



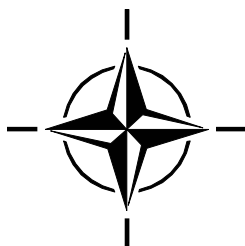
RTO TECHNICAL REPORT

TR-HFM-128

Human Behavior Representation in Constructive Simulation

(La représentation du comportement humain
dans la simulation constructive)

Final Report of Task Group 128,
Consolidating the Findings of Task Group 143.



Published September 2009



RTO TECHNICAL REPORT

TR-HFM-128

Human Behavior Representation in Constructive Simulation

(La représentation du comportement humain
dans la simulation constructive)

Final Report of Task Group 128,
Consolidating the Findings of Task Group 143.

The Research and Technology Organisation (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote co-operative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective co-ordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also co-ordinates RTO's co-operation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of co-operation.

The total spectrum of R&T activities is covered by the following 7 bodies:

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS System Analysis and Studies Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These bodies are made up of national representatives as well as generally recognised 'world class' scientists. They also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier co-operation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

The content of this publication has been reproduced
directly from material supplied by RTO or the authors.

Published September 2009

Copyright © RTO/NATO 2009
All Rights Reserved

ISBN 978-92-837-0074-6

Single copies of this publication or of a part of it may be made for individual use only. The approval of the RTA Information Management Systems Branch is required for more than one copy to be made or an extract included in another publication. Requests to do so should be sent to the address on the back cover.

Table of Contents

	Page
List of Figures	viii
List of Tables	ix
List of Abbreviations	x
Glossary of Terms	xvii
Research Study Group HFM-128 Programme Committee	xxii
Acknowledgement	xxiii

Executive Summary and Synthèse	ES-1
---------------------------------------	-------------

PART ONE – HUMAN BEHAVIOR IN MODELING

Chapter 1 – The Scope of this Report	1-1
1.1 The Goal of HFM-128 on Representation of Human Behavior in Constructive Modeling	1-1
1.2 Target Audience	1-1
1.3 Main Deliverable	1-1
1.4 A Brief Outline	1-2
1.5 Limits of this Study	1-2
1.6 Summaries of Preceding Reports	1-2
1.7 A Scenario Demonstrating the Use of HBR	1-5
Chapter 2 – Human Factors, Artificial Intelligence, and Cognitive Architectures	2-1
2.1 Human Factors, Human Performance, and Human Behavior	2-1
2.2 HF Framework	2-1
2.3 Representation of Human Information Processing	2-2
2.4 Representation of Physical Drivers for Behavior	2-4
2.5 Artificial Intelligence and Human Behavior Representation	2-5
2.6 Choosing a Modeling Architecture	2-6
2.7 Architectural Reviews	2-6
2.8 Architecture Comparison	2-8
Chapter 3 – Modeling Effects Based Operations	3-1
3.1 Effects Based Operations in Practice and in Modeling	3-1
3.2 Technical Architecture of Operational Models	3-2
3.3 Requirements for the Analysis Environment	3-3
3.4 Hierarchy of Aims	3-4
3.5 Operational Performance	3-5

3.6	Handling Intelligence in the Modeling Structure	3-8
3.7	Performance and Learning from Experience in EBO	3-9
3.8	Learned Behavior	3-10
3.9	Sequence of Events	3-10

PART TWO – DISCUSSION OF KEY HUMAN FACTORS

Chapter 4 – Perception 4-1

4.1	Introduction	4-1
4.2	Basic Concepts	4-2
4.2.1	Sensation versus Perception	4-2
4.2.2	The Senses	4-2
4.2.3	Locus of Information	4-3
4.2.4	Attention	4-4
4.3	Models of Perception	4-5
4.3.1	Stimulus Properties	4-5
4.3.2	Physiological Properties	4-6
4.3.3	Psychological Properties	4-7

Chapter 5 – Cognition in Human Behavior Representation 5-1

5.1	Introduction	5-1
5.2	Cognitive Science Contributions to HBR	5-3
5.3	High-Level HBR Cognitive Capabilities	5-4
5.3.1	Situation Awareness	5-4
5.3.2	Problem Solving and Decision Making	5-6
5.3.3	Planning	5-7
5.4	Low-Level Cognitive Processes	5-8
5.4.1	Memory	5-8
5.4.1.1	Episodic Memory	5-8
5.4.1.2	Short Term Memory	5-9
5.4.1.3	Semantic Memory	5-10
5.4.1.4	Prospective Memory	5-11
5.4.1.5	Categorization	5-11
5.4.1.6	Learning	5-12
5.4.2	Reasoning and Thinking	5-13
5.4.3	Judgment and Choice	5-17
5.4.4	Belief and Confidence	5-18
5.5	Observations	5-19
5.6	Additional Reading	5-20

Chapter 6 – Sources of Variation in Human Cognition, Behavior and Performance 6-1

6.1	Introduction	6-1
6.2	Moderators	6-1
6.3	Moderators in Military Simulations	6-2

6.3.1	Important Moderators	6-2
6.4	Characterizing the Effects of Moderators	6-4
6.5	Representation of External Moderators	6-5
6.5.1	Performance Metrics	6-5
6.5.2	Operator State as a Predictor of Performance	6-6
6.5.3	Use of Performance Shaping Functions	6-7
6.6	Modeling the Effect of State on Task Performance	6-7
6.6.1	Engineering Approach	6-10
6.6.2	Constructive Modeling	6-10
6.7	Multiple Stressors	6-10
6.8	Implementation Issues	6-10
6.9	Modeling Task Demand	6-11
6.10	Representation of Internal Moderators	6-12
6.10.1	Systematic Component	6-12
6.10.1.1	Illustrative Example	6-12
6.10.2	Non-Systematic Component	6-14
6.11	Team and Organizational Moderators	6-15
6.11.1	Teams	6-15
6.11.2	Cohesion	6-16
6.11.3	Leadership	6-16
6.11.4	Team Diversity	6-16
6.11.5	Social Support	6-16
6.11.6	Interactions	6-16
6.11.7	Structured Teams	6-17
6.11.8	Unstructured Groups	6-19
6.11.9	Crowds	6-19
6.11.10	Culture	6-20
6.12	Moderator Summary	6-21
Chapter 7 – Motor Task Requirements for Human Behavior Representation		7-1
7.1	Introduction	7-1
7.2	Eye Movements or Oculomotor	7-4
7.3	Speech or Vocal	7-6
7.4	Hands or Manual	7-7
7.5	Legs, Trunk, and Whole Body Movement	7-10
7.6	Motor Tasks, Workload Channel Limitations and Inter-Channel Conflicts	7-11
Chapter 8 – Task Execution and Errors		8-1
8.1	Task Networks: Representation, Selection and Resource Limits	8-2
8.2	Fundamentals of Error	8-6
8.3	A Network-Centric View of Error: The Gap between the Real and the Perceived World	8-9
8.4	Representing Error in Task Execution	8-10
8.5	Additional Error Factors	8-11

PART THREE – THE FUTURE OF HUMAN BEHAVIOR REPRESENTATION

Chapter 9 – Operationalization of Human Behavior in the Model Environment	9-1
9.1 A Proposed Framework for Human Factors in Simulated Operations	9-1
9.2 Good Modeling Practice	9-1
9.3 Implementation of the Good Practice in an Operational Model Architecture	9-3
9.4 Suitability for Application Fields	9-5
9.5 Operationalization Conclusions	9-6
9.5.1 HBR is on its Way	9-6
9.5.2 Observed Weaknesses	9-6
9.5.3 Human States and Task Performance	9-7
9.5.4 New Developments in Modeling	9-7
9.5.5 Architectures are Still Stove Piped	9-7
9.5.6 A Recommended Practice to Involve HF	9-7
 Chapter 10 – Verification and Validation	 10-1
10.1 Verification	10-1
10.2 Validation	10-1
10.3 Model Components	10-3
10.4 Validation and Standardization	10-3
10.5 Statistical Methods in Experimentation and Simulation	10-5
 Chapter 11 – Future Challenges	 11-1
11.1 Scalability, Aggregation, and Levels of Analysis	11-1
11.2 Society Level Modeling	11-2
11.2.1 Introduction	11-2
11.2.2 From Individuals to Societies	11-2
11.2.3 Operational Needs	11-3
11.2.4 Subject Matter Domains	11-4
11.3 Data Acquisition	11-6
11.4 Current Significant Gaps	11-6
11.5 Rapid Prototyping	11-7
 Chapter 12 – Conclusions and Recommendations	 12-1
12.1 Modeling Practice	12-1
12.1.1 Conclusions	12-1
12.1.2 Recommendations to the Scientific Community	12-1
12.1.3 Recommendations to NATO	12-1
12.2 Modularity	12-2
12.2.1 Conclusions	12-2
12.2.2 Recommendations to the Scientific Community	12-2
12.2.3 Recommendations to NATO	12-2
12.3 Validity	12-2
12.3.1 Conclusions	12-2

12.3.2	Recommendations to the Scientific Community	12-3
12.3.3	Recommendations to NATO	12-3
12.4	Human Error	12-3
12.4.1	Conclusions	12-3
12.4.2	Recommendations to the Scientific Community	12-3
12.4.3	Recommendations to NATO	12-3
12.5	Interoperability	12-4
12.5.1	Conclusions	12-4
12.5.2	Recommendations to the Scientific Community	12-4
12.5.3	Recommendations to NATO	12-4
12.6	Functional Extensions	12-4
12.6.1	Conclusions	12-4
12.6.2	Recommendations to the Scientific Community	12-5
12.6.3	Recommendations to NATO	12-5

Chapter 13 – References	13-1
--------------------------------	-------------

List of Figures

Figure		Page
Figure 2-1	An Approach to Modeling Operator Performance Incorporating HF	2-2
Figure 2-2	The Widely Used Modified Stage Model for Human Information Processing	2-3
Figure 2-3	Graphical Representation of MDS Analysis of Architecture Comparisons	2-9
Figure 3-1	A Schematic of a Generic Modeling and Simulation Environment	3-3
Figure 3-2	An Operational Model in an Analysis Environment	3-4
Figure 3-3	Goal Driven Behavior of the Organization as a Hierarchy of Sub-Goals (Aims)	3-5
Figure 3-4	Likelihood that a Simulation can be Run Based on Availability of Key Personnel	3-8
Figure 3-5	Goal Directed Behavior	3-9
Figure 3-6	Modification of Behavior Through Training	3-10
Figure 4-1	Considerations for Perceptual Modeling – the Links between the Physical Environment, the Sensations and Perception	4-5
Figure 5-1	Scheme for Implementing HBR in Operational Studies	5-3
Figure 5-2	Mapping Cognitive Concepts and Processes on the Stages of SA	5-5
Figure 6-1	Conceptual Model of Perception	6-4
Figure 6-2	The HFM-128 Philosophy Approach to Modeling Operator Behavior	6-6
Figure 6-3	Conceptual Flow of Variable Influence from Environment to Performance	6-7
Figure 6-4	Task Execution Leads to Performance	6-7
Figure 6-5	Model Chain for Sleep Deprivation and Circadian Rhythm	6-9
Figure 6-6	Illustrative Approach to Coping Style	6-13
Figure 6-7	State Variable Decision Effectiveness and Performance	6-14
Figure 6-8	Team Characteristics, Leading to Team Effectiveness	6-17
Figure 6-9	Relationship between Moderators, Teamwork Behaviors and Team Performance	6-18
Figure 8-1	Notional Discrete Event Task Network with Branching	8-2
Figure 8-2	Behavior Selection and Resulting Performance Outcome Dependency on Task Priority	8-4
Figure 8-3	Task Performance is Moderated by Resource Demand and Availability	8-5
Figure 8-4	Three Levels of Accident Evolvment (Adapted from K��ppler, 2006, 2008)	8-7
Figure 8-5	The “Swiss Cheese” Model of Human Error (Adapted from Reason, 1990)	8-8
Figure 8-6	The HFACS Framework (Adapted from Wiegmann and Shappell, 2003)	8-9
Figure 9-1	Synthesized Model Including HF in Operational Performance	9-4
Figure 9-2	Modeling Requirements Analysis for Various Military Applications	9-6
Figure 11-1	Time Scales of Human Action and Culture-Related Research Areas	11-3

List of Tables

Table		Page
Table 3-1	Definitions in the Operational Context	3-7
Table 6-1	Environmental Stressors Important for Military Simulations	6-3
Table 6-2	Characteristics of Individuals and Groups Important for Military Simulations	6-3
Table 6-3	Summary of Moderator Models	6-22
Table 7-1	Integrative Architectures, Motor Column of Table 3.1 in Pew & Mavor (1998)	7-2
Table 7-2	MTM: Fifteen Fundamental Motions for Body Movements (After Schlick, 2006)	7-3
Table 7-3	Eye and Head Movements (ARL, 2005)	7-5
Table 7-4	Speaking and Listening Rates (ARL, 2005)	7-7
Table 7-5	Cognitive Activity Durations (ARL, 2005)	7-8
Table 7-6	Motor or Hand Control Duration (ARL, 2005)	7-8
Table 7-7	Properties Important for Realistic Body-Driven Simulations (Templeman et al., 2006)	7-11

