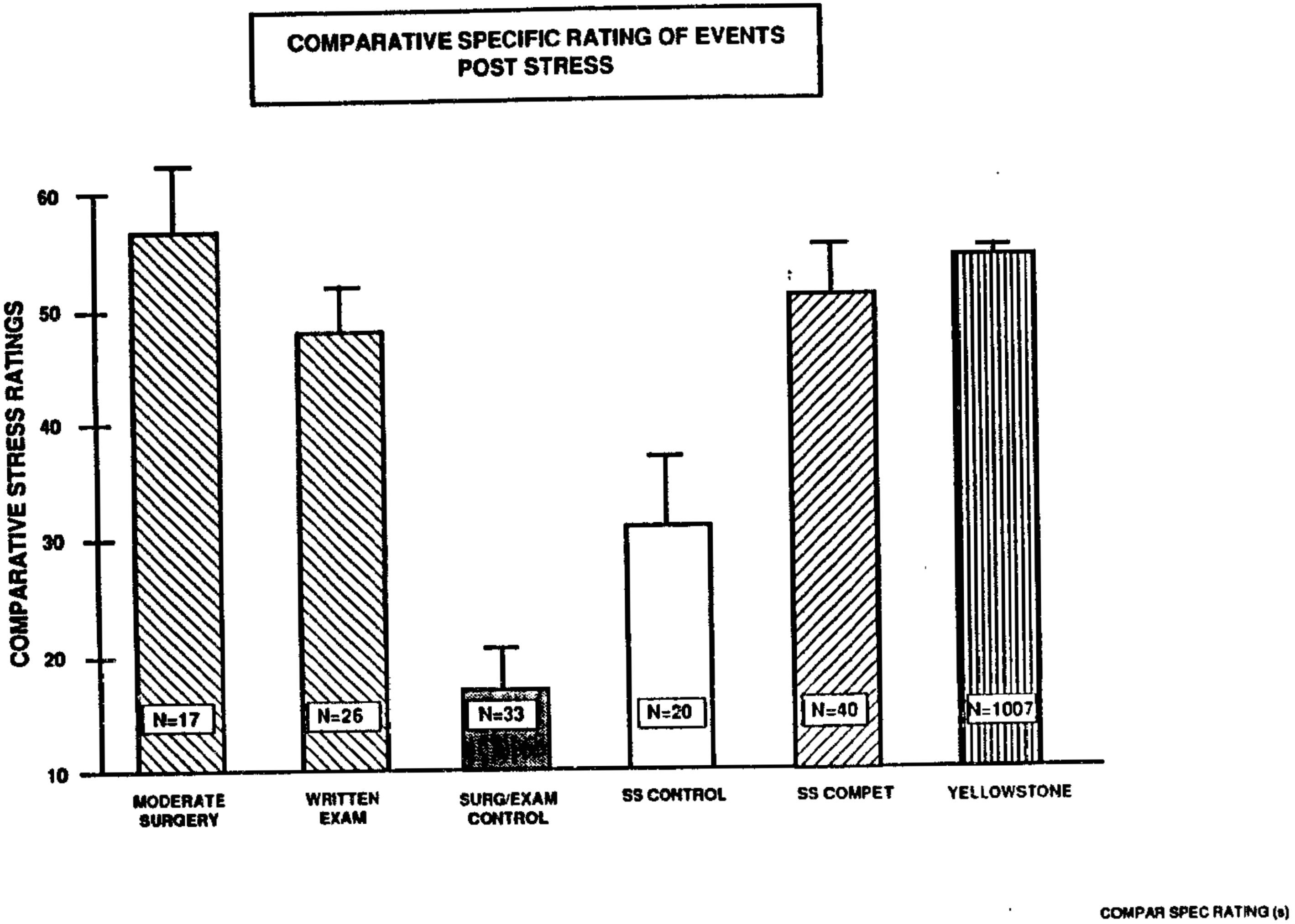
Figure 2 displays the event ratings broken out by the task forces (TF) into which the Yellowstone fire fighters were organized (TFs A through E), or into which they could be logically placed (TF F, the aviation personnel). Except for TF F, the demographics of the TFs were similar. Note that TF C had by far the highest stress event rating we have recorded to date. This TF was extensively involved in structure protection during their period in Yellowstone.

The MAACL-R scores for Anxiety (Figure 3), Depression (Figure 4), Hostility (Figure 5), Positive Affect (Figure 6), and Sensation Seeking (Figure 7), broken out by TF reveal that TFs C and F, although generally at opposite ends of the response spectra, are clearly distinguishable from the other TFs on the measures. This was also true for ratings of the Life Threatening aspects of these duties (Figure 8), and for Coping Efficacy (Figure 9), but is less clear for our Success, the subjective measure of performance effectiveness (Figure 10). This measure was collected by asking the soldiers how successful, on a scale of 0 to 100, they felt about getting the job done.

Thus, the measures of stress have been shown to be sensitive to variations in stress levels in this field situation, and to naturally and experimentally induced stress (1). The present results are less useful than might otherwise be the case due to lack of true baseline and objective performance measures and to the time delay in obtaining access to the soldiers, shortcomings which we plan to correct in subsequent efforts.

A stepwise regression analysis using the subjective measure of performance, Success (SUCC, see Figure 10), as the dependent measure was