List of Abbreviations

3D	Three dimensional
AAMC	Army Aviation and Missile Command (USA)
ACME	Analogical Constraint Mapping Engine
ACT-R	Adaptive Control of Thought-Rational, http://act.psy.cmu.edu/
ACT-R/PM	Adaptive Control of Thought-Rational/Perceptual Motor
AFRL	Air Force Research Laboratory (USA)
AGARD	Advisory Group for Aerospace Research and Development (NATO)
AI	Artificial Intelligence
AMBR	Agent-based Modeling and Behavior Representation
AMCOM	Aviation and Missile COMmand (U.S. Army)
ANN	Artificial Neural Network
ARIADNE	http://www.ariadne-sms.com/
ARL	Army Research Laboratory (USA)
ART	Adaptive Resonance Theory
ARTMAP	Adaptive Resonance Theory (MAP)
BAE	British AErospace
BCDMEM	Bind Cue Decide Model of Episodic Memory
BDI	Belief, Desire, Intent
BRAHMS	Business Redesign Agent-based Holistic Modeling System
BRIMS	Behavior Representation In Modeling and Simulation (Conference)
C2	Command and Control
C3I	Command, Control, Communication and Intelligence
C4ISR	Command, Control, Communication, Computers, Intelligence, Surveillance, Reconnaissance
CAN	Canada
CAEn	Close Action Environment
CBR	Case Based Reasoning
CBRN	Chemical, Biological, Radiological, Nuclear
CCIS	Command and Control Information Systems
CGF	Computer Generated Force
CHS	Centre for Human Sciences (formerly DERA, now QinetiQ)
COGNET/iGEN	COGnition as a NETwork of Tasks

COPE	Coping Orientation for Problem Experiences
COPYCAT	http://www2.psy.uq.edu.au/CogPsych/Copycat/Tutorial/overvw.htm
COVIS	COmpetition between Verbal and Implicit Systems
DCOG	Distributed COGnition
DIGuy	http://www.diguy.com/diguy/index.htm
DIME-PMESII	Diplomatic, Informational, Military and Economic actions to accomplish Political, Military, Economic, Social, Infrastructure and Informational effects
DM	Decision Making
dMARS	Distributed Multi-Agent Reasoning System
DMSO	Defense Modeling and Simulation Office (USA)
DMT	Distributed Mission Training
DND	Department of Defence (CAN)
DoD	Department of Defense (USA)
DRDC	Defence Research and Development Canada (CAN)
D-OMAR	Distributed Operator Model ARchitecture
DSTO	Defence Science and Technology Organization (AUS)
DV	Dependent Variable
EASE	Elements of ACT-R, SOAR and EPIC
EBO	Effects Based Operations
EFC	Emotion Focused Coping
EPIC	Executive Process Interactive Control
EWS	Electronic Warfare System
EXODUS	http://fseg.gre.ac.uk/exodus/
FEDEP	FEderation Development and Execution Process (HLA)
FLAMES	FLexible Analysis, Modeling, and Exercise System (http://www.ternion.com/)
FOM	Federation Object Model (HLA)
FSM	Finite State Machine
G	Gravity (acceleration)
GBR	United Kingdom
GCI	Ground Controlled Intercept
GOMS	Goals, Operators, Methods and Selection rules
GPS	Global Positioning System
GRADE	GRaphical Agent Development Environment
GTACS	Ground Theater Air Control System

GT-ASP	Georgia Tech Aegis Simulation Program
GUI	Graphical User Interface
HARDMAN	Hardware versus Manpower
HB	Human Behavior
HBM	Human Behavioral Models
HBR	Human Behavioral Representation
HF	Human Factors (Ergonomics)
HFACS	Human Factors Analysis and Classification System
HFE	Human Factors Engineering
HFM	Human Factors and Medicine
HIL	Human-In-the-Loop
HLA	High Level Architecture, http://www.dmsomil/
HOS	Human Operator Simulator
HOMER	Human Operator Modeling Expert Review
HPM	Human Performance Model/Modeling
HQ	HeadQuarters
HTA	Hierarchical Tasks Analysis
HUM	HUMAN Resources and Performance (TTCP)
IA	Intelligent Agent
ICT	Institute for Creative Technologies, http://www.ict.usc.edu
IDP	Initial Development Phase
IED	Improvised Explosive Device
IEEE	Institute of Electrical and Electronics Engineers (www.ieee.org)
IFOR	Intelligent FORces
iGEN	Trademark name, Chi Systems Incorporated
I/ITSEC	Interservice/Industry Training, Simulation and Education Conference
IJCAI	International Joint Conference on Artificial Intelligence
IMCN	Intelligent Mission Controller Node
IMPRINT	IMproved Performance Research INtegration Tool
IP	Information Processing
IP/PCT	Information Processing/Perceptual Control Theory
IPME	Integrated Performance Modeling Environment, http://www.maad.com
IUSS	Integrated Unit Simulation System
iWars	Infantry Warrior System

JACK	JAVA Agent Compiler and Kernel
JAVA	Trademark name, Sun Microsystems Limited
JBASS	JSIMS Breakout Analysis Spread Sheet (USA)
JFACC	Joint Forces Air Combat Commander
JORN	Jindalee Operational Radar Network
JSAF	Joint Semi-Automated Forces (USA)
JSIMS	Joint SIMulation System (USA)
JWARS	Joint WARfare System (USA)
JWFC	Joint WarFighting Center (USA)
KA/KE	Knowledge Acquisition/Knowledge Engineering
KADS	Knowledge Acquisition, Documentation and Support
KBS	Knowledge Based System
LISA	Learning and Inference with Schemas and Analogies
NAWCTSD	Naval Air Warfare Center Training Systems Division
MAAD	Micro Analysis And Design Inc.
MALO	Maritime Air Littoral Operations
MAV	Micro-Air Vehicle
MDAPS	Major Defense Acquisition ProgramS
MDS	Multi Dimensional Scaling
METT-T	Mission Enemy Time Troops Terrain
MHP	Maritime Helicopter Project
MIDAS	Man-machine Integration Design and Analysis System (http://human-factors.arc.nasa.gov/dev/www-midas/)
MoD	Ministry of Defence (GBR)
ModSAF	Modular Semi-Automated Forces
MoE	Measure of Effectiveness
MOOTW	Military Operations Other Than War
MOUT	Military Operations in Urbanized Terrain
MoM	Measure of Merit
MoP	Measure of Performance
MNS	Mission Need Statement
MRE	Mission Rehearsal Exercise
M&S	Modeling and Simulation
MTM	Methods-Time Measurement (Ergonomics)

MTRS	Mission-Training and Rehearsal System
Myriad/ SIMULEX	http://www.crowddynamics.com/Egress/simulex.html
NASM	National Air and Space Model (USA)
NAF	Numbered Air Force
NATO	North Atlantic Treaty Organisation
NBC	Nuclear, Biologic, Chemical
NLP	Natural Language Processing
NRC	National Research Council
NSI	Natural Selection Inc.
OA	Operations Analysis
OMAR	Operator Model ARchitecture
OneSAF	One Semi-Automated Force
ONR	Office of Naval Research (USA)
OOTW	Operations Other Than War
OPFOR	OPposing FORce
OR	Operations Research
ORD	Operational Requirements Document
OT&E	Operational Test and Evaluation
P2NBC2	Physiological and Psychological Effects of NBC and Extended Operations on Crew Program
PAC	Personality-based Architecture for Cognition
PAM	Preparatory Attentional Memory
PFC	Problem Focused Coping
PM	Perceptual-Motor
PMFServ	Performance Moderator Function Server
POP	Prediction of Operator Performance
POPIP	Prediction of Operator Performance Information Processing
PSF	Performance Shaping Function
PSYOPS	PSYchological OPERATIONs
R&D	Research and Development
RAAF	Royal Australian Air Force
REM	Retrieving Effectively from Memory
RESA	Research, Evaluation, and Systems Analysis (USA)
RO	Radar Officer

ROE	Rules of Engagement
RPICN	Role Player Intelligent Controller Node
RPD or RPDM	Recognition-Primed Decision Making
RSSB	Railway Standards and Safety Board (GBR)
RTI	Run-Time Infrastructure (HLA)
RTO	Research and Technology Organisation (NATO)
SA	Situation Awareness or Situation Assessment
SAF	Semi-Automated Forces
SAFE	System for Aircrew Fatigue Evaluation (QinetiQ Ltd., GBR)
SAFEWORK	http://www.safework.com/solutions.html
SAFTE	Sleep, Activity, Fatigue, and Task Effectiveness (SAIC Ltd., USA)
SAIC	Science Applications International Corporation
SAINT	Systems Analysis of Integrated Networks of Tasks
SAMPLE	Situation Awareness Model for Pilot-in-the-Loop Evaluation
SAP	Situation Awareness Panel
SAS	System Analysis and Studies
SBA	Simulation Based Acquisition
SBIR	Small Business Innovative Research
SC	Soft Computing
SCOPE	Soldier Capability Optimization for Projected Efficacy
SCOTT	Synthetic Cognition for Operational Team Training
SECO	Synthetic Environment Co-ordination Office (GBR)
SEM	Sensor Employment Manager
SIMNET SAF	SIMulation NETwork Semi-Automated Force
SIMNET-T	SIMulation NETwork - Trainer
SM	Semantic Memory
SME	Subject Matter Expert
SOAR	State, Operator And Result, http://ai.eecs.umich.edu/soar/
SOP	Standard Operating Procedures
SP	Syntagmatic-Paradigmatic
STEMM	Simulation of TEam Mental Models
STM	Short Term Memory
STOW	Synthetic Theatre Of War
SUSTAIN	Supervisory and Unsupervisory STRatified Adaptive Incremental Network

SuW	Surface Warfare
TADMUS	Tactical Decision Making Under Stress (USA)
TAMSS	Tactical Aviation Mission System Simulation (CAN)
TACCO	Tactical Coordinator
TAP	Technical Activity Proposal
TBMCS	Theater Battle Management Core System
TLAR	That Looks About Right
T&E	Test and Evaluation
TCT	Time Critical Target
TODAM	Theory Of Distributed Associative Memory
TOEFL	Test Of English as a Foreign Language
TP	Technical Panel (TTCP)
TR	Technical Report
TTCP	The Technology Cooperation Program
U.S., USA	United States of America
USAF	United States Air Force
USARIEM	United States Army Research Institute for Environmental Medicine
USN	United States Navy
VAE	Virtual Air Environment
VACP	Visual, Auditory, Cognitive, Psychomotor (demand)
VR	Virtual Reality
VV&A	Verification, Validation and Accreditation
VVE&T	Verification, Validation, Evaluation and Testing
WARSIM	WAR fighter's SIMulation
W/Index	Workload Index
WISSARD	What If Simulation System for Advanced Research and Development
WOC	Wing Operation Center
XML	Extensible Markup Language

Glossary of Terms

Alertness	Watchfulness: the process of paying close and continuous attention. From: http://wordnet.princeton.edu/perl/webwn .
Architecture	<p>Design: the way components fit together to form a unified system. May be conceived of any complex system such as “software architecture” or “network architecture” [Free On-line Dictionary of Computing]. An IT architecture is a design for the arrangement and interoperation of technical components that together provide an organization (of) its information and communication infrastructure. From: http://www.balancedscorecard.org/basics/definitions.html.</p> <p>(Cognitive) architecture: (1) specifications of the main modules and mechanisms underlying human cognition; (2) the computer program implementing these specifications (Ritter et al., 2002, p. 4).</p>
Attention	<p>The cognitive process of selectively concentrating on one thing while ignoring other things. From: http://en.wikipedia.org/wiki/Attention.</p> <p>The process whereby a person concentrates on some features of the environment to the (relative) exclusion of others. From: http://wordnet.princeton.edu/perl/webwn?s=attention.</p>
Behavior	Purposeful actions of a human being in a meaningful situation including observable (physical) actions as well as unobservable (cognitive) processes.
Categorization: Information Integration Tasks	Optimal categorization performance requires the integration of several variables to define a decision boundary. Unlike rule-based tasks, in information-integration tasks, the optimal strategy for combining variables is not easy to verbalize (such as describing the visual features that differentiate one person’s face from all others).
Categorization: Prototype Distortion Tasks	Sets of stimuli are generated by first creating a prototype for each category and then generating exemplars through random distortion. To assess category learning, the categories are new to the subjects, with patterns of dots commonly being used as. Two basic paradigms are frequently used: the (A, not A) paradigm in which subjects are given patterns belonging to category A, and are asked to decide whether or not new stimuli belong to this category and the (A, B) paradigm in which labeled exemplars of each category are presented, with subjects requiring to categories novel items accordingly.
Categorization: Rule Based Tasks	The subject’s task is to learn the rule that distinguishes one category from another. A common method to use in the laboratory is the Wisconsin Card Sorting Test in which subject learn the rule that separates piles of cards. The basis on which objects are assigned to categories can range in complexity from fairly low-level perceptual characteristics to higher level semantic ones, and as such can require the use of several memory systems to perform the task.
Categorization: Weather Prediction Tasks	Categorical membership is probabilistic rather than deterministic. During training, an example may be described as belonging to category A 60% of the time, and category B for the remaining 40%.

Categorization: Weather Prediction Tasks (cont'd)	Weather prediction tasks differ from the other tasks mentioned above primarily in the variance of the strategies that subjects can use to solve the task. According to Gluck, Shohamy, and Myers (2002), subjects can utilize non-verbal information-integration rules, explicit rules, or explicit memorization. Thus, the neural categorizations pathways that are employed, and hence the models that would predict performance, vary with the individual subject as a choice of strategy, rather than being a choice that is related to the type of categorization task the subject is doing.
Cognition	<p>High level functions carried out by the human brain, including comprehension and use of speech, visual perception and construction, calculation ability, attention (information processing), memory, and executive functions such as planning, problem-solving, and self-monitoring.</p> <p>The conscious processes of the mind by which individuals perceive, think, and remember.</p> <p>Cognition encompasses perception, imagination, judgment, memory, and language. It includes the processes people use to think, decide, and learn.</p> <p>The mental functions and mental processes of people, with a particular focus toward the study of comprehension, inferencing, decision making, planning, learning, abstraction, generalization, concretization/specialization and meta-reasoning involving such concepts as beliefs, knowledge, desires, preferences and intentions. Adapted From: http://en.wikipedia.org/wiki/Cognition.</p>
Cohesion	In teams, the commitment of a team's individuals to work together towards a common goal in a supportive manner.
Computer Generated Forces (CFG)	A generic term used to refer to computer representations of forces in simulations that attempt to model human behavior sufficiently, so that the forces will take some actions automatically (without requiring man-in-the-loop interaction). Also referred to as semi-automated forces (SAF). From: http://www.sedris.org/glossary.htm#C_grp .
Constructive Models	Abstractions from the reality to simulate life events, in which elementary processes are modeled and connected in a constructive way to allow for simulation of a variety of composed events.
Crowd	A collection of people in a common location acting as individuals but whose actions, beliefs and goals may be coupled.
Culture	Patterns of human activity and the symbolic structures that give such activity significance ... includes technology, art, science, as well as moral systems. From: http://en.wikipedia.org/wiki/Culture .
Decision Making	The cognitive process leading to the selection of a COA among variations ... decision making is a reasoning process that can be rational or irrational, and can be based on explicit assumptions or tacit assumptions . From: http://en.wikipedia.org/wiki/Decision_making .
Effects Based Operations	<p>A process for obtaining a desired strategic outcome or effect on the enemy through the synergistic and cumulative application of the full range of military and non-military capabilities at all levels of conflict.</p> <p>An approach to planning, executing and assessing military operations with an explicit focus on effects as opposed to targets or even objectives.</p>

Error	Lack of correlation between the intended consequences of an action and the actual consequences of that action. From: http://en.wikipedia.org/wiki/Error .
G	Acceleration expressed relative to gravitational acceleration at sea level; more common in aviation.
Goal	An objective or desired intermediate state; may be an internal state of the operator or an external, environment state, but a goal always represents the operator's internal perception of the desired condition.
Human Behavior	The collection of behaviors exhibited by human beings and influenced by culture, attitudes, emotions, values, ethics, authority, rapport, hypnosis, persuasion, coercion and/or genetics. From: http://en.wikipedia.org/wiki/Human_behavior .
Human Behavior Representation (HBR)	A computational model of a human being to achieve embodied goals and predict performance, expressing observed variability in behavior attributable to differences in the person's characteristics, to differences in the situation or to the interplay of both, mapping characteristics of empirical phenomena into values of parameters and models in an artificial world. From: RTO-TR-047 AC/323(SAS-017)TP/25.
Human Factors (HF)	<p>The scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and other methods to design in order to optimize human well-being and overall system performance. From: International Ergonomics Association, August 2000.</p> <p>The variables, parameters and relationships among human performance or behavior, human attributes and characteristics, and the environment in which the human is working.</p>
Human Performance	Measurement or assessment, usually quantitative, of the execution of a task for a given set of conditions.
Knowledge Acquisition	Includes the elicitation, collection, analysis, modelling and validation of knowledge for knowledge engineering and knowledge management projects.
Model	<p>A simplified representation at a conceptual level of (a part of) the real world and/or the way it behaves, that suffices to make some deductions concerning the real world and/or its functioning. A model consists of components and the relationships between those components, which are generally cause-effect relationships. From: RTO-TR-047 AC/323(SAS-017)TP/25.</p> <p>Theories implemented as computer programs or represented mathematically to apply to specific situations or types of situations (Ritter et al., 2002, p. 4).</p>
Moderator	An operator state, trait or relationship among variables that affects behavior or performance. Environmental conditions (states) affect operator performance through moderators rather than affecting performance directly.
Motor Tasks	Tasks that require an organism to utilize their skeletal muscles effectively in a goal-directed manner. Motor skills and motor control depend upon the proper functioning of the brain , skeleton , joints , and nervous system .
Multitasking	The apparent process of performing more than one task, action or mental process at the same time.

Ontology	A formalised representation of the knowledge in a domain taken from a particular perspective or conceptualisation. The main use of an ontology is to share and communicate knowledge, both between people and between computer systems.
Perception	<p>In psychology and the cognitive sciences, perception is the process of acquiring, interpreting, selecting, and organizing sensory information. From: http://en.wikipedia.org/wiki/Perception_(album).</p> <p>Awareness of the world and its contents through sensory experience.</p> <p>The active psychological process in which stimuli are selected and organized into meaningful patterns.</p> <p>The psychological ability to process or use information received through the sense organs.</p>
Performance Shaping Factors	A factor or phenomenon that changes the predicted outcome of a model. It is also colloquially applied to the underlying model that may be used to determine how that factor evolves in a simulation.
Performance	<p>Executing and finishing a certain task.</p> <p>Expression of the competence, moderated by some personal or situational variables.</p> <p>A description of how well a COA is executed.</p> <p>From: RTO-TR-047 AC/323(SAS-017)TP/25.</p>
Scenario	An account or synopsis of a projected COA, events or situations. Scenario development is used when organizations wish to test strategies against uncertain future developments. The detail in what is called a scenario differs between nations.
Semiautomated Forces (SAF)	<p>See Computer Generated Forces.</p> <p>Semiautomated refers to the need for human intervention to provide higher-level decision making and goal setting while lower-level behaviors occur automatically according to preset, usually simple, rules.</p>
Sensation	<p>The initial contact between the perceptual system and the relevant stimuli within the external environment.</p> <p>The interactions between the physiological mechanisms (i.e. sense organs) that respond directly to external stimulation.</p> <p>In psychology, sensation is the first stage in the chain of biochemical and neurologic events that begins with the impinging of a stimulus upon the receptor cells of a sensory organ, which then leads to perception. From: http://en.wikipedia.org/wiki/Sensation.</p>
Simulation	<p>A method to implement a model in some environment or in a device that may be totally or partially artificial (instead of real).</p> <p>A technique for analyzing, testing, evaluating the effect of some values of the parameters of the model on other parameters. For example, a decision making process (e.g., choose a tactical plan among three alternatives) can be simulated on a computer, using some algorithm (among others, a decision tree).</p> <p>From: RTO-TR-047 AC/323(SAS-017)TP/25.</p>

Situation Awareness	Situational awareness is the result of the perception of a number of elements in the environment within a given time frame and space, their meaning with respect to the mission at hand and their possible evolution in the near future that must be taken into account in determining one's own behavior. (Endsley, 1995).
State	An internal operator attribute that varies over the duration of the simulation. State variables are context dependent and may be a trait variable in different application.
Stressor	An external state change (such as a change of temperature or salinity) that results in a physiological response from an organism required to maintain homeostasis .
Task	A purposeful unit of operator activity, usually with an observable output, although a reasoning task may have no associated behavior until a decision is acted upon by the operator.
That Looks About Right (TLAR)	A fanciful form of "validation" that relies on subjective opinion of suitability and representativeness without supporting objective measures of performance; a crude approximation to the Turing test and often the default level of validation for agent models.
Team	At least two people, who are working together towards a common goal, where each person has been assigned a specific role or function to perform, and where completion of the goal requires some dependency among group members. Dyer (1984)
Trait	An operator attribute that remains constant for the duration of the simulation. Traits are context dependent and what may be a trait in one application may be a state in another application.

Research Study Group HFM-128

Programme Committee

Chairman

Dr. Wouter Lotens, PhD.
Senior Advisor, TNO Human Factors / Human Performance
PO Box 23, 3769DE Soesterberg
NETHERLANDS
Telephone: +31 346 356 203 / Facsimile: +31 346 353 977
E-mail: wouter.lotens@tno.nl

Members

Dr. Laurel Allender, PhD.
U.S. Army Research Laboratory
Human Research and Engineering Directorate
Attn: AMSRD-ARL-HR-SE
Aberdeen Proving Ground, MD 21005-5425
USA
Telephone: 1-410-278-6233
Email: laurel.allender@us.army.mil

Mr. Joe Armstrong
CAE Professional Services
1135 Innovation Drive, Suite 200
Kanata, Ontario K2K 3G7
CANADA
Telephone: 1-613-247-0342
Facsimile: 1-613-225-4281
Email: jarmstrong@cae.com

Dr. Andrew Belyavin, PhD.
Centre for Human Sciences
A50 Building, Cody Technology Park
Ively Road, Farnborough
Hampshire GU14 0LX
UNITED KINGDOM
Telephone: +44 1 252 393 401
Facsimile: +44 1 252 392 097
Email: ajbelyavin@qinetiq.com

Mr. Brad Cain, M.A.Sc.
DRDC Toronto, Human System Integration Section
Human System Engineering Group
1133 Sheppard Avenue West
Toronto, Ontario M3M 3B9
CANADA
Telephone: 1-416-635-2195
Facsimile: 1-416-635-2013
Email: brad.cain@drdc-rddc.gc.ca

Mr. M. Castor, M.Sc.
Senior Researcher
Department of Man System Interaction
Swedish Defence Research Institute, FOI
Gullfösgatan 6
164 90 Stockholm
SWEDEN
Telephone: +46 8 55503097
Facsimile: +46 8 55503700
Email: martin.castor@foi.se

Dr. Kevin Gluck, PhD.
Senior Research Psychologist
S&T Advisor
Cognitive Models and Agents Branch
Air Force Research Laboratory
6030 S. Kent Street
Mesa, AZ 85212-6061
USA
Telephone: 1-480-988-6561 x-677
DSN 474-6677
Facsimile: 1-480-988-2230
Email: Kevin.Gluck@mesa.afmc.af.mil

Dr. Wolf Dieter K  ppler
Research Institute for Communication
Information Processing and Ergonomics
Neuenahr Str. 20
53343 Wachtberg
GERMANY
Telephone: +49 228 9435 401
Facsimile: +49 228 9435 508
Email: k  ppler@fgan.de

Dr. Peter Kwantes, PhD.
DRDC Toronto, Human System Integration Section
1133 Sheppard Avenue West
Toronto, Ontario M3M 3B9
CANADA
Telephone: 1-416-635-2028
Facsimile: 1-416-635-2013
Email: Peter.Kwantes@drdc-rddc.gc.ca

Mr. Mikael Lundin
Department of Man System Interaction
Swedish Defence Research Institute, FOI
Gullfossgatan 6
164 90 Stockholm
SWEDEN
Facsimile: +46 8 55503700
Email: mikael.lundin@foi.se

Dr. Gina Thomas-Meyers, PhD.
Air Force Research Laboratory
AFRL/RHCS
2698 G Street
Wright Patterson AFB
OH 45433-7604
USA
Telephone: 1-937-255-9474
Email: Gina.Thomas-Meyers@wpafb.af.mil

Mr. Niklas Wallin
Department of Man System Interaction
Swedish Defence Research Institute, FOI
Gullfossgatan 6
164 90 Stockholm
SWEDEN
Facsimile: +46 8 55503700
Email: niklas.wallin@foi.se

Acknowledgement

Several issues presented in this report have been discussed at the HFM-143 Specialists' Meeting on "Human Behavior Representation in Constructive Modeling", prior to finishing the report and recommendations. The HFM-128 panel is grateful for the contributions received from the participants of the HFM-143 meeting.