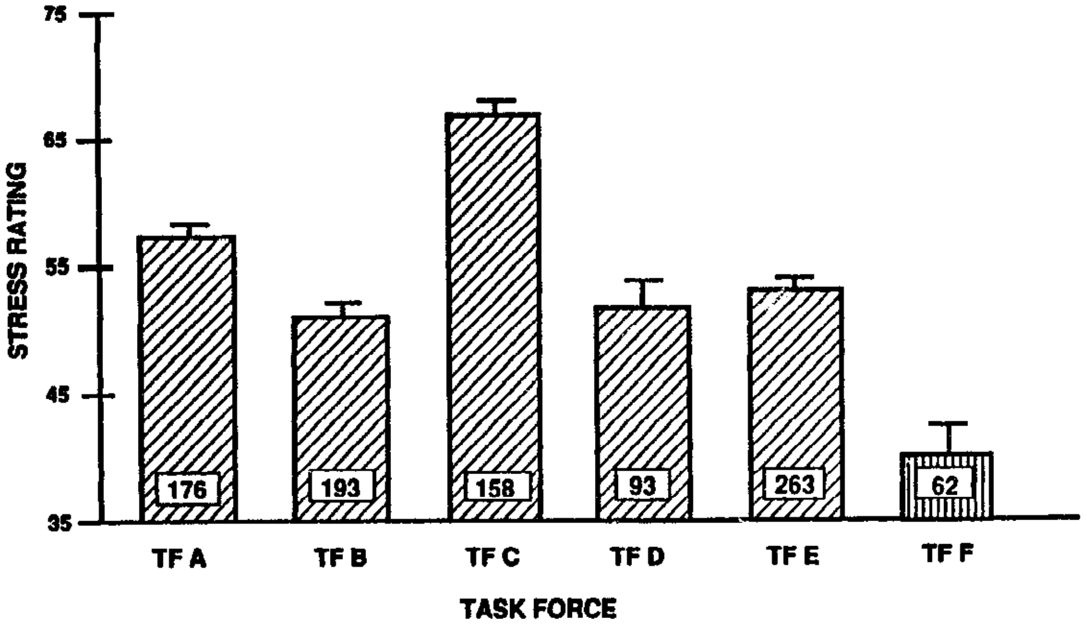
Figure 1



COMPARATIVE SPECIFIC RATING OF EVENTS POST STRESS

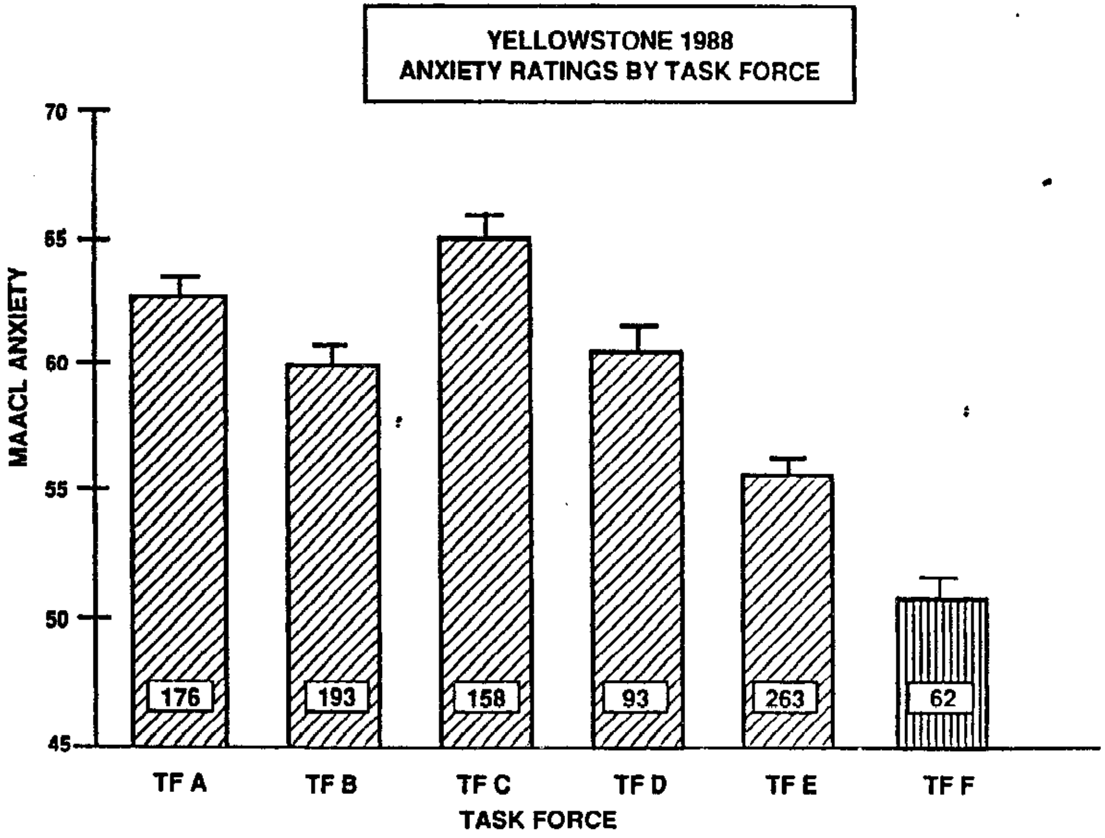
Figure 2

YELLOWSTONE 1988
SPECIFIC RATING OF EVENTS BY TASK FORCE



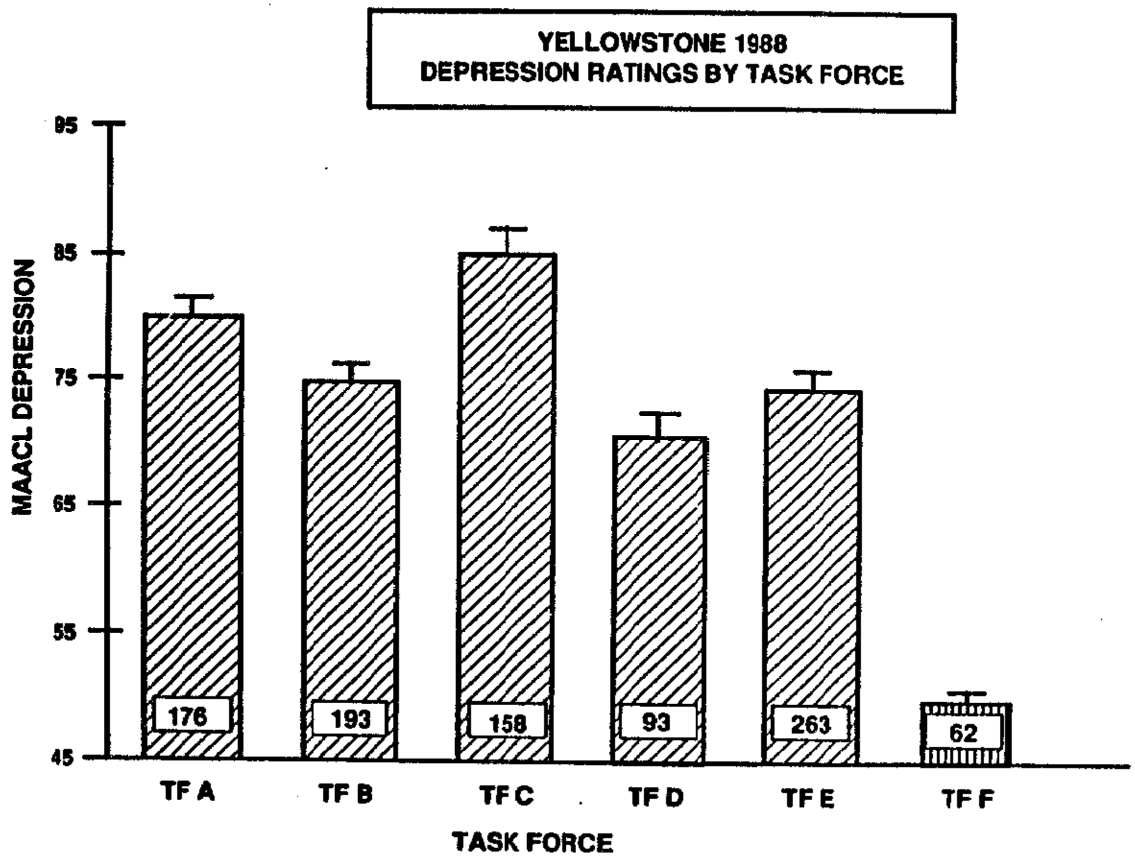
YS SPECIFIC RATING (s)

Figure 3



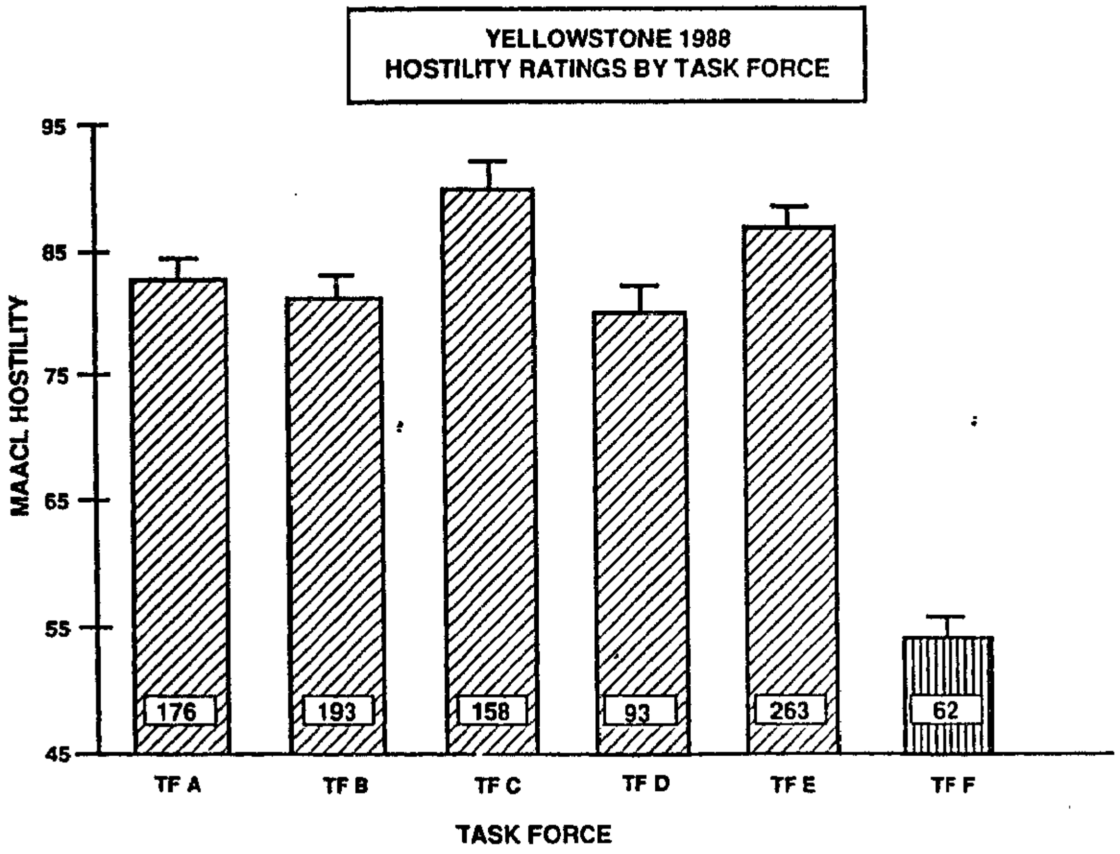
YS ANXIETY TF (s)

Figure 4



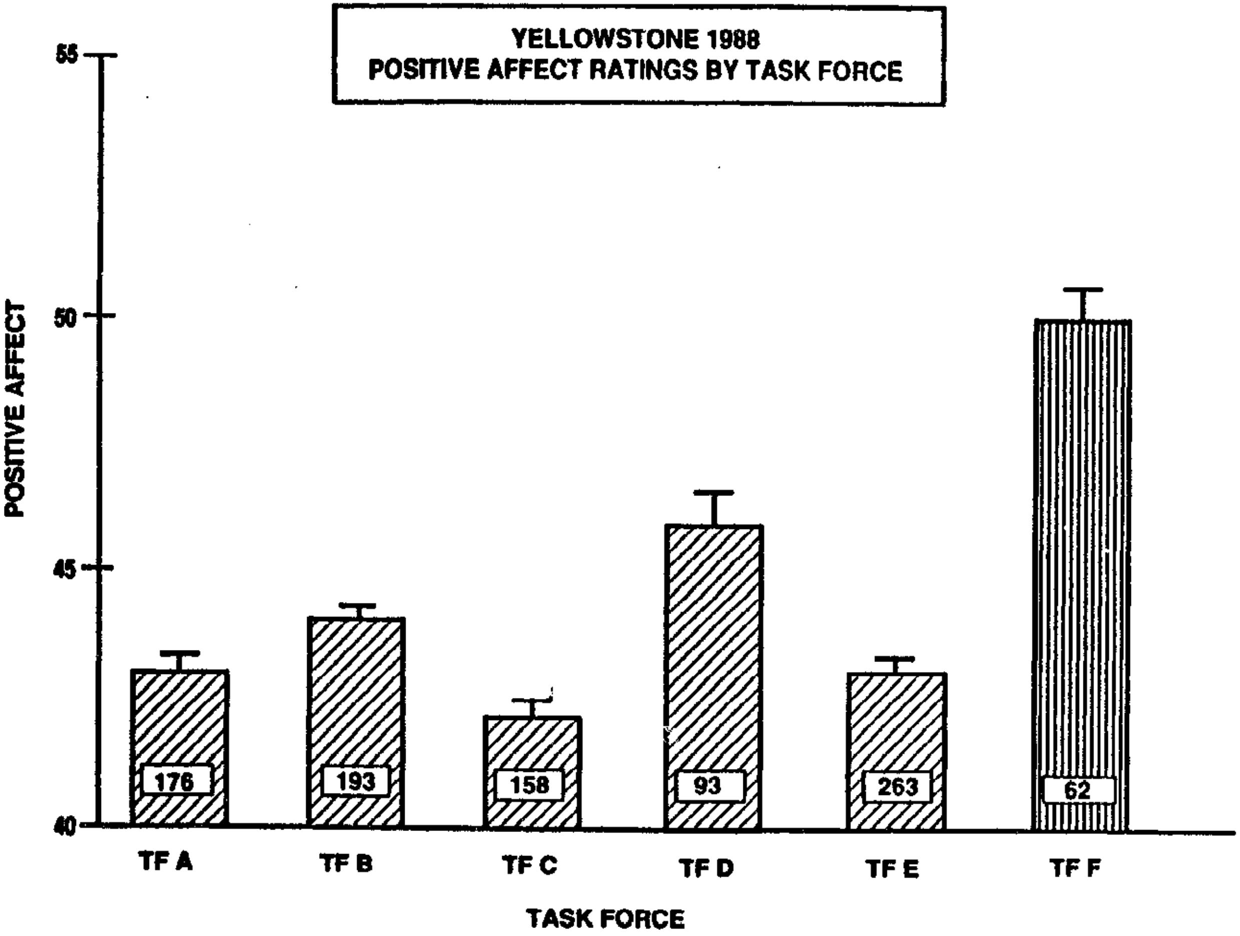
YS DEPRESS TF (s)

Figure 5



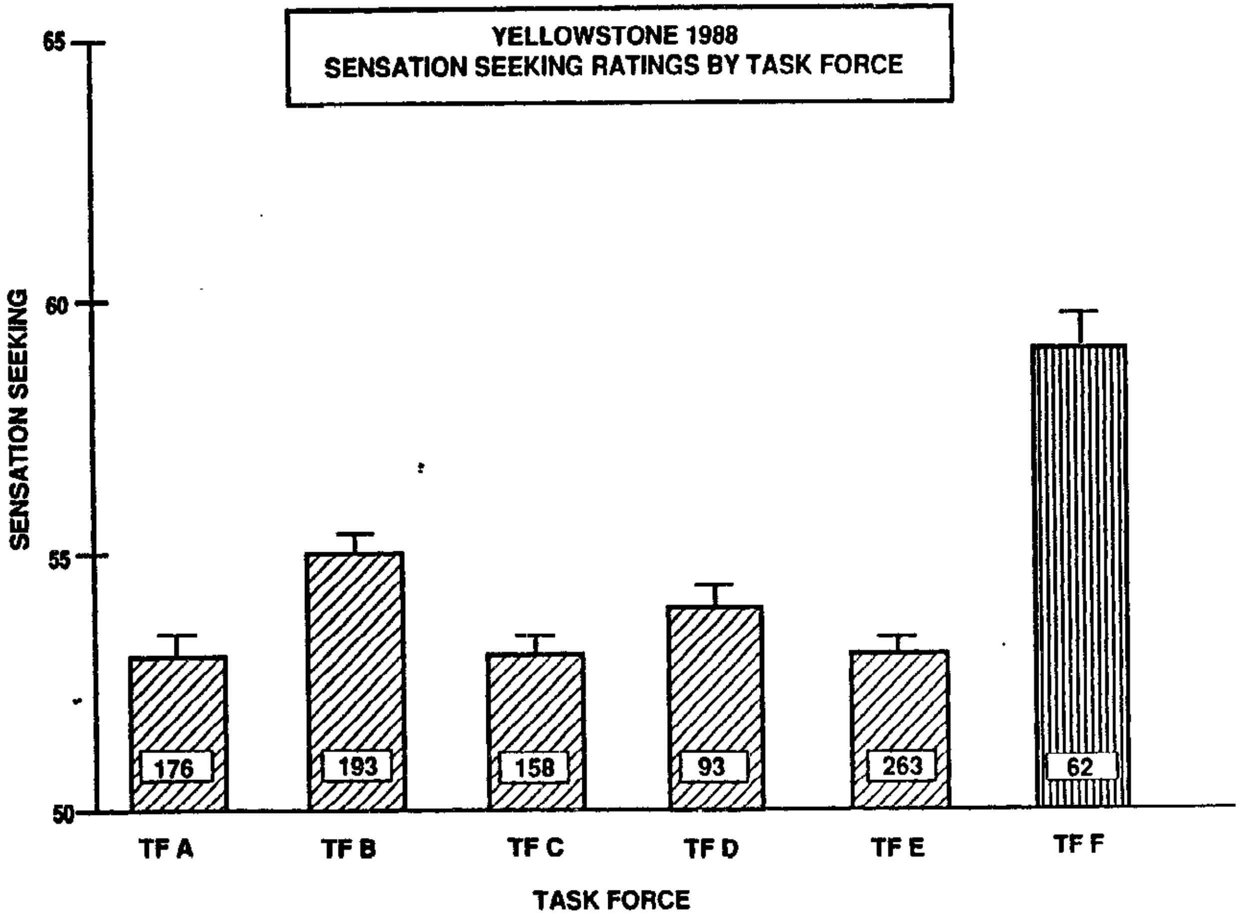
YS HOSTILITY TF (s)

Figure 6



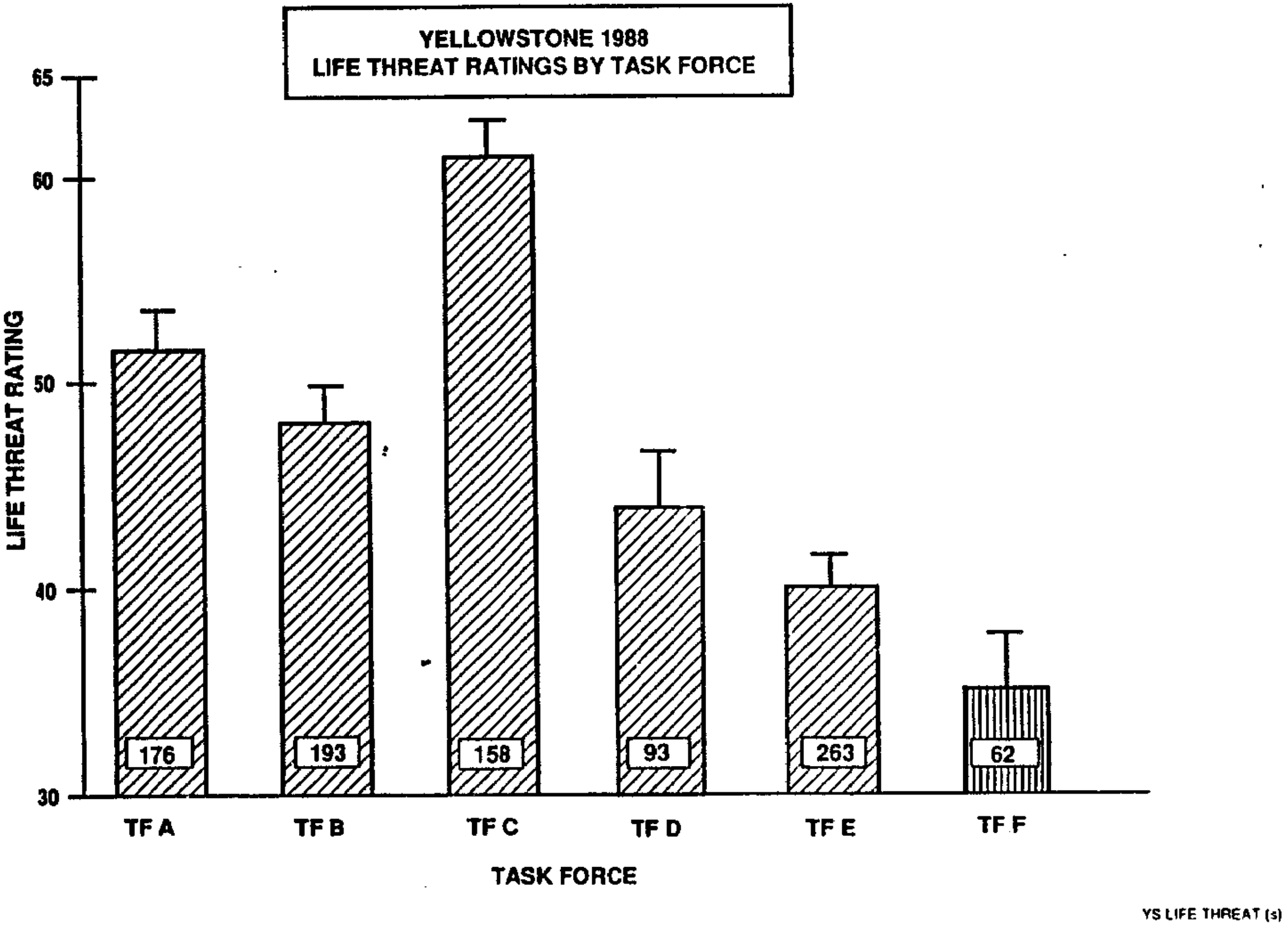
YS POS AFFECT (s)