Human Behavior Representation in Constructive Simulation

(RTO-TR-HFM-128)

Executive Summary

The Issue of this Study

In current military operational models, the human aspect is still often represented in a mechanistic way, bearing little resemblance to observations, as if all humans always act the same way in a situation much as a machine would. In reality, human behavior is not deterministic. Without proper representation of behavior, and the reasons behind the behavior, the validity of the model may be seriously flawed, making its performance and predictions questionable.

Scope

Human Factors (HF) that are relevant to operations have been identified in previous NATO reports and in books on human behavior representation with weaknesses or shortfalls pointed out in existing military simulations. In general, the knowledge of HF is more advanced than the implementation in these models. The difficulty that model developers face is how to incorporate HF knowledge without creating models of such complexity that they become unusable. This report provides a number of suggestions to limit the complexity, while using solid HF knowledge.

Considerations and Recommendations

A key recommendation is to represent human states as explicit variables that affect performance in a transparent way. Human states can be measured and provide a basis for expressing the stresses experienced by personnel that are important to military outcomes (exhaustion, thermal load, mental workload, etc.). It is time for the operational military modeling community to begin incorporating advances from HF and cognitive science into Human Behavior Representations (HBRs) for constructive simulation.

Special considerations are necessary when aggregated units are used as entities in a constructive simulation. Teams and larger units have additional HF properties that do not exist at the individual level. It is recommended that the simulation be scaled by aggregating elements where necessary and incorporating the associated HF appropriate to that aggregation.

Weaknesses in HF knowledge are revealed by the demands of the latest international operations. The effects of soldier behavior on the attitude of the local population is receiving increasing interest in the field, but the scientific knowledge in this field is still fragmented and has not reached a useful level of modeling.

Effects based operations have a profound impact on the way an operational problem is solved and consequently also on the requirements placed on simulation models. Modeling EBOperations requires increased representation of cognition, in which coordinated units must have their own representation of cognitive processes that capture assessment, judgment, and decision making. The choices allow for variable behavior. This report attempts to provide some guidance on how human behavior models can be extended to capture these effects in military modeling and simulation.

The validation of models remains a concern. Substantial funding should be allocated for this stage of the development, particularly if data must be collected. It is important that the integrated model be tested with

use cases to ensure validity, and we strongly recommend describing the validation method used when presenting the simulation results so that the audience can make an informed judgment about the suitability of the conclusions.

Military/NATO Significance of the Study

This document attempts to provide concrete guidance regarding integrated HF modeling. These considerations are synthesized in an overall scheme and a 19-step process to guide practitioners and analysts through an HF reinforced study case, called good practice. We have chosen to avoid the arrogance of referring to our recommended process as “best practice,” but we hope the reader will find it useful.

La représentation du comportement humain dans la simulation constructive

(RTO-TR-HFM-128)

Synthèse

Sujet de l'étude

Dans les modèles opérationnels militaires actuels, l'aspect humain est encore souvent représenté d'une façon mécanique, ayant peu de rapports avec les observations, comme si tous les humains réagissaient toujours de manière identique dans une situation donnée, comme le ferait une machine. Dans la réalité, le comportement humain n'est pas déterministe. Sans une représentation adéquate du comportement, et les raisons expliquant ce comportement, la validité du modèle peut être sérieusement mise en doute, rendant ses performances et ses prévisions contestables.

Champ d'application

Les facteurs humains (FH) intéressant les opérations ont été identifiés dans de précédents rapports de l'OTAN et dans des ouvrages traitant de la représentation du comportement humain, qui mettent en exergue les faiblesses ou les lacunes des simulations militaires existantes. En général, les connaissances en matière de FH sont plus avancées que leur mise en œuvre dans ces modèles. La difficulté rencontrée par les développeurs de modèles concerne la manière d'incorporer les connaissances sur les FH sans pour autant créer des modèles tellement complexes qu'ils en deviendraient inutilisables. Le présent rapport apporte un certain nombre de suggestions en vue de limiter la complexité, tout en intégrant de solides connaissances en matière de FH.

Considérations et recommandations

L'une des recommandations clés est de représenter les états humains comme des variables explicites affectant les performances de manière transparente. Les états humains peuvent être mesurés et fournir une base permettant d'exprimer les contraintes subies par le personnel qui ont une importance pour les résultats des opérations militaires (épuisement, charge thermique, charge de travail mental, etc.). Il est temps que la communauté de la modélisation militaire opérationnelle commence à intégrer les progrès des FH et de la science cognitive dans les Représentations du comportement humain (RCH) de la simulation constructive.

Des considérations particulières sont nécessaires lorsque des unités agrégées sont utilisées comme entités lors d'une simulation constructive. Les équipes et les unités plus importantes doivent compter avec des FH supplémentaires qui n'existent pas au niveau individuel. Il est recommandé d'ajuster la simulation en agrégeant des éléments lorsque c'est nécessaire et en incorporant les FH associés adaptés à cette agrégation.

Les exigences des dernières opérations internationales ont révélé les faiblesses des connaissances en matière de FH. Les effets du comportement des soldats sur l'attitude de la population locale rencontrent un intérêt grandissant, mais les connaissances scientifiques en ce domaine sont toujours fragmentées et n'ont pas encore atteint un niveau de modélisation utile.

Les opérations basées sur les effets influencent fortement la manière dont un problème opérationnel est résolu et, en conséquence, également les exigences relatives aux modèles de simulation. La modélisation des Opérations basées sur les effets nécessite une meilleure représentation de la cognition, dans laquelle les

unités coordonnées doivent avoir leur propre représentation des processus cognitifs, englobant l'évaluation, le jugement et la prise de décisions. Les choix tiennent compte de comportements variables. Le présent rapport tente de fournir des instructions quant à la manière dont les modèles de comportement humain peuvent être élargis en vue d'intégrer ces effets dans la simulation et la modélisation militaires.

La validation des modèles demeure un problème. Des ressources substantielles doivent être allouées à cette étape du développement, en particulier s'il s'avère nécessaire de recueillir des données. Il est important que le modèle intégré soit testé dans des cas d'usage afin d'en garantir la validité, et nous recommandons vivement de décrire la méthode de validation utilisée lors de la présentation des résultats de la simulation, afin que l'auditoire puisse porter un jugement informé sur la pertinence des conclusions.

Importance de l'étude pour l'OTAN/l'Armée

Le présent document tente de fournir des directives concrètes quant à la modélisation intégrée des FH. Ces considérations sont synthétisées dans un projet général et un processus en 19 étapes visant à guider les praticiens et les analystes au travers d'une étude de cas de FH renforcée, appelée bonne pratique. Nous avons choisi d'éviter l'arrogance du terme de « meilleure pratique » pour nous référer à notre processus recommandé ; nous espérons néanmoins que le lecteur le trouvera utile.

Part One – Human Behavior in Modeling

PART ONE – HUMAN BEHAVIOR IN MODELING



Chapter 1 – THE SCOPE OF THIS REPORT

1.1 THE GOAL OF HFM-128 ON REPRESENTATION OF HUMAN BEHAVIOR IN CONSTRUCTIVE MODELING

The human has always been recognized as a key factor in military operations. Numerous examples illustrate how the actions of an individual or a small team turned defeat into victory or how difficult it is to sustain operations in extreme climates. One would expect in this era of performance measurements, modeling and simulation (M&S), and mission rehearsal, that human factors (HF) would play a significant role when representing human behavior and its variation in military simulations. One might expect operator models based on human science knowledge, yet that expectation has not been fulfilled. Most simulations handle military units as if they are robots, carefully representing only mechanical qualities. Recently, combat models and training simulations have begun to include some HF in computer generated forces (CGFs) and synthetic teammates where they should have human like behavior, but often these have fallen short of what is truly required. This can actually result in a poorer experience for the trainees, for example, than no human-like behavior at all.

This modeling shortfall is not caused by a lack of HF knowledge. The human has been studied in great detail and many aspects have been represented in models, for visual detection, for thermal load, for energy cost, and for many other factors. More recently, aspects of human cognition have been modeled, including decision making, sense making, error and more. Current simulations might suggest that it is difficult to include HF in operational models even though such knowledge does exist. A generic approach for including human behavior in military simulations is indeed missing and a method is required to identify the requirements for HF models with a link to operational performance in military M&S.

The goal of this report is to demonstrate how HF models could interface with operational models by enumerating the basic, existing models and how they fit into a generic framework. We are seeking a method that enables practitioners in operational modeling to apply HF knowledge to produce more plausible behavior in their simulations. The formal terms of reference for HFM-128 RTG are as follows: “The objective of the technical team is to explore the potential of knowledge on human behavior and performance to expand simulation of military operations with simulated individual or aggregate human entities and improve their validity”.

1.2 TARGET AUDIENCE

Although this paper may be interesting reading for HF specialists, the target audience is intended to include analysts, scientists, and engineers in the field of military operations modeling. This paper guides analysts through a number of the HF that we feel could be important and how they might be implemented in constructive simulations. It is our hope that once armed with this knowledge, military operations modelers will be better able to choose the analysis environment they want to set up and the important HF they should consider.

1.3 MAIN DELIVERABLE

This report is this study group’s main deliverable, with its primary practical application being a process for considering and including key HF variables in constructive simulations. Some of the main HF areas are discussed, explaining why they are used and identifying formal models in the literature.

THE SCOPE OF THIS REPORT

We will focus on a structure for the balanced inclusion of HF in operational models that are more suited for these evolving military roles. We do not propose a standard methodology for including HF in military simulation, rather an approach that might be considered good practice that may someday lead to a standard. We do not attempt to be prescriptive, well aware of the different needs of different studies. The HF included in a study must be determined by the user according to the purpose of the application, and hopefully, this report will help inform these users in their model design.

1.4 A BRIEF OUTLINE

Part One deals with brief explanations of new trends and issues in human modeling. A practical approach is taken, illustrated by examples, and discusses the set up of models in current military applications, focusing on HF, architectures and effects based operations. An example scenario runs through the report, illustrating applications.

Part Two provides details about several HF that are relevant to modeling humans in constructive simulation. The chapters in Part Two are organized to match the information need of the reader, addressing perception, cognition, performance moderators, motor tasks and the cause of error.

In **Part Three**, we summarize what we learned from this study for the promotion of efficient use of modeling resources. Part Three wraps up with future challenges, recommendations to NATO and references cited in this document.

1.5 LIMITS OF THIS STUDY

Not all aspects of HF are addressed in this report. The effects of drugs on performance and the effects of biological and chemical agents on physiological status are not discussed, nor are the effects of injuries, emotion, or motivation on performance. The study is limited to the HF that are likely to have formal representations (computational models) and that span the normal range of human physical and mental functioning. This study focuses primarily at the individual human level, although there is some coverage of the influence of teams, crowds and socio-cultural influences on behavior.

This study attempts to build on several technical papers that have been published on human behavior representation (HBR). The authoritative book of Pew and Mavor (1998) collected and summarized a large amount of information that is still valid. Since publication, some new developments have emerged, particularly in the cognitive and emotional domains. Changes in operational activities have occurred over the past 10 to 20 years, with effects based operations and network enabled capabilities as prime examples. These concepts directly affect even the lowest echelons in the military hierarchy.

1.6 SUMMARIES OF PRECEDING REPORTS

The team led by Pew and Mavor (1998) set the standard on analyzing the future of modeling human behavior in their advice to the (U.S.) Defense Modeling and Simulation Office. Their report is largely dedicated to cognition in military operations, focusing on cognitive functions, planning, and behavior at the unit level, although they recognize the importance of representing movement, detection, and identification as basic military behaviors. Behavior must be interpretable to the observer as a choice from a range of behaviors, representing response to threats. Behavior is modified by individual and organizational moderators. An integrative architecture is needed to accommodate these elements in a model.

It is doubtful that a single integrative model could be developed to meet all needs, but a common philosophy may help to make models more valid. In the short term, real-world data must be collected, identifying procedures and measures of performance that are poorly developed. In the intermediate term, task analyses must be expanded, supporting simulation performance metrics. Published component models must be validated by independent sources in specific domains. In the long term, behavioral research is needed to reinforce theory. The Pew and Mavor team analyzed a number of major cognitive models (ACT-R, COGNET, HOS, SAINT, SOAR and others) for their architecture and applicability, concluding that none of them is precisely what is needed for military simulations, although together, they cover many of application requirements. Validation at the system architecture level is a significant challenge. Partial or component models may be well validated, but there is no consensus in the field regarding the right methodology for validating entire cognitive architectures. Pew and Mavor discuss the need for attention and multitasking, memory and learning, human decision making and planning in detail.