Figure 7



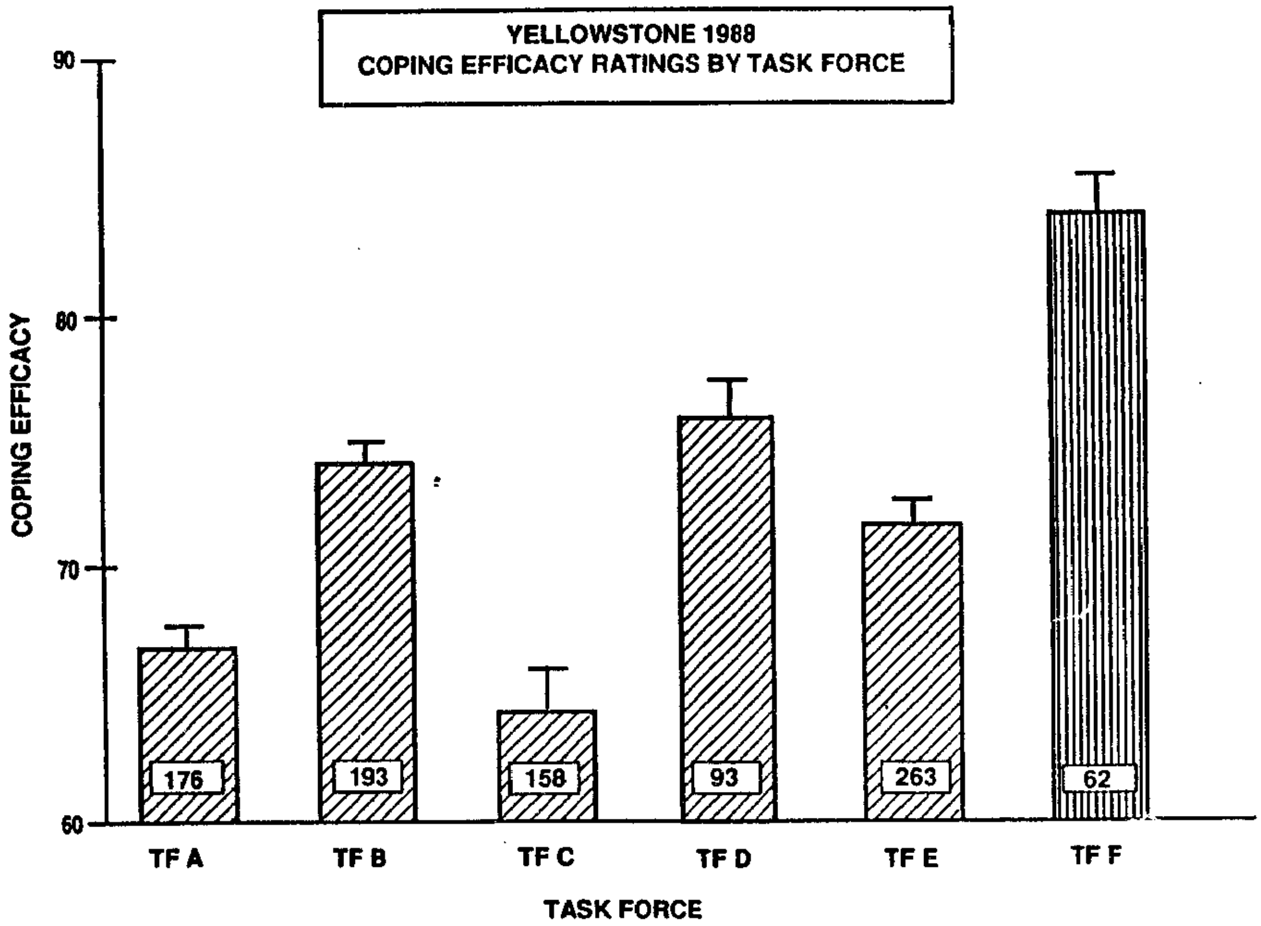
YS SENSATION SEEK (s)

Figure 8



YELLOWSTONE 1988
LIFE THREAT RATINGS BY TASK FORCE

Figure 9

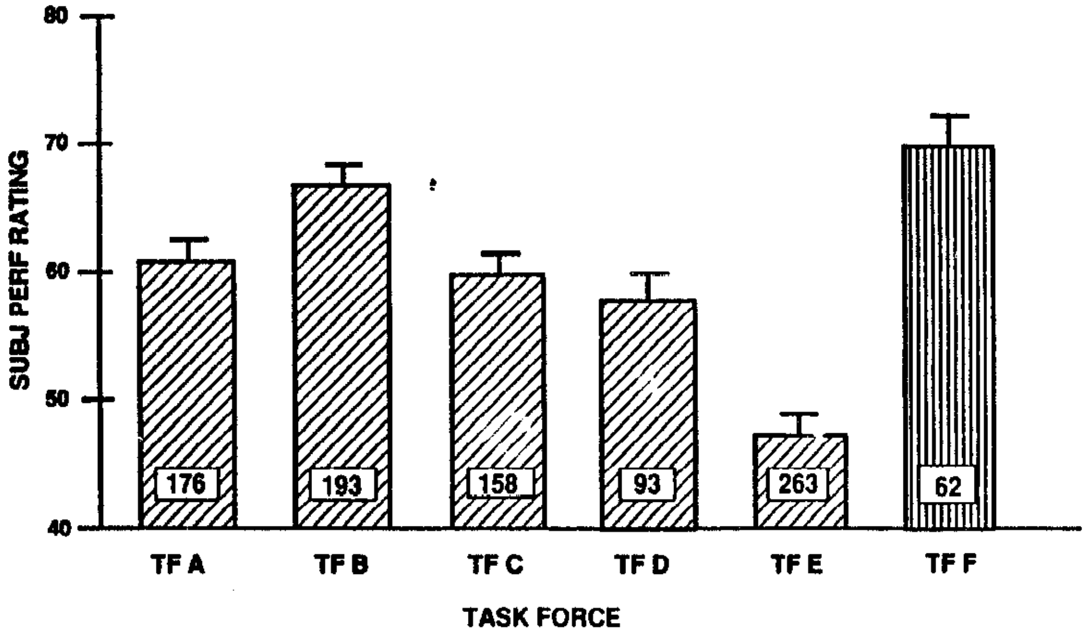


YELLOWSTONE 1988
COPING EFFICACY RATINGS BY TASK FORCE

YS COPING EFFICACY (s)

Figure 10

YELLOWSTONE 1988
SUBJECTIVE PERFORMANCE RATINGS BY TASK FORCE
("HOW SUCCESSFUL DID YOU FEEL. . . ?")



conducted for subjects in all TFs. The final model, which was based on 860 cases, involved as variables Length of Service (LOS), Life Threatening (LT, see Figure 8), Coping Efficacy (COPE, see Figure 9), Anxiety (ANX, see Figure 3), Hostility (HOS, see Figure 5), Positive Affect (PA, see Figure 6), and the Rating of Events (SRE, see Figures 1 and 2). SUCC and the predictor variables LT, COPE, and SRE can range in value from 0 to 100. For this sample, a reasonable range for LOS is 0 to 20 years. For this population, ANX can range from 37 to 168, HOS can range from 39 to 237, and PA can range from 23 to 88. The multiple R was 0.421, which accounted for 17.7% of the variance in the subjective performance rating. The model itself, which was highly significant ($p < 0.001$), is given below.