The focus of the NATO RTO SAS-017 report (Dompke, 2001) is on the long term forecast of technology areas that will support HBR. HBR is seen as an affordable alternative to human role players for instruction and training, defense planning, operational support and acquisition, particularly for uncertain future environments. The use of HBR is considered a potential winner between human agents (limiting flexibility) and over-simplistic automated agents. It is observed that models for collective behavior and organizational behavior are underdeveloped. Also, validity and interoperability need attention. The technology areas to be developed are adaptive-intelligent coaching, goal oriented information processing, automated explanation of behavior, distributed constructive modeling, and the reuse of knowledge in composable models.

The span of entities acting in various military simulation applications are individuals, teams, groups of various sizes, organizations, and crowds. Behavior is interpreted as a result of mental functions and the relationship between these functions and the entities' traits and states, including abstract performance moderators for groups. For the aggregation of individuals to group behavior, two lines of development are distinguished:

- 1) Aggregation of individuals by inter-relationships (by task networking or multiple agents); and
- 2) Assignment of aggregate properties to the group (at its infancy).

Identified scientific weaknesses at the time were team performance moderators, communication, backup behavior, Measures of Performance and taxonomy of grouping. The report discusses model properties and structures, identifying numerous shortfalls of current approaches such as:

- 1) Knowledge acquisition is expensive and often undocumented.
- 2) Current efforts are granularity locked, but there is a need to be able to change the granularity of the model during development if not actually in use.
- 3) Composability of modeling effort is lacking and necessary in relation to model development and revision.
- 4) Scalability of models is lacking.
- 5) Lack of required behavior moderators in current models produces models that are brittle, non-adaptive, and predictable.
- 6) Current models require extensive human manipulation and are typically difficult to control adequately during simulations.
- 7) No intent determination is used within current operational models.

THE SCOPE OF THIS REPORT

- 8) No integrated cognitive processing architecture (situation assessment, decision making, and planning) is present.
- 9) Learning is limited in current models, if addressed at all.
- 10) No explanation capability is present to allow understanding of current model decision processes.
- 11) Verification, validation, and accreditation are not currently addressed.

Some of these topics are addressed in the current report, recognizing that many shortfalls are related to shaping the modeling environment to accommodate operational problems.

The NATO SAS-027 report (NATO, 2004) concerns C2 assessment in general, of which modeling is one method. Regardless of method, a consistent set of measures of merit (MoM) is needed, each of which is observable and quantifiable. C2 in high intensity combat is regarded as easier to model than in operations other than war (OOTW), where uncertainty and complexity are greater. A single monolithic model of C2 systems is considered less feasible than a federation of entity models. Consequently, there is a need for common MoM and common parameters to link specialized models. The inclusion of human behavior in C2 modeling is considered essential and some guidelines are as follows:

- 1) Provide a repertoire of adversary intents to recognize.
- 2) Represent headquarters explicitly.
- 3) Represent operational scenarios so that deception, shock and surprise can be considered.
- 4) Treat information as a commodity that can be transmitted through communication systems.
- 5) Make information processing explicit to allow for aggregation and fusion.

The need is stated, but the ways this can be done are left to specialist readers. Among the strengths of current C2 modeling mentioned are the general state of understanding in high intensity combat (not in OOTW), evolutionary model development, progress in linking and federating models, progress in describing morale, fatigue, and training proficiency, interface protocols and decision support tools. Among the challenges are mastering the complexity and confusion of OOTW, mixing human and constructive entities in the simulation, representing human decision making processes in the constructive entities so that stress, fatigue, shock and morale have their impact, supplying data to support these effects, establishing similarity in modeling the capabilities and limitations of blue and red parties, increasing the validity of the models, and performing cost-effective sensitivity analyses.

Ritter, Shadbolt, Elliman, Young, Gobet, and Baxter (2002) extend the review of Pew and Mavor, examining work done after publication of the NRC report as well as broadening the perspective to include HBR architectures created outside the U.S. Ritter et al. review a number of features deemed important to consider in HBR architectures, such as emotion and error, in addition to the standard, perception and rational cognition concepts so that more plausible performance is observed from the simulation. The authors identify and discuss a number of architectures (Ritter et al., 2002, summarized in Table 5.1, pp. 45-52), particularly those that incorporate models of perception and affective features. They note that, in most cases, the models developed in these architectures had not been verified against human data yet they also note that with the proliferation of architectures, lessons are being learned and it is becoming increasingly easy to create plausible and useful simulations. Ritter et al. (2002) finish by identifying some activities that begin to address a number of the limitations in HBR to be suitable for military M&S, from architecture elements through modeling philosophy and component model requirements, notably perception, emotion and error.

The Gluck and Pew (2005) book documents the human data, models and lessons learned from the U.S. Air Force Research Laboratory's Agent-based Modeling and Behavior Representation (AMBR) project. This research effort involved four modeling teams using different architecture-based modeling systems (ACT-R, COGNET/iGEN, DCOG, and EASE), an independent supervisory team (BBN Technologies), and several related architecture-based model development, evaluation, and validation efforts over several years. The processes and performance levels of computational cognitive models are compared to each other and to human operators performing equivalent tasks. The tasks are variations on a simplified en route air traffic control hand-off task and emphasize multi-tasking, workload, and category learning. The descriptions of the models go well beyond computer code that is included with the book, and detailed accounts of both the simulation architectures and the models are provided. Quantitative analyses comparing the performance data from the models and human subjects are included.

The book is divided into three sections. The first section of the book is background material. It starts with an overview of the effort, followed by a description of the method and results from the human experiments, the rationale for the choice of tasks, a detailed description of the task software and its dynamics, the human operator requirements and how the software was set up to allow seamless introduction of either a human operator or a computational process model that simulates the human operator, and the way in which the models were connected into the simulation.

The second section of the book includes a separate chapter for each of the models developed in response to the modeling challenges. The authors of these chapters were given a detailed structure to follow to assure that the chapters would cover similar topics and so that the reader would find it easier to follow the model descriptions and modeling results. At the end of each of these chapters, the authors answered a set of summary questions about their models.

The last third of the book presents a qualitatively oriented discussion of the practical and scientific considerations that arise in the course of attempting this kind of model development and validation effort. It starts with a discussion of how the architectures and models were similar and different and how they performed the target tasks as compared with human data. The book ends with three chapters of general interest to those working in the area of human behavior representation. Love (2005, Chapter 9), compares the AMBR models of category learning to other models of category learning in the contemporary psychological literature and discusses the arguments and requirements for architected versus non-architected modeling approaches. Campbell and Bolton (2005, Chapter 10) cover a variety of important issues associated with the validation of computational process models, emphasizing the distinction between construct validity and application validity. The book ends with reflections on the results of the project and proposes a research agenda to carry the field of human behavior representation forward. The final chapter calls for programmatic and budgetary emphases on formative and summative evaluations of validity and robustness, which unfortunately are more the exception than the rule in current practice.

1.7 A SCENARIO DEMONSTRATING THE USE OF HBR

Throughout this paper, you will find excerpts of a fictional scenario tied to the application of HBR to help you understand what is and what is not important for your own study. The main scenario is presented here.

Operational Scenario

The year is 2020. The NATO countries are preparing for a humanitarian effort in a mountainous country near the equator. They will be working to combat the effects of drought, famine, and poverty

THE SCOPE OF THIS REPORT

in the midst of political unrest and frequent skirmishes among the local factions. Colonel Henry Schmidt is leading this multinational joint coalition. As final preparations are made to move into the area, he reflects on the confidence he has in the equipment the troops will be using, the training they have received, and the mission rehearsals they have just completed – and he knows that confidence is well-founded.

What he may not know is the extent to which M&S – in particular human HBR in M&S – served as the foundation of effectiveness throughout the lifecycle and fielding of the systems and equipment he was using, as well as the operational analyses used to plan for the mission and train his command.

New Equipment Concept Exploration. Virtually all Soldier equipment begins as a concept and this is where the HBR M&S are first employed.

Colonel Schmidt surveys his situational update display. He can determine not only the precise locations of the various NATO units under his command but also has the information to assess the readiness of the various units. The readiness level includes Soldier factors such as cognitive preparedness, how recently training was completed, and how rested the Soldiers are. At the same time, just outside the capital city, Lieutenant Anderson readies his platoon to deliver truckloads food and medicine supplies to the medical center located near the city center. He has received the reconnaissance reports from a swarm of robotic assets giving the all-clear. The information is displayed via a light weight, flexible head-mounted display, the latest in technology. Further, he knows that the information has been sent simultaneously to the other platoons in his company and has been fully coordinated with the other supporting NATO forces.

When Colonel Schmidt and Lieutenant Anderson's equipment was first being explored in the concept stage, HBR modelers performed a series of M&S-based activities side by side with the military experts and designers responsible for developing new equipment concepts intended for use by a future multi-national coalition. Task-level models representing human performance at the granularity of minutes were used to predict which mission functions would benefit from the introduction of new technology capabilities. For example, using task data for a range of Soldier tasks – mission planning, squad navigation, and casualty care – the usefulness of individual Soldier global positioning systems (GPS) accuracies of 0.5 versus 10 meters was estimated using metrics such as time to perform a mission, message load, and probability of mission success. These studies were largely based on SOPs and required few human data beyond typical task completion times and error rates. HBR models of military personnel – at a granularity of less than a few seconds – in various occupations were used to determine the “cognitive fit” of display concepts such as helmet-integrated on-demand information displays. These studies required support from the human sciences, introducing psychologically plausible models of perception, decision making, and planning that were used to test concepts in a variety of scenarios. These HBR operators were further stressed by the introduction into the simulation of bright sunlight and hot environments, requiring additional moderator models of visual sensation and thermal physiology. On the basis of these early HBR performance predictions, certain technologies were ruled out and for the technologies that showed promise, specific design recommendations were made to improve operational performance while minimizing operator load.

Design Evaluation. As equipment and system concepts mature, the first prototypes emerge as software mock-ups. HBRs are able to interact with software mock-ups for a first look at Soldier-system performance.

Lieutenant Anderson appreciates that the information from his helmet-mounted display does not interfere with his ability to perceive and act on information coming directly from his immediately surrounding environment. This is especially the case as the platoon travels through one particular

neighborhood en route to the medical center. Although the robotic assets had given the all-clear, Lieutenant Anderson has a direct visual sighting of a crowd that has gathered and he is not sure of their intent. This is an area where insurgents have been active and he has witnessed attacks against other patrols. The crowd seems very excited, although from training and experience, he recognizes that these behaviors are often associated with positive events in this culture and the atmosphere seems different from that when attacks were imminent. He queries the network and his display re-activates instantaneously to show the message confirming that the crowd is not a threat, that they are gathered for the opening of a new school. The display fades and he decides on a detour so as to avoid any appearance of aggression, leading the platoon on toward the medical center by an alternate route that had been identified for such contingencies.

HF engineers and psychologists working for the governments of several NATO countries and for the contractors hoping to win the final production contract use integrated perceptual-cognitive-motor models to assess both the cognitive and physical fit of proposed systems. Some HBRs were used in constructive simulations to assess higher level, essential design aspects in selected scenarios, testing the predicted performance sensitivity to known human differences, simulating representative missions conducted by hundreds or even thousands of different personnel. Other HBRs fill the role of missing teammates or adversaries, freeing personnel from role-playing duties in the simulations. These HBRs interact with military experts who assess prototypes in mixed-reality simulations, using Natural Language Processing and speech production communications, exploring team performance during key aspects of the military's roles. Effective HBRs in these applications will require more detailed models of human cognition to provide adequate sensitivity to scenario variables and to produce sufficiently rich assessments that the military experts can relate the simulation to real-world experiences or plausible events. Effects of stressors on performance are investigated in depth such as noise on communication and vibration on display legibility or operator fatigue on cognitive tasks. In both the standalone and the integrated modes, a standardized and validated common HBR framework was used by government and industry alike to evaluate competing designs on common, objective metrics that reflect human abilities and limitations. This evaluation work was critically important to assure that the value added of the technology worked seamlessly in dynamic operational environments.

Safety. Making predictions about possible Soldier-system errors early in the conceptual or design phase allows designers to improve designs before production or to incorporate fail-safe and contingency features such as automated warnings, reducing the likelihood that such errors will lead to accidents in the field. HBRs can be used to examine the kind of rare events that lead to catastrophic failures through repeated simulations that would be impracticable for standard Operational Test and Evaluation (OT&E).

The delivery of the supplies to the medical center is well underway when Colonel Schmidt receives an urgent message about a possible hostage situation in a remote section of the city. He immediately realizes that this implies a number of steps to prepare for his new tasking. While he is in the midst of checking unit status reports to support replanning for the rescue mission, he is interrupted again with a message about a power outage in a remote neighborhood. This may indicate insurgent activity or be just another infrastructure glitch. Command staff reports no other indicators of hostile activity on the situation displays, so he decides on the latter for now and returns to checking the unit status reports. The system prompts him regarding exactly what information he has reviewed so far and highlights important aspects of the remaining information that should be reviewed to ensure getting the total picture. Colonel Schmidt is reminded of several important details he had forgotten during the interruption but quickly refocuses on the problem without missing information or losing time.