$$\mathrm{SUCC} = -1.963 + 0.401*\mathrm{LOS} + 0.133*\mathrm{LT} + 0.331*\mathrm{COPE} + 0.162*\mathrm{ANX} - 0.176*\mathrm{HOS} + 0.625*\mathrm{PA} + 0.119*\mathrm{SRE}$$

Another stepwise regression analysis was conducted using only the soldiers in TF C, which appeared to have been the most stressed of the TFs. The final model, based on 127 cases, included as variables COPE and HOS. The multiple R was 0.466, which accounted for 21.7% of the variance in the subjective performance rating. The model, which was highly significant, follows.

$$\mathrm{SUCC} = 41.408 + 0.477*\mathrm{COPE} - 0.136*\mathrm{HOS}$$

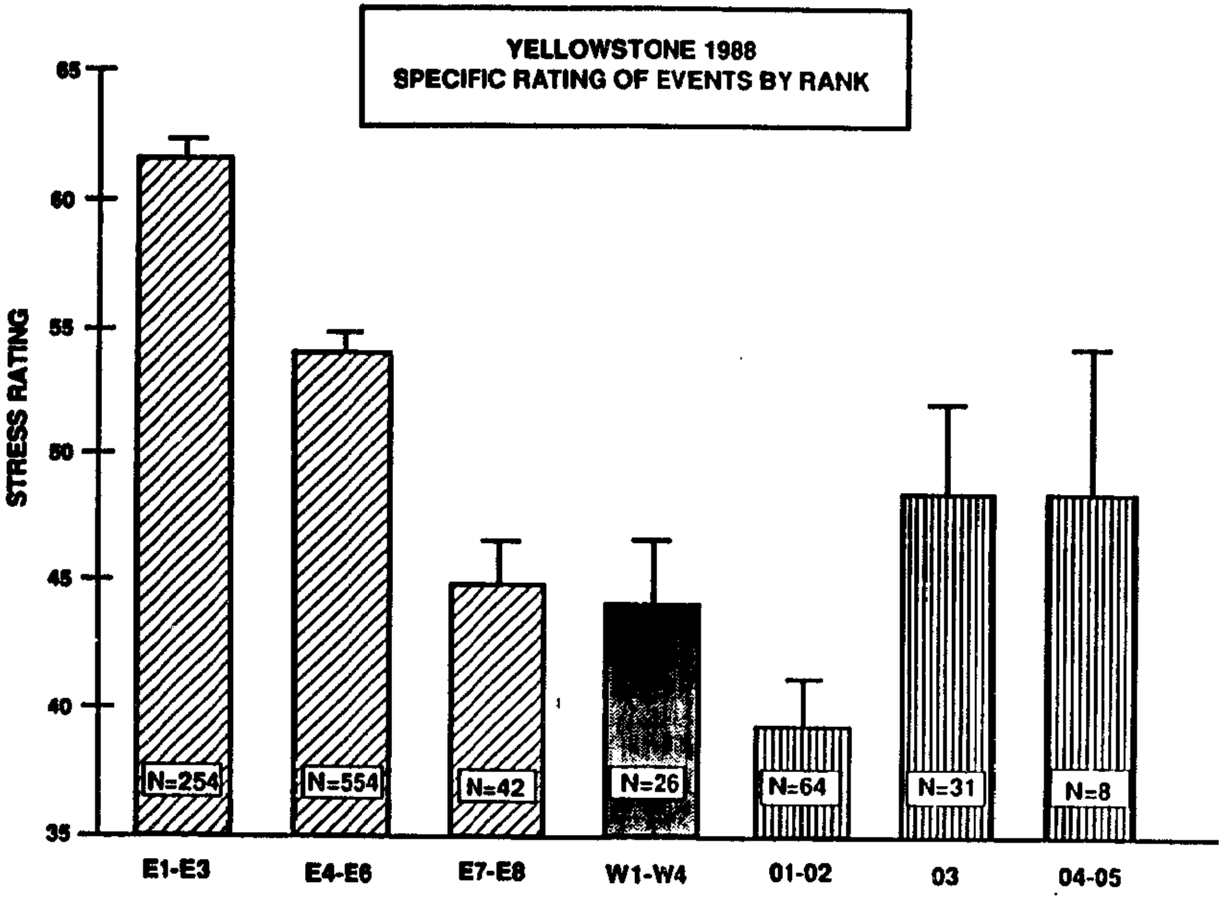
It is instructive to note that COPE and HOS contribute substantially to both of the models. This is consistent with the results we have obtained in a variety of other situations (1).

It is also worth noting that the soldiers perceived the situation differently from their leaders. This point is made rather strongly by Figure 11, which presents the Rating of Events by Rank, and by Figure 12, which displays our subjective performance rating, Success, by Rank.

The comments made by the soldiers bear on the issue of using operations such as the Yellowstone National Park fire fighting experience as a model of combat. Roughly 40% of the subjects were either unwilling or unable to compare their Yellowstone experience to either their experiences in or their ideas of combat. Those willing to make such a comparison noted that the Yellowstone operation shared several common factors with combat. These included the deployment process, family separation, the need for leadership, teamwork, and discipline at the unit level, and the requirement to manage individual differences in stress responses. Other common factors included the sustained nature of the work, with alternating periods of intense activity and boredom, unfamiliar terrain with limited ingress and egress routes and dangerous animals, the physical strain of fire fighting and the long (10 to 14 mile) marches to fire-fighting sites, complications arising from communications, and the unpredictable nature of the fire itself.

Based on these results, we will prepare revised surveys, and will strive to include baseline measures, measures taken closer in time to the event, and objective measures of performance. These improved surveys will be used when opportunities arise in the future to study soldiers in other potentially stressful operations. Our experiences to date suggest that

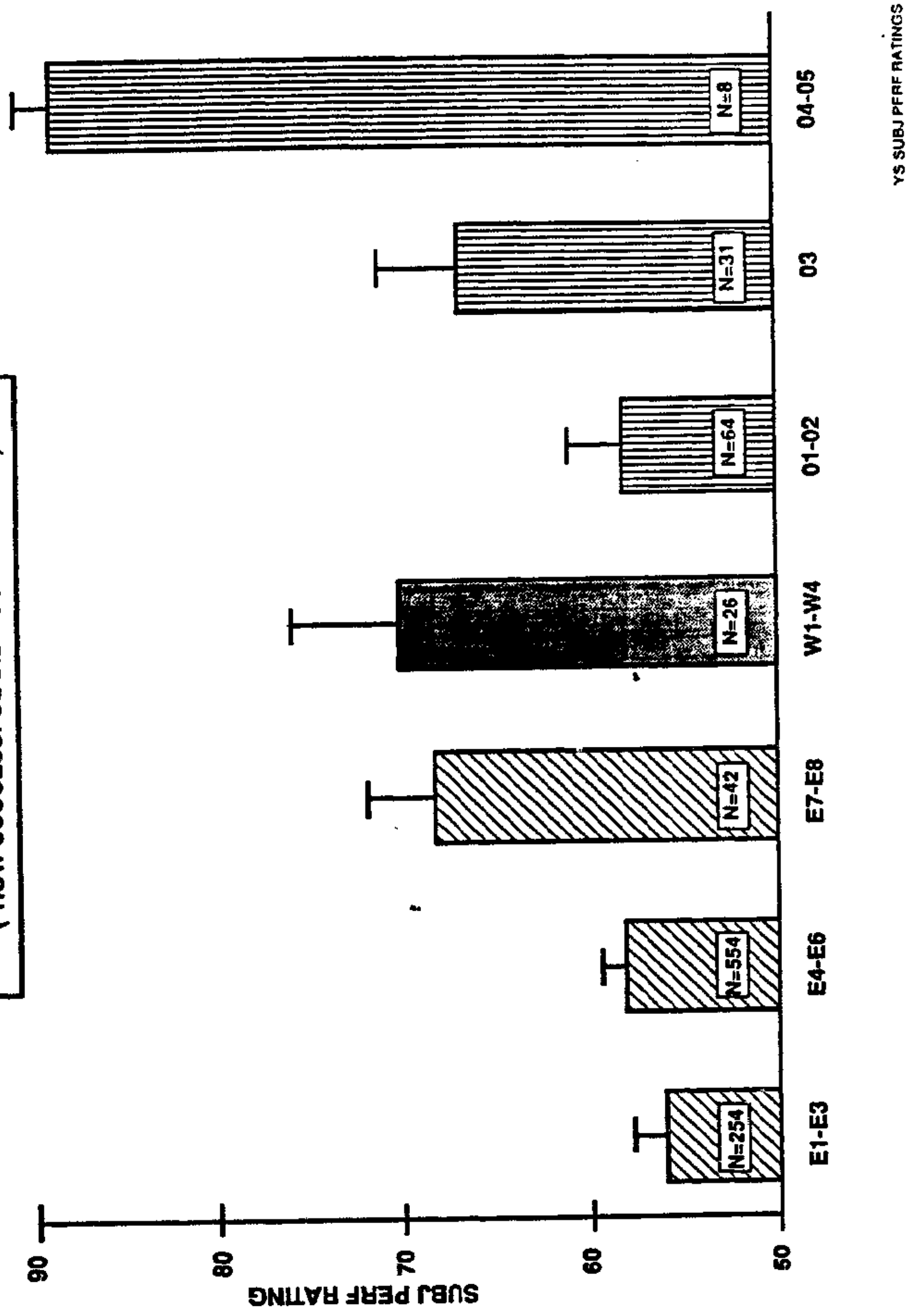
Figure 11



YS SPEC RATE EVENTS (s)

Figure 12

YELLOWSTONE 1988
SUBJECTIVE PERFORMANCE RATINGS BY RANK
("HOW SUCCESSFUL DID YOU FEEL . . . ?")



much valuable information relevant to the behavior and performance of soldiers and their leaders in combat can be collected in situations such as that offered by the Yellowstone National Park fires, because, unlike training, these situations involve real hazards, real dangers, and real consequences in a real world setting. The fire, unlike a human enemy, is neither alive nor is it motivated to defeat the soldiers, but it is, none the less, a dangerous and unpredictable foe. These findings further suggest that it is advisable to maintain a team prepared to collect data from soldiers in situations analogous to the Yellowstone operation. With the addition of more performance data to the collection effort, such undertakings will be able to provide a steady flow of information on human performance in operational settings to the modeling community.

References

1. Fatkin, L.T., Hudgens, G.A., King, J.M., Torre, J.P., Jr., Wansack, S., Mazurczak, J., Myers, J., Slager, S.E., and Chatterton, R.T., Jr. (1989). The use of competitive marksmanship as a stressor in soldier/equipment testing. In Human Behavior and Performance as Essential Ingredients in Realistic Modeling of Combat - MORIMOC II. Alexandria, VA: Military Operations Research Society.

2. Hockey, G.R.J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.). Handbook of Perception and Human Performance. Volume II Cognitive Processes and Performance. New York, NY: John Wiley and Sons, 44-1 to 44-49.

3. Hudgens, G.A., Torre, J.P., Jr., Chatterton, R.T., Jr., Wansack, S., Fatkin, L.T., and DeLeon-Jones, F.A. (1986). Problems in modeling combat stress: A program to meet the challenge. In G.E. Lee (Ed.). Proceedings, Psychology in the Department of Defense, Tenth Annual Symposium. Colorado Springs, CO: U.S. Air Force Academy.

4. Hudgens, G.A., Chatterton, R.T., Jr., Torre, J., Jr., Slager, S.E., Fatkin, L.T., Keith, L.G., Rebar, R.W., DeLeon-Jones, F.A., and King, J.M. (1989). Hormonal and physiological profiles in response to a written examination. In O. Zinder and S. Bresnitz (Eds.). Molecular Biology of Stress. UCLA Symposia on Molecular and Cellular Biology, New Series, Volume 97. New York, NY: Alan R. Liss, Inc., 265-275.

5. Torre, J.P., Jr. (1966). The Effects of Stress on the Performance of Riflemen. Aberdeen Proving Ground, MD: Human Engineering Laboratory.

6. Torre, J.P., Jr., Wansack, S., Hudgens, G., King, J.M., Fatkin, L.T., Mazurczak, J., and Myers, J. (in preparation). Effects of Competition and Mode of Fire on Physiological Responses, Psychological Reactions, and Shooting Performance. Aberdeen Proving Ground, MD: Human Engineering Laboratory.

7. Zuckerman, M. and Lubin, B. (1985). Manual for the MAACL-R: Multiple Affect Adjective Check List - Revised. San Diego, CA: Educational and Industrial Testing Service.

A STRATEGY FOR MODELING COMBAT RELATED PERFORMANCE DECREMENTS: DOSE EQUIVALENCY

Robert S. Kennedy*, Dennis R. Baltzley*,
Norman E. Lane*, Marshall B. Jones**

*Essex Corporation
1040 Woodcock Road, Suite 227
Orlando, Florida 32803

**Milton S. Hershey Medical Center
Pennsylvania State University
Hershey, Pennsylvania 17033

ABSTRACT

The most straightforward way to measure operational performance is to monitor it on the job. However, when such measures are employed, they are generally insensitive to all but the most powerful stressor effects. We believe the chief impediment to the development of suitable battlefield performance models is the lack of reliability of the measures. Insufficient attention to reliability can lead to reduction of statistical power, higher sample size requirements, increased cost of experiments, and when hazard and discomfort are involved, other problems. We propose two linking approaches to address these difficulties: Surrogate Measures and Dose Equivalence. Surrogate measures are those which are related to, or predictive of, a construct of interest, but are not direct measures. Dose Equivalence is a technique which scales performance deficits and indexes them to graded levels of agents which adversely affect humans. Standardized treatments are empirically studied and transfer functions developed against "dosages." Battlefield performance decrements related to stress, for example, which are less amenable to experimental study, are related by equivalency of effect.

This paper describes these strategies for modeling combat-related performance decrements and our progress with the programmatic development of a microcomputer-based performance test battery. Because the tests of this battery have all of the requisite metric features, and are to some extent standardized, performance loss can be equated between agents. From these relations one can infer that, for example, 18,000 feet is like three ounces of alcohol (in terms of performance). Applications of this strategy to the adverse conditions of combat and battle stress are discussed.

INTRODUCTION

We hypothesize that the chief problem in the assessment of operational performance is the lack of reliability of the measures. It is our opinion that not enough attention has been paid to the reliability of criteria or dependent variables in experimental studies. The consequence of this omission can be seen in the well-known correction-for-attenuation formula reported by Spearman in 1904 (1). The equation is expressed as:

$$r_t = \frac{r_{xy}}{\sqrt{(r_{xx})(r_{yy})}} \tag{1}$$

where R_{xy} is the predictive (obtained) validity, R_{xx} is the reliability of the predictor, R_{yy} is the reliability of the criterion, and R_t is the true relationship. With low reliability in either variable x or variable y, the greatest possible relationship between the two is substantially limited. For measuring operational performance, this relationship is considerable.

Correction for attenuation is a commonly employed relation in classical test theory (2). Although it is found in all statistics texts, the equation is not commonly employed in theory development nor practice in human performance or operations research. We believe that this equation and its variant (e.g., Equation 2), properly applied, can have important implication for combat modeling.

$$r_{xy} \leq (r_{xx}r_{yy})^{1/2} \tag{2}$$

Insufficient attention to reliability can lead to reduction of statistical power, higher sample size requirements, cost of experiments, and when hazard or discomfort are involved, human use problems. Utilizing the correction-for-attenuation formula often changes conclusions. For instance, the true predictability of operational criteria from paper-and-pencil aptitude tests are often misinterpreted because of criterion unreliability. Relatedly, treatment effects (e.g., radiation sickness fatigue) have markedly different reliabilities so that their influences cannot be simply combined.

In our experience with operationally-based performance measurement as the criteria, reliability of the criteria are rarely higher than $\underline{r}$ = 0.50 and usually much lower. This factor limits predictive relationships because when low correlations are obtained, they are interpreted incorrectly as lack of relationship. When attempts are made to use these low-obtained relationships to create combat models, the outcomes have little predictive power. Application of the correction-for-attenuation formula allows one to sidestep this error and reinterpret the relationship. To address this issue we have prepared a look-up table (Table 1) to illustrate the difficulties. We have assumed that operational predictors and criteria have limited reliability and have shaded out the correlation possibilities higher than 0.50. Likewise, predictor reliabilities greater than 0.70 are also shaded. The correlation coefficients which remain are examples of the maximum relationships possible under the cosntraints given. It is clear that if a surrogate or proxy work measure (3) is substituted (and these relationships are <u>not</u> considered) one might conclude wrongly that the selected measure may not be appropriate. Moreover, as the next table shows (Table 2), unless sample sizes exceed 40 subjects, most experiments of this type lack statistical power to discover anything.