Colonel Schmidt's transition back to his main task after the interruption was made seamlessly due to HBR efforts. The HBRs used previously in both the concept development and design evaluation stages were

THE SCOPE OF THIS REPORT

tailored and were run hundreds, even thousands, of times, in order to understand the effects of specific types of errors. What if a commander omitted an action, for example, forgetting to check units both for fuel supplies and for cognitive preparedness? What if information is presented ambiguously and an alternative interpretation is plausible that differs from the intended message? The task models were modified to explore consequences of such errors, employing advanced models of human reasoning to test for misinterpretations and human biases rather than relying on artificial intelligence (AI) approaches that attempt to reach an optimal conclusion. Then, those HBRs were used to identify possible error correction tasks and to inform the design of decision aids, alerts, and warnings. The same modeling approach was used to explore the safety consequences for the full range of Soldier-system errors, including errors of commission, over- and under-reliance on system automation, and even higher echelon doctrine or policies that can influence overall system safety.

Performance under the Effects of Stressors. Despite various cultural and military differences found in multi-national coalitions, the effects of stressors – e.g., heat, noise, vibration, fatigue, and sleeplessness – are universal and prevalent. Such stressors induce changes in physiological states and, consequently, performance effects that must be accounted for when developing a system and when training to use it.

The delivery of food and medicine supplies to the medical center is accomplished on time and without incident. Now Colonel Schmidt devotes his undivided attention to the hostage rescue. An insurgent faction led by a key figure is holding local, friendly citizens and United Nations inspectors in an abandoned warehouse. Colonel Schmidt is concerned about his attention to detail since he has now been without sleep for 20 hours. The situation update display, being tuned to recognize possible performance decrements, implements an increased level of decision aiding automation. Out in the city, Lieutenant Anderson, having completed the supply delivery, is now near the industrial district. He receives an order to support the hostage rescue. The platoon has been exposed to the effects of extreme vehicle vibration and now is suffering from the heat of day. In order to ensure their effectiveness on this next mission, the platoon takes on additional water and specially packaged nutrition supplements; their information displays are reconfigured automatically for better readability, and their micro-cooling is enabled. Colonel Schmidt is alerted by real time, operational readiness assessments that estimate the levels of fatigue and thermal strain due to weather during the resupply mission. Coupled with the additional equipment required for the rescue mission, the readiness assessment model predicts the platoon's abilities to move quickly and adjust to the uncertain situation are reduced. Should the situation degrade, immediate action could save lives, prompting Colonel Schmidt to deploy a fresh, fast response Special Ops Team to lead the rescue assigning Lieutenant Anderson's platoon to a supporting role, securing the perimeter from strong points.

A rich database of performance under the effects of a full complement of stressor types was invoked to modify and shape baseline HBRs. Thus, without subjecting soldiers to hardship unnecessarily and without the large cost of conducting field evaluations, HBR modelers were able to evaluate the effects of stressors – individual stressors, stressors in combination, and extremes of the stressor environment – on performance, using models such as thermal physiological response to working in hot environments coupled with vigilance and decision making models that are moderated by body temperature and hydration status. An additional factor, individual Soldier traits, was also invoked during the HBR M&S process to capture the variability of performance due to differences in soldier resiliency, coping mechanisms, and tolerance for uncertainty. As important, possible mitigations, such as design changes, personnel or manning fixes, and training enhancements were evaluated so that a full range of design guidance and system recommendations could be made that were based on quantitative projections.

Predicting Operational Outcomes. While it is critical to evaluate specific Soldier-system interactions via the use of HBR, the larger operational picture must be assessed also, providing in-theatre personnel with the latest predictions of what to expect. It may be unrealistic to predict all possible eventualities, but exercising a breadth of scenarios with non-standard actions will reduce the risk of applying outdated, stereotypical responses to novel situations. In order to update Cold War era war games and fully address effects-based operations, HBR must be included to go beyond the stereotypical SOP threats of the last war.

Colonel Schmidt is thinking that, despite all his experience and training, today's events have presented unexpected challenges. He finds himself continually assessing options and considering risks when directing his command. On the other side of the city, Lieutenant Anderson, is thinking much the same thing, that today's events are a new test for his leadership and military skills, however, he finds that his focus is on the immediate environment and his decisions must be made based on instinct with little time to consider options.

Although neither Colonel Schmidt's precise scenario nor Lieutenant Anderson's exact mission had been modeled, a series of M&S exercises predicting operational outcomes had been conducted. HBR modelers worked with large-scale M&S developers to make sure that the appropriate granularity was incorporated. They were able to draw from a database of various levels of HBRs, from individual soldiers and small teams that could act with a large degree of autonomy in simulations, to brigade- and battalion-sized forces that required command intervention by personnel. Likewise, they were able to select from a database of standard tasks including navigation and shooting, to reconnaissance and higher level decision making, all drawing on component models of human perception and cognition to generate plausible actions given uncertainties in the simulation rather than idealistic responses that arise from being aware of everything at every moment. In this way, HBR helped to inform the trade-offs that must occur in the application of effects-based tactics and force employment against the higher level metric of operational effectiveness.

Training Soldiers with Simulated Teammates. It is said that all but war is simulation. Thus, soldiers' training must be sufficiently realistic to achieve the curriculum objectives, including interacting with other teams, units, and forces in a natural manner, avoiding contrived or simplistic actions of other participants in the scenario. Representation of other entities in training simulations is another application of HBRs that can be developed to varying levels of resolution as required.

Even before arriving in country, Colonel Schmidt and Lieutenant Anderson had trained together even though their home units were in different NATO countries. Their operating styles as NATO partners and their shared understanding about the common mission – including both support and combat missions – grew out of distributed team training with simulations. Their experiences included training with simulated HBR teammates from their own country and from the various NATO countries, all in the context of a common environment that included enemy and neutral parties, also simulated by HBRs. As this day unfolds, they appreciate the rich set of teammates, simulated and real, with whom they had trained.

HBR modelers provided representations of teammates at all echelons, from commanders to squad members, customized according to the training objective. Simulation of coalition partners from other countries included modifications to HBR parameters such as cultural differences that lead to a likelihood to make certain types of decisions and variations in tactics. In some cases, it even included changes to fundamental aspects of cognition such as the factors affecting recall from memory and what aspects are considered important in situational assessment and decision making, all leading to different behaviors that broaden the team's perspective. HBR for enemy and neutral parties has been a long term challenge for the HBR community using simple rule-sets, but incorporation of more flexible models that learn and reason, leading to behaviors that the

THE SCOPE OF THIS REPORT

analyst had never considered, provides broader experiences and richer simulations for training, including actions that seem out-of-the-box compared to conventional military tactics and doctrine.

Mission Rehearsal. With a firm foundation of HBR used in system concept exploration, design and training, the stage is set for use of HBR in mission rehearsal.

As the hostage rescue unfolded, the mission rehearsal from the early hours of the morning came in handy. Colonel Schmidt had prepared Lieutenant Anderson's platoon for the planned mission, the delivery of supplies, through readily available simulations that provided for surprise and replanning. This exercise prepared the Soldiers for the rescue mission, even though it was not explicitly gamed, by broadening their expectations. The HBRs with which they had interacted had provided cues as to what to expect at multiple levels, from interaction with a crowd to interaction with specific individuals. Armed with this realistic experience, along with good equipment and information, Colonel Schmidt sees his plan succeed and Lieutenant Anderson leads his platoon into the abandoned warehouse to coordinate with the Special Ops Team to rescue the hostages without casualties. Colonel Schmidt makes a mental note to commend Lieutenant Anderson's platoon on their professional response under difficult circumstances.

In the end, the rescue mission was successful, due in no small way, to the use of HBR in all stages of system acquisition and force employment in simulated operational settings.

Chapter 2 – HUMAN FACTORS, ARTIFICIAL INTELLIGENCE, AND COGNITIVE ARCHITECTURES

The year is 2020. The NATO countries are preparing for a humanitarian effort in a mountainous country near the equator. They will be working to combat the effects of drought, famine, and poverty in the midst of political unrest and frequent skirmishes among the local factions... A variety of HBRs played a part in preparing Colonel Schmidt and the multinational coalition forces for their humanitarian mission this hot day in 2020. HBR was used years before as a part of concept exploration and design evaluation for equipment being used. More recently, HBR was used in training and operational analysis, and just this morning, HBR in synthetic teammates was used as a part of the mission rehearsal. Decades of basic research in psychology, HF, and AI and applied development activities all contributed to the development of HBRs for use in M&S. It is not important that Colonel Schmidt know the details; however, it is vital for mission success that the details were addressed via rich, robust, valid, and militarily -relevant HBRs.

2.1 HUMAN FACTORS, HUMAN PERFORMANCE, AND HUMAN BEHAVIOR

It will be useful to clarify how the terms human factors (HF), human performance (HP), and human behavior (HB) are used in this document. HF is the collective name for the full range of intrinsic human characteristics and capabilities, determined through scientific study to be important for understanding and predicting performance of a human-machine system. Basic HF research can focus on thresholds or limits to what humans can perceive, learn, or accomplish generally for specific tasks in context. HP is related to the execution of a specific task under specific circumstances. Performance does not bear on relevancy of the task or whether a person will choose to perform a task under certain circumstances; rather, it refers to a measure or measures of the outcome achieved during task execution. The impact of the performance is subject to the operational requirements and performance may be compared with criteria that assess the acceptability of the performance for a given context. HB, on the other hand, refers to the human actions that ultimately are observed in the world; that is, HB describes what people actually do.

HBR encompasses elements of all three of these definitions in the creation of a computational model to predict human activity. HBR is often broader than HP in that it often attempts to integrate numerous factors into a performance relationship or function rather than considering a single set of conditions. HBR is broader than just HB since it also encompasses the underlying reasons for those actions. An important element of HBR is the rules or principles that direct the observed behavior and the degree to which these rules are based on valid, scientific knowledge from the HF field. Fortunately, there is a wealth of basic HF and psychology research to draw upon, although the results are not always in the form needed for modeling.

2.2 HF FRAMEWORK

In the field, goals or objectives are pursued in an operational environment (in the battle lab, a synthetic environment) and the actions taken evolve with the situation. Plans are generated and orders issued to be carried out by the personnel involved. Each entity has activities to perform and the outcome depends on their skills, circumstances within the environment, and the actions of the opponent.

In this report, the interaction between a human agent and the environment is assumed to be moderated by the agent's state (Figure 2-1). The execution of tasks in the given environment creates stress in the humans.

Stressors (*properties of the operational environment or tasking that disturb state variables from their normal values*) dynamically change the agent's state while traits (*characteristics of the entity that remain constant over the operation*) affect the agent's state or performance in a static way. The activities are performed better or worse, depending on the state of the entity. States and moderators are discussed further in Chapter 6. As an example, the activity of marching with pack (stressor) will exhaust the entity, but physical training (trait) moderates exhaustion. Exhaustion is then a state variable (changing over time) that degrades the marching performance. This variation in behavior is formally described in Performance Shaping Functions. Slowing down or stopping altogether are two behaviors, but the list may be long. For instance, threat causes another behavior and decisions on the course of action (COA) yet other. The collection of individual behaviors and the interaction between individuals result in higher level (operational) effects. Task networking is a common tool to scale up to collective achievements. Tasks are treated in detail in Chapter 8.

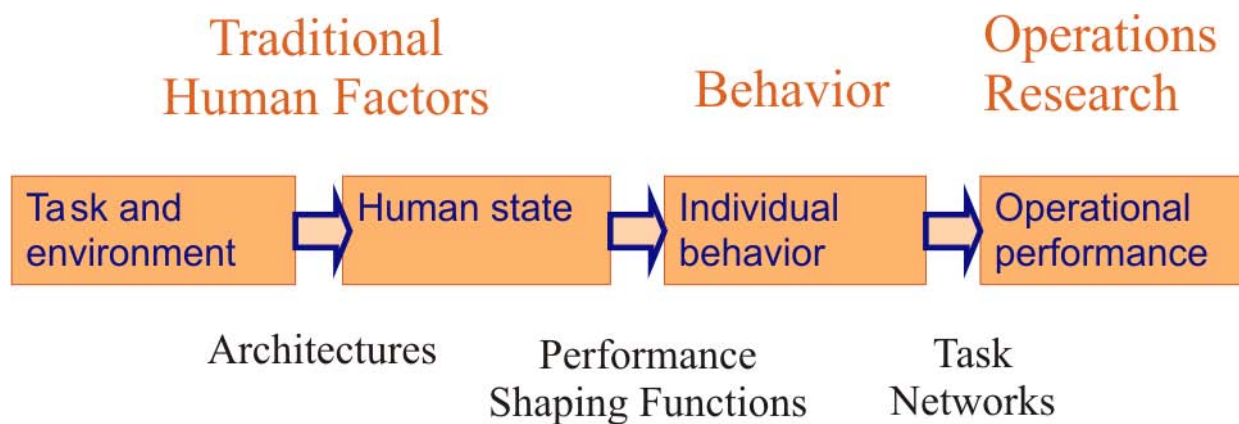


Figure 2-1: An Approach to Modeling Operator Performance Incorporating HF.

The relationship between task and operational performance is established via two unique model features: the human state and individual behavior generation. Human states are predicted with models of human responses to stressors integrated with capabilities in modeling architectures. Performance Shaping Functions incorporate operator traits to predict the effect of operator state on behavior. Task networking is used to model the nominal actions of entities in the simulation. The final outcome is operational performance. The various research fields involved in the various stages are shown above the relevant functional blocks.

Larger entities (teams, squads) may have additional characteristics relating to the collective. Team cohesion and team diversity do not exist at the individual level, yet they will be exhibited by specific behaviors. Individual differences are largely ignored in these larger entities unless they are represented as differences in the aggregate factor values. Aggregation of entities and scaling of operations is a relevant issue, which is far from resolved (Chapter 11).

2.3 REPRESENTATION OF HUMAN INFORMATION PROCESSING

Most information processing models apply a variation of the stage model shown in Figure 2-2. Depending on the specific application, the working memory plays a central role in the information processing and arrangement of the functional blocks, as well as the information flow paths vary. Some models use a blackboard representation as a substitute for working and long term memory, while others make a distinction between working and long term memory. Working memory is often considered the bottleneck for information processing in these models (ACT-R) although some approaches, such as the EPIC model, assume that

perception and motor action are the limitations. Models differ in how competing processes are handled. Most assume that multiple tasks or processes may be attended to by switching attention, but the trigger to switching is a point of debate. ACT-R assumes that for each problem, production rules can be selected on a cost-benefit basis and that the production rules create a procedure for task switching. In COGNET tasks compete for attention and the priority is a dynamic process, depending on the external world and the human state. SOAR adds learning to the model by feeding the long term memory with production rules that solved a problem.

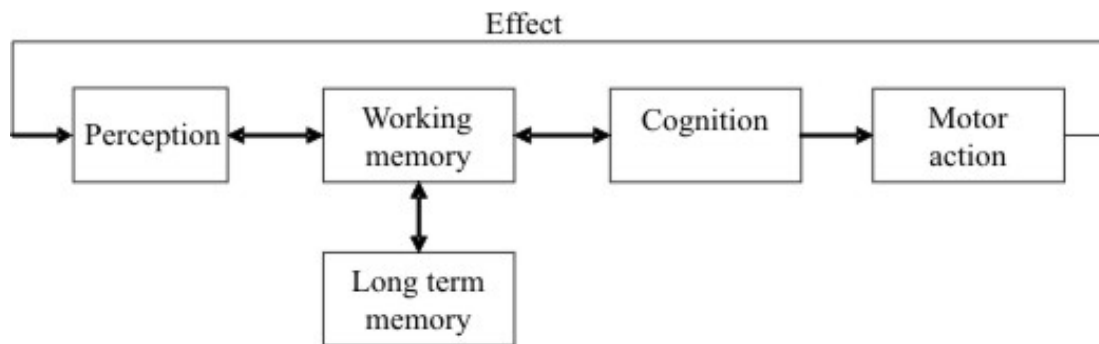


Figure 2-2: The Widely Used Modified Stage Model for Human Information Processing.

Of particular interest are militarily meaningful concepts such as situation awareness (SA) and decision making. The widely used and intuitive term SA is scientifically difficult to handle, since there is no mental substrate that compares to SA. SA capitalizes on the working and long term memories since it uses declarative knowledge (reference models stored in long term memory). It also exploits reasoning and sense making to understand information in the context of the reference model, and perceptual attention is focused to search for confirming information. Finally, reasoning is used to infer consequences of possible events, evaluating the expected outcome. In the decision making cycles, the metric for optimization is established by reasoning and by exploiting episodic memory. Obtaining SA thus is using the full range of cognitive functions and can be considered a higher order process, while cognitive research attempts to attribute steps in this process to distinct mental functions, as depicted in Figure 2-2.

Perception is presented in Figure 2-2 as an integral part of information processing. Perception goes beyond the study of human senses (sensation) and includes some processing and interpretation of the sensation. Each entity representation of a human in a model must have the competence to see and hear, to search and recognize, to identify and to understand the environment, similar to what personnel experience. The model must be limited in range of perception and be enabled by instruments (radio, night vision) similar to the constraints on a human operator.

The decision making embedded in Figure 2-2 is often regarded as a logical process although it is known that the process varies among individuals and situations. The effort invested in the decision making processes results from the interest and competence of the decision maker and the time available. More often than not, simple strategies are applied, guided by experience, to make timely decisions, sometimes at the expense of the quality of the decision.

The traditional view of decision making considers and weighs explicitly all factors, trying to evaluate options to choose the option with the highest evaluation. Although linear weighing functions fit well in some situations, the internal cognitive process is not straightforward. When such a deliberation occurs, it is an

iterative process where explicit and implicit evaluations need to match. The decision maker circles around the problem, trying to define the factors at stake and trying to find factor weights that are in accordance with intuitive evaluations.

The implicit part of decision making is important for modeling behavior realistically, even under seemingly rational circumstances. Purely cognitive and more emotional elements need to come together in the decision making mechanism. Hoyer and MacInnis (1997) document many aspects of the behavior associated with decision making including non-rational elements such as groupthink, dominance, leadership, overconfidence, tunnel vision, etc. Consumers are known to vary their decision strategy, depending on the uniqueness of a purchase and the time and knowledge available. Attitudes and emotions seem to play a significant role in many of these decisions, and there is little reason to doubt that similar effects occur in what is typically represented as purely rational thought.

Indeed, the traditional or optimal formal reasoning approach is seldom observed, particularly decisions made by experts under time pressure. Instead of considering all factors, weighting options and selecting the most beneficial or least detrimental option, expert decisions are often made based on experience, focusing on one or a small number of salient factors, as reflected in Recognition Primed Decision Making (Zsombok & Klein, 1997) or Fast-and-Frugal (Gigerenzer & Goldstein, 1996) decision making models.

2.4 REPRESENTATION OF PHYSICAL DRIVERS FOR BEHAVIOR

Physical activity is a very common element of operations and exhaustion is a limiting factor in the momentum of the operation. Military are being physically trained to sustain longer and the limits may be considerably shifted. The time to exhaustion is dependent on fitness, but also on the type and severity of the activities. To cope with such different physical activities as jumping, long distance running or wilderness survival, the human energy system draws from various metabolic processes, each with a specific time span (McArdle, Katch & Katch, 2001).

Most military activity is in the seconds to hours time frame, exploiting both anaerobic and aerobic metabolism. Predicting exhaustion requires the monitoring of blood glucose and muscle lactate by a metabolic model that may include working muscle, resting muscle (not used in the particular activity but important for recovery), blood circulation, liver function, and glucose supply by the intestine. Since fitness plays a large role in sustainability, the expression of fitness in aerobic capacity, parasitic lactate production, glycogen depletion tolerance, and lactate tolerance must be specified. Predictive formulas for walking take body weight, load, slope, terrain and speed into account (see Pandolf, Givoni & Goldman, 1977 and succeeding papers). Indeed, movements are important military tasks, but movements under threat such as crawling have not received extensive attention. More common tasks such as digging, loading, lifting, and construction may be found in work physiology text books but usually for self paced work (Astrand, Rodahl, Dahl & Stromme, 2003). Hard physical work, sustained for 8 hrs a day, is usually carried out at 450 W of metabolic power, but exhaustive work, sustained for a minute may be done at much higher power (up to 4 kW). Tracking of the physiological parameters is required to adequately capture the effects at this level.

Other physical factors can be important in include in an HBR in addition to gross motor limits and capabilities. These factors include the speed and skill of fine motor movement (e.g., typing or aiming), eye movements (e.g., scanning the environment or a computer display), motor activity associated with speech, and coordination of eye, hand, and whole body movement. Thus, depending on the way in which physical factors are represented, the HBR may need to consider physiological factors, task-goal drivers, and other concurrent task demands, and may need to be sensitive to environmental variables.

2.5 ARTIFICIAL INTELLIGENCE AND HUMAN BEHAVIOR REPRESENTATION

For someone not being a specialist in the AI or HBR field, it might be somewhat confusing to sort out how AI and HBR relate to one another. The current report is on human behavior representation, but development in the general AI field is definitely of interest for any HBR specialist since a great deal of human behavior can be and has been described in computer algorithms that fit well into operational models. Unfortunately, resulting behaviors in many models are humanlike but mechanistic, and the repertoire of behaviors is limited such that the responses are unrealistic.

In AI, problems are solved by rule sets that lead to a solution and typically attempt to find an optimal solution. That may be different from what humans do. Observed human strategies lead, more often than not, to suboptimal solutions that are considered to be acceptable solutions, at least in the short term. The question for the human behavior model in military simulation, then, is not to find the best solution, but to find the solution the human would find: the model needs to be equally wrong as the human and for similar reasons. This makes cognition in HBR a greater challenge than decision making in AI.

The HBR practitioners who worked on the systems used by Colonel Schmidt and Lieutenant Anderson were able to invoke very high fidelity HBR-based entities, very different from the early days when entities in combat simulations behaved in ways that would be unlikely to be demonstrated by humans. For example: in a past simulation, “a constructive infantry unit encounters a river and walks along the river’s bank to find a bridge. A constructive enemy force is holding the bridge but the simulated infantry stubbornly try to cross until the last man is killed. The constructive enemy continues to fire on the dead infantry unit.” Shortfalls in HBR such as the three elementary behavioral weaknesses in this vignette (the strategy to cross the bridge did not anticipate the opponent’s tactics; the decision to cross the bridge should have been aborted in light of the new information after the enemy opened fire; and the constructive enemy should have been capable of discriminating dead from live entities) have now been addressed, thus ensuring the credibility of the model outcomes.

The AI research field is considered by many to be divided into a number of subfields that specialize in a certain type of algorithms and knowledge representation formalism. Fuzzy logic, Bayesian networks, artificial neural networks (ANN), genetic algorithms, finite state machines (FSM), case based reasoning (CBR), belief-desire-intent architectures (BDI), and production rule systems are some examples that have received widespread attention. AI and HBR share in part the representation techniques, adding to the confusion. However, some techniques are more used in the one or the other.

In Russell and Norvig’s (2002) comprehensive overview of the AI field, most chapters map directly to the challenges of human behavior representation with problem-solving, knowledge representation and reasoning, planning, learning, reasoning with uncertain knowledge, communication, perception and action execution being covered. Thus the visionary goal of both AI and HBR are very similar, and indeed, AI attempts to accomplish things that people already do, although perhaps using different mechanisms.

HBR could be considered a subfield of AI, but whereas several AI subfields have focused on the development of effective algorithms for problem solving, the goal of most HBR specialists is to develop models that to some extent are exhibiting humanlike behavior and exhibit that behavior due to a reasoning process that at least partly resembles human decision making. An old and basic, although rather blunt distinction between approaches described in Russell and Norvig (2002) is the distinction between:

- a) Systems that act like humans;

- b) Systems that think like humans;
- c) Systems that act rationally; or
- d) Systems that think rationally.

Modeling goals and model applications of different HBR projects differ, but regardless of this, different AI algorithms provide the basis and valuable tools for HBR model development. Especially in projects with decomposable models, AI components can be combined with HBRs. For instance, the modeling of G-force or sleep deprivation effects in pilots might be the research topic in a project, in which AI path finding algorithms for pilot navigation are providing good-enough input to the models.

2.6 CHOOSING A MODELING ARCHITECTURE

One key issue in any human behavior representation project is the choice of modeling architecture for the implementation of models. In many instances, selection of an architecture for a project is governed by previous projects, i.e., modelers will choose an architecture they are familiar with and that they feel will be “fit for purpose” to develop a model that satisfies the project goals.

A model is a simplified representation of aspects of the real world or how it works. The model needs to be detailed and correct enough so that it can be used to draw conclusions concerning the real world or to provide effective transfer of training. Important attributes of a model are that it reduces the complexity of the real world, emphasizing aspects that are deemed necessary or interesting. A model consists of components and relations between these components, usually in some sort of causal relation, and the relation between different components of the model is very explicit. Models are often designed in order to discover or test hypotheses of how the real world works, adding or deleting components, changing parameters or their values, or changing the relations among parameters.

The successful implementation of a conceptual model in a given modeling architecture or tool depends on the constraints and capabilities of the architecture. Trade-offs between model complexity and development costs need to be considered and the trade-offs depend on the project goal. For the choice of any tool or architecture, care must be taken so that relevant aspects are included in the chosen architecture and that non-relevant aspects have been abstracted away. Even important, relevant aspects can sometimes be ignored if it can be assumed that the effects are constant during the course of a simulation, particularly if these aspects are difficult and costly to model.