The relationship in the denominator of Equation 1 suggested to us a focus on developing highly reliable measure sets such as the Automated Performance Test System (APTS) (4) separate from the operational criteria,

Table 1. Maximum Predictive Validities Possible Based On Combinations of Retest Reliabilities for Predictor and Criteria Less Than r = 1.00

PREDICTOR RETEST RELIABILITY	CRITERION RETEST RELIABILITY					
	.99	.77	.55	.33	.11	.00
.99	.99	.87	.74	.57	.33	.00
.77	.87	.77	.65	.50	.29	.00
.55	.74	.65	.55	.43	.25	.00
.33	.57	.50	.43	.33	.19	.00
.11	.33	.29	.25	.19	.11	.00
.00	.00	.00	.00	.00	.00	.00

Table 2. Comparison of Magnitude of Predictive Validity Correlations For Two Significant Levels As A Function of Sample Size When One r Calculated

Sample size	$p < .05$	$p < .01$
10	.576	.708
15	.482	.606
20	.423	.537
25	.381	.487
30	.349	.449
40	.304	.383
50	.273	.354
100	.195	.254
500	.088	.115
1000	.062	.081

but highly similar to the criteria in skill requirements. We call this method Surrogate and we stipulate five formal requirements for this approach (listed below). If the measures correlate well with the criteria, and behave similarly under changing task conditions, perhaps they could be used in place of the criteria; for example, as a surrogate in the case of assessing fitness for duty, a highly reliable measurement set of basic psychomotor and cognitive functions could be used to assess the operational criteria because the sets would be tailored to skill requirements.

The Surrogate Measurement Approach

Surrogate measures are those which are related to or predictive of a construct of interest but are not direct measures. In our plan, surrogate measures are composed of tests in batteries which exhibit five characteristics:

1. test performances are stable so that "what is being measured" is constant;
2. scores are correlated with the performance construct;
3. they are sensitive to the same factors that would affect performance as the operational performance variable would;
4. they are more reliable than field measures; and
5. they require minimal training time.

Surrogate measures differ from conventional performance measures in that tests need not involve operations in common with the performance measures, only components or factors in common. They also differ from "synthetic" or "job samples" because the surrogate takes little practice and is simply scored. Given the great difficulty of obtaining reliable enough field measures to carry out stressor sensitivity studies on an operational task, the case for using a surrogate is strong. A large portion of variance and extremely complex tasks can be predicted from performance on relatively simple tests. However an external test or battery cannot be as "valid" as the measure itself from a practical standpoint, but may have more of the true variance of field performance because its reliability is much greater.

Many of the most critical stressors produce important but relatively small effects on overall task performance. Combining the small effects with the well-known statistical unreliability of performance measures in battlefield exercises makes it difficult to detect effects unless experiments are based on very large samples of field performances. Second, when performances on indicator tasks fail to stabilize with practice, or require extended practice to become sufficiently stable for analysis, they also suffer from this particular problem. Stability on the task is rarely achieved on the complex performances of operational tasks, even among seasoned military operators. Use of these data for the generation of combat models, coupled with the diversity of military occupational specialties -- which may differentially interact with battlefield stressors -- is rife with problems. Our proposed solution proceeds from our work with a microcomputer performance battery which has been shown to be sensitive to the adverse environmental and toxic conditions encountered in military life.

Dose Equivalency: As an operational strategy for implementation of surrogate measures we advocate the use of degraded performance indexed against specific toxic or environmental stressors to provide a dose response relationship. Under NASA and NSF support, we have representative data for alcohol and partial pressure of oxygen (altitude) on a standardized portable microcomputer test battery. Following the surrogate logic, the tests in the battery have been indexed to measures which themselves are related to military jobs (ASVAB, ACT, WAIS, Wonderlic), and such measures may be used as surrogates for operational performances. Other agents like fatigue, sleep loss, and combat stress can be indexed to these agents for purposes of generating human exposure standards. Illustrations of this "dose equivalency" approach are provided where, among others, halon, drugs (e.g. scopolamine), and sleep loss are "marked" against alcohol and hypoxia.

Both *surrogate* and *dose* equivalency strategies require tests, and we employ the APTS (4). APTS development has followed the assumptions of classical test theory (2,5) and the empirical findings of the Performance Evaluation Test for Environmental Research (PETER) program (6). Begun by the Navy, the chief outcomes of the PETER program were statistical methodologies with which to evaluate tests for repeated-measures applications and a series of "good" or "qualified" mental tests. Over the last decade, and through the support of organizations such as NASA, National Science Foundation, and corporations in the private sector, a program of study for microcomputer implementation, mechanization, and psychometric development has been carried out. The APTS psychometric development has included culling of each test's properties to produce a menu of acceptable tests of cognition, information processing, psychomotor skill, memory, mood, and others. From this menu of tests a battery may be tailored to suit specific applications.

Sensitivity Studies. The APTS has undergone empirical validation in more than a dozen sensitivity studies. Six of these have sufficient commonality that they invite meta synthesis and will be discussed. Four of the studies use each subject (in a "placebo" condition) as his or her own control, and the two others use either matched cohorts (chemoradiotherapy) or control groups (drugs). The APTS has also been used in studies involving sleep loss, cave dwelling, and exercise interventions for fatigued airline pilots, but these results are not yet analyzed.

From the sensitivity studies we have attempted a preliminary model using a series of dose equivalent matrices (Table 3). Although available on a limited number of individuals and conditions, they are illustrative of what in our judgment should be a long-range plan for the purpose of obtaining human performance data which can be employed in the creation of combat or battlefield performance models. On the one hand, we have nine "good" tests, most of which have been used on all environments and treatments. Also, we will compare: (1) three blood alcohol levels, .05, .10, .15; (2) simulated altitude at 15-20K at 23-25K; (3) motion sickness drugs, scopolamine and a combination of scopolamine and dexedrine; (4) effects of chemoradiotherapy, reported as an average decrement in treated versus untreated cohorts (donors); (5) two experimental drugs; and (6) halon gas decrement (averaged across 24-hour exposures). We suggest that this relation be used as a

Table 3. Approximate Percent Decrement Across
Six Sensitivity Studies Involving the APTS

Percent				Types of Intervention Levels of Treatment and Percent Functional Loss from a Placebo							
APTS Based Tests	Alcohol			Altitude		MS Drugs		Chemo	Drugs*		Halon*
	.05*	.10	.15	15-20K	23-25K	Scop	Comb	Avg.	X	Y	Avg
Short Battery < 15 Min.											
NONPREFERRED HAND TAP				0	4	3	0	10	--	--	--
4-CHOICE REACTION TIME	5	10	23	--	--	--	--	12	11	10	8
CODE SUBSTITUTION	7	9	25	--	--	--	--	37	3	2	4
GRAMMATICAL REASONING	9	9	20	4	22	9	7	39	17	7	4
PATTERN COMP. (SUC)	0	4	10	1	12	2	3	17	2	3	6
MANIKIN	6	9	14			2	0	15	1	0	2
TWO-HAND TAP	--	--	--	0	4	--	--	4	--	--	--
Medium Battery < 20 Min.											
MATH PROCESSING	6	4	10	--	--	--	--	--	--	--	4
PATTERN COMPARISON	--	--	--	--	--	--	--	--	--	--	--

Legend:
* Data tentative, under analysis.
Drugs = Antihistimine
X = Drug X
Y = Drug Y
Halon = Fire extinguishing agent
MS = Motion Sickness Drugs
Chemo = Chemoradiotherapy
Scop = Transdermal Scopolamine
Avg = Average decrement across the evaluation

preliminary marker to index other comparable effects, calculated as percentage of baseline. This approach is advocated for providing guidance regarding strength of relationships and "dose equivalency," not for statistical testing. It is well known that percentages lack sufficient statistical power and are generally to be avoided for analytic purposes. We use percent decrement here to provide a basis for comparison of effects across different treatments. Note that what we are showing is illustrative of the approach. One would like many more subjects before the data become normative.

When these rational and experimentally controlled data from an alcohol study are used to "calibrate" or mark the other results, it would appear that the chemoradiotherapy treatments exhibit the strongest effect, although we also know (not shown) that this effect recovers when the subjects who survived the treatment were tested 12 months later. Note also that while scopolamine alone has a slight (and mildly significant) effect, when scopolamine is combined with amphetamine this effect is lessened. The altitude study shows a similar relation and even at the highest altitude obtained (23-25,000 feet, the approximate height of Mt. Everest) the effect is no stronger than we found with 2-3 drinks of alcohol (i.e., 0.05 - 0.10

BAC). Although the data are too sparse to conclude confidently, the pattern of the changes is illustrative of the conclusions which may be possible with a larger data base. For example, Grammatical Reasoning (symbol manipulation) appears to be most sensitive in all treatment conditions. Reaction Time, a response speed measure, also appears sensitive. Whether other treatments will show the same effect or not is problematic and awaits further study.

We believe that a completely filled matrix of tests X agents X dosages X mental factor would be extremely useful in providing the estimates of individual and group battlefield performance for entry into combat models.

ACKNOWLEDGMENTS

Support for this project was provided under National Aeronautics and Space Administration Contract NAS9-17326 (COTR Dr. Frank Kutyna), National Science Foundation Grant ISI-8521282 (COTR Dr. Joseph Young), and by Essex IR&D funding.

REFERENCES

1. Spearman, C. (1904). The proof and measurement of association between two things. American Journal of Psychology, 15, 72-101.

2. Allen, M. J., & Yen, W. M. (1979). Introduction to measurement theory. Monterey, CA: Brooks Cole Publishing.

3. Lane, N. E. (1986). Issues in performance measurement for military aviation with applications to air combat maneuvering (NTSC TR-86-008). Orlando, FL: Naval Training Systems Center.

4. Bittner, A. C., Jr., Carter, R. C., Kennedy, R. S., Harbeson, M. M., & Krause, M. (1986). Performance Evaluation Tests for Environmental Research (PETER): Evaluation of 114 measurers. Perceptual and Motor Skills, 63, 683-708.

5. Gulliksen, H. (1950). Theory of mental tests. New York: Wiley.

6. Kennedy, R. S., Wilkes, R. L., Dunlap, W. P., & Kuntz, L. A. (1987). Development of an automated performance test system for environmental and behavioral toxiology studies. Perceptual and Motor Skills, 65, 947-962.

DISCUSSION OF PAPERS PRESENTED IN SESSION IIIA: "PREDICTING HUMAN PERFORMANCE/AVAILABILITY IN COMBAT ENVIRONMENTS"

DISCUSSANT: MG Ennis Whitehead, USA (Retired),
Burdeshaw Associates, Ltd.

There are two major points to be made. First, we must remember that models are tools, not an end unto themselves. If the analyst can answer the decision maker's questions without human variables being represented, then we do not need to include them in the model. The second point is that all human factors are not of equal importance in assessing combat outcome.

If we wish to include human factors in models, which level of model deserves our priority efforts? Figure 1 shows the typical hierarchy of models, lists several current Army models and identifies some of their uses. In my opinion, we should concentrate our human factors improvement efforts on the higher resolution models (one on one or few on few). The higher level models are so aggregated (a Red regiment or Blue division is often the smallest representation) that human factors lose their meaning (an entire division will not be equally fatigued or leaderless at a given moment). The major exception to this guideline is in representing nuclear weapons employment in which a number of battalions maybe incapacitated or otherwise affected simultaneously. Hence the combat power is reduced by more than equipment and immediate personnel losses.

COMBAT MODELS

HIERARCHY	USE
A. MANY ON MANY	RESOURCES
• FORCEM	CAPABILITY
• VIC	COMBAT DEVELOPMENT
B. FEW ON FEW	WEAPONS
• CASTFOREM	EFFECTIVENESS
C. ONE ON ONE	PERFORMANCE
• ENGINEERING	TRADES
• MAN IN THE LOOP	HUMAN FACTORS

FIGURE 1

In deciding which human variables to include, Figure 2 shows what I believe to be the dominant external human factors in combat. Therefore, these are the factors which should be introduced first. Yet if we are examining future capabilities, we do not have the data to represent these factors in the model.

DOMINANT HUMAN FACTORS IN COMBAT

- **LEADERSHIP**
- **TRAINING**
- **CONDITIONING**

THESE MORE THAN ALL OTHER HUMAN FACTORS IMPROVE OR INHIBIT SOLDIER PERFORMANCE

FIGURE 2

In sum, add human factors to models only when they are needed and make sense. In deciding which variables to include, start with the important ones.

DISCUSSANT: Franklin L. Moses,[1] US Army Research Institute, Alexandria, VA

The goal of this session was to display different ways of measuring, modeling, and predicting operational performance.

I've joined you to discuss the presentations as the spirit of Earl Alluisi or, if that's difficult to achieve, than at least as his surrogate. As many of you know, Dr. Alluisi has a keen personal interest in human performance and models. He currently has oversight responsibility in the Office of the Secretary of Defense for R&D in Training and Personnel Systems Technology. He very much looked forward to being here, but had to cancel because of illness in his family.

Today, I'll make a few observations about the session, raise a few questions about what we're trying to achieve, and reflect on how well we are doing. I do this from a background of a research psychologist with experience in applications of human factors, including modeling.

Predicting human performance and capturing data for modeling purposes is a challenge. It is not an easy task to decide on the right measures to use, and in many cases, it is even more difficult to use them successfully. What I heard in today's papers leads me to pose two questions that help focus the points made in the session:

(a) Are the tasks well enough defined so that the purpose of making predictions about human performance is clear?

(b) What measures are appropriate and how may we improve on them?

Both questions relate to the fact that everyday human performance, the kind that we seek to model, occurs under complex contexts and conditions -- people plus their task(s) plus their environment and their interactions. Thus, realistic measures to go along with realistic performances are hard to achieve. For example, you already can appreciate the complexities, even without my elaboration, by considering two of the performance areas reported on today -- flying a helicopter and fighting forest fires.

My co-discussant, MG (R) Ennis Whitehead, and I clearly agree that the session's papers overall are a good sign in the modeling community. Much needed human performance data are being produced and the variety of topics illustrate some good directions -- particularly for individual and small crew performance.

---

[1] The views and opinions in this discussion are those of the author and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other authorized documents.

At least some of the papers, however, are attempting to do the harder job before the easier one: modeling individual instead of group behavior. The variability in measures of individual behavior is always greater than that in measures of group behavior, and individual behavior is always much harder to model reliably. That generalization holds for the modeling of all behaviors, be they of individual human beings or of individual biological or physical particles.

Another concern is to ensure that the measures of behavior and performance are the ones of interest and not just the available or "convenient" measures. For example, measures of physiologically induced changes and stress-induced changes may be reliable indicators of performance only in the extremes and not for typical performances obtained in "usual" circumstances.

In landmark research using appropriate measures of individual <u>and</u> group performance, Dr. Aliuisi has published findings on the behavioral effects of infectious disease. A summary of that work serves to illustrate the challenges of measuring and modeling human performance. The experiment involved two groups of military volunteers (Army Medical Corpsmen) who were exposed to infectious disease. One member of each group of five was not infected, but rather served as a double-blind control. Two types of measures were taken: biochemical and performance. The results showed that the infected participants had essentially identical biochemical responses -- they were equally ill with the infectious disease. The performance measures varied to a maximum -- with high probability, one subject out of eight would show insignificant drops in performance, whereas another one of the eight would show maximum deterioration in performance -- he would become nonresponsive to the physical environment for at least a two to four-hour period during the height of illness. The other subjects demonstrated various degrees of deterioration in their performances, between these two limits. And, remember, the biochemical data indicated all the subjects were objectively, and equally, ill. Average performance, measured over a group with as few as eight volunteer subjects, was quite reliable in spite of the wide spread in individual performances.

I believe that clearly defining tasks and selecting the right measures of human performance to use constitute the core of our modeling challenge. Based on that, I propose four issues in considering the lasting impact of the kinds of data and concepts presented in, or stimulated by, this session:

- How good are questionnaire-estimation techniques for determining likely performance?
- How good are physiological measures as predictors of work-rest requirements, ability to carry a load, and so on?
- How is it possible to measure the *strain* factor in human performance based on information about *stress* in a situation?
- When is it appropriate to make measurements under controlled or simulated conditions and when must we use field conditions?

Overall, what I believe we want to develop is human performance data for use in models that predict military outcomes -- wins and loses. This session suggests to me that we are headed that way. The modeling community is doing its job by including human performance data. The need, as I understand it, is to work with the measurement experts -- the psychologists, physiologists, psychometricians, statisticians, and many others -- to get more group data. Those experts and the modeling experts must work together as teams to ensure that the right data are made available to be used, in the right way, in the right kinds of models, to predict those militarily meaningful behaviors that make the difference between winning and losing.

MAILING LIST
FOR
MMII PROCEEDINGS
as of 8/30/90

NAME	NUMBER OF COPIES
James Bexfield ** IDA Washington, D.C.	1
Jane Clayborne c/o Director: TRAC-LEE ATTN. ATRC-L Ft. Lee, VA 23801-6140 (804-734-3449	1
Dr. Russell Coile VRC Corporation 2150 Garden Road Suite B3 Monterey, California 83940-5327 (408) 372-3439	1
John Friel ** Rand Corporation Los Angeles, California	1
Mr. Ed Jones * Jones and Company Post Office Box 7516 Carmel Valley, CA 93924 (408) 373-8507	1
Dr. Gary Klein Klein Associates, Inc. Post Office Box 264 Yellow Springs, Oklahoma 45387 (513) 767-2691	1
Stephen A. Murtaugh Calspan Corporation Post Office Box 400 Buffalo, New York 14225 (716) 631-6824	4
Mr. William D. Randolph PRB Associates, Inc. 47 Airport View Drive Hollywood, MD 20636 (301) 373-2360	1
Jim Rice Opti-Metrics, Inc. 2008 Hogback Road Ann Arborn, Michigan 48105 (313) 973-1177	1

* Please note change of address.

** Address given incomplete - to be provided by MORS Office.