The HF or phenomena of interest for the human behavior representation community range over at least what Anderson (2002) calls “seven orders of magnitude,” where both neurological phenomena best described on a timescale of tens of milliseconds to social behaviors where hundreds of hours is a more appropriate timescale, can be of interest. Given this and the wide range of project and research goals, a large number of different architectures and tools have been developed by the modeling and simulation community for various purposes. No single architecture has been found to be adequate to represent human behavior for all military M&S purposes.

2.7 ARCHITECTURAL REVIEWS

It is not the intention or role of the NATO RTO group to recommend a particular modeling architecture or tool, but in the research literature several reviews and comparisons can be found. All the reviews presented

below provide important criteria and information to consider when a project is faced with the issue of evaluation whether a certain architecture or tool fit the goals of the project.

In Pew and Mavor (1998) 11 modeling architectures are described with regard to the following aspects: Purpose and use, Underlying assumptions, Functionality, Operation, Features of implementation, Support environment, Validation and Applicability.

McClanahan et al. (2001) describe 20 broad technologies and modeling architectures relevant to HBR using a detailed set of criteria based on the following aspects: Scientific basis of applicability for different types of behavior (physical, psychological, organizational), Technological maturity, Computational performance, Availability, Funding, Costs and Range of applicability.

In Archer et al. (2003), five different task network based tools/architectures are compared with respect to the following aspects: Principal purpose, Features supporting manpower, personnel and training requirements analysis, Additional analytic and data capabilities, Platform and Distribution.

Morrison (2003) reviewed 19 cognitive architectures and provides some recommendations concerning the application of certain architectures to certain research questions. Attributes considered are Principal Metaphors and Assumptions, Cognitive/Behavioral Functions Represented, Technical considerations, Input and input aids, Model output and analysis tools, Implementation language and interfaces, Evaluation, Model Purpose and History of Development and Applications.

Ritter et al. (2002) review modeling developments and issues including seven architectures that emerged after the Pew and Mavor (1998) review. The architectures are not described using as specific criteria as in the reviews above, but the review provides a wealth of information and guidance relevant to choosing a modeling architecture.

Similarly, the description of the parallel model development in four different architectures conducted in the Agent-based Modeling and Behavior Representation (AMBR) project (Gluck & Pew, 2005) provides very valuable information and experiences that should be considered when choosing tools or performing a comparison of architectures.

The report on human performance models for prediction of human error by Leiden et al. (2001) also uses more generic headings when describing the applicability of nine different modeling architectures and three vision models to the modeling of human error.

In an AGARD report, Alnaes et al. (1998) described an expert system called HOMER (Human Operator Modeling Expert Review), intended for assessing the suitability of a given architecture or modeling tool to a certain project is presented. The HOMER concept demonstrator included 13 modeling tools and architectures current at the time. The content of a HOMER-like tool would need updating if this line of thought and advice system development were pursued again, but many of the criteria used would likely still be valid.

It is worth noting that very few commercial CGF tools (for example STAGE, VR Forces, STRIVE, or FLAMES) are included in the eight reviews mentioned above, even though these tools are used to study many system areas involving HBR or at least HF elements.

It is also important to distinguish between the merits and features of a specific architecture and the availability of out-of-the-box or built-in behavior sought by some projects. For a certain project, one architecture or tool

could be developed with more appropriate features or abstraction level, while another might come with more out-of-the-box behavior and this could be a factor affecting the choice of architecture.

As often the case, there is no such thing as “the universally best tool.” There are many different tools in the toolbox and it is the goal of the specific modeling project and the type of questions that need to be answered that govern whether a certain architecture or tool is “fit for purpose.”

2.8 ARCHITECTURE COMPARISON

A search on the internet and in the proceedings of relevant conferences unearthed a large number of different modeling approaches used by academia and industry to model human behavior. Knowledge about the properties of the many different approaches takes a long time to learn. In order to provide some rough guidance on the similarities of differences of architectures for the uninitiated, the NATO RTO group asked seven HBR specialists (both from the group and external experts, from research and industry, with an average of 10 years of experience) to rate the similarity of 16 modeling architectures. The statistical technique MDS (Multi Dimensional Scaling) was used to analyze the results.

Each HBR specialist made pair wise ratings of similarity between 16 architectures on a seven point scale. They were also asked to rate their familiarity with each modeling architecture. The 16 modeling architectures compared were:

- ACT-R act-r.psy.cmu.edu
- AI-Implant www.biographictech.com
- Brahms www.agentisolutions.com
- COGNET/iGEN www.chisystems.com
- DI-Guy www.bostondynamics.com
- EPIC www.umich.edu/~bcalab/epic.html
- FLAMES www.ternion.com
- IMPRINT www.arl.army.mil/ARL-Directorates/HRED/imb/imprint/Imprint7.htm
- IPME <http://www.maad.com/index.pl/ipme>
- MicroSaint <http://www.maad.com/index.pl/products>
- MIDAS human-factors.arc.nasa.gov/dev/www-midas
- ModSAF homepage no longer available
- OneSAF www.onesaf.net
- Soar sitemaker.umich.edu/soar/home
- Strive www.cae.com
- VR-Forces www.mak.com

This selection does not imply any recommendation on behalf of the RTO group, and the different architectures were chosen in order to include approaches with different features and intended uses. Only one web reference has been selected for inclusion in the list above, but several other references could be cited.

The 16 architectures basically form four clusters, one in each quadrant of the Euclidean space as shown in Figure 2-3. The cluster in the lower left quadrant contains commercial CGF or scenario generation tools. The cluster in the lower right quadrant contains architectures often labeled as cognitive architectures. The upper right quadrant contains the task network architectures (although two contain micro-models of cognitive performance). The upper left quadrant contain two architectures that seem to focus on visualization of behavior or advanced digital mannequins. This indicates that even though all the 16 architectures compared here are used to model human behavior, they show rather different approaches and views upon how human behavior and cognition are represented. The different clusters also represent clearly different application areas where human behavior representation of some form is used.

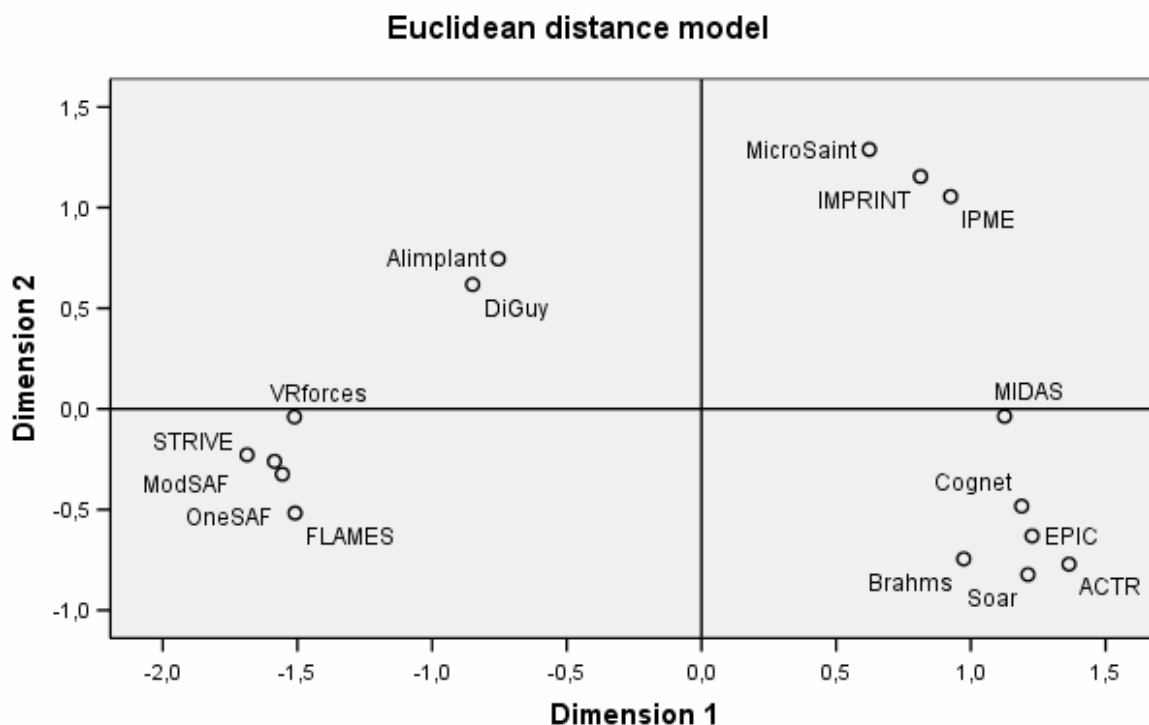


Figure 2-3: Graphical Representation of MDS Analysis of Architecture Comparisons.

Given that no modelers were expected to have extensive knowledge of all architectures, Figure 2-3 should not be analyzed at any finer level of detail than what has been provided above. It should be noted that for FLAMES, AI Implant, Brahms, DI Guy and STRIVE, the mean familiarity ratings were much lower than for ACT-R, IMPRINT, IPME, Soar and MicroSaint.

The dimensions in Figure 2-3 are not easy to interpret. Dimension 1 seems to distinguish between operational orientation (left) versus cognitive orientation (right). The meaning of dimension 2 is less clear but could be task orientation (top) versus human plausibility (bottom), at least for the two quadrants to the right. The specialists mentioned the following aspects as leading in their comparisons:

- Level of representation of behavior (e.g., cognitive architecture versus task network model versus scripted entity behaviors versus AI solution).
- Psychological fidelity and degree of human plausibility.

- Application area, typical use, and target market.
- Focus on cognition, perception, or behavior.
- Visualization.
- Model concepts.

The most significant feature in Figure 2-3, however, is the clustering itself. Apparently, schools of models exist that are not well integrated. Alternatively, the judges may have been familiar with their own school and perceive the distance with the other schools as large, implicitly making the same point. This could be seen as an underlining of the desire of HFM-128 to bring several stakeholders in modeling together.

Chapter 3 – MODELING EFFECTS BASED OPERATIONS

Before being deployed to their current location, Lt Anderson and his platoon had participated in several training exercises tailored to this country and the current situation. The training for this kind of humanitarian operation included not only the usual convoy procedures and possible interaction with armed insurgents (modeled via HBR) but also training to prepare for other possible effects – both positive and negative. The delivery of medical supplies is not as simple as it seems. For example, the delivery must be pre-coordinated with the proper local hierarchy so that they can be part of the acceptance process. If the supplies are given directly to a hospital storeroom clerk, the action will be seen as impolite or worse, as an attempt to thwart the power structure. This level of detail and its influence on daily operations was included in the training simulation via the decision making in the HBR entities of the local inhabitants, combined with the mission environment of the scenario. This level of detail is important in the context of effects-based operations.

3.1 EFFECTS BASED OPERATIONS IN PRACTICE AND IN MODELING

Increasingly, effects based operations (EBO) are gaining prominence in the way the military operates. The traditional concept of attrition warfare for combat has lost significance now and it is inappropriate for peacetime operations. Reduction of casualties and collateral damage has broadened the range of military options and the use of violence is just one means to impose one's will on the adversary. Other, less lethal or destructive means may be used to achieve the objectives. EBO thus changes the planning of an operation, investigating more military options and evaluating the effects and costs of each option. The emphasis on the decision process calls for a higher cognitive effort. EBOs require more staff for the evaluation and decision making process, but it is often more efficient to delegate specific, local decisions to subordinate units, who have the required intelligence on the spot. Issuing orders is thus replaced with assigning aims: aims to achieve rather than things to do. Recruitment of cognitive powers at the lower level has a profound effect on the unit, resulting in faster decision making, increased flexibility and greater trust between teams and leadership.

Effects based models need to follow this trend. For models of entities within EBO simulations, this means that SA, tactical option generation and decision making occur at lower levels in the military organization and less of the entity model can be adequately represented by drills and standard operational procedures. It also means that working with a fixed COA is becoming less realistic and that (plausible) variation in behavior is playing a larger role. The need for modeling the decision making process is increasing and the military shortlist of factors to be considered in decision making becomes explicit: tasking, terrain, enemy, resources, materiel and training. Sharing tactical information enables the development of a common operational picture and the aforementioned factors support the generation of alternative actions. For models the increased need for representing the decision making process is a challenge. Providing entities with simple condition-action rules is not sufficient any more for valid modeling results. An entity needs to be capable of making its own decision, and decision making in the model becomes more distributed. In Chapter 3.5, we will address this in more detail.

When Lt Anderson sought information about the crowd that was in his way and decided to make a detour, he made a decision that was not obvious in the past. The EBO training made him aware of unnecessary adverse effects of moving the crowd. In general, his thinking has changed. It may be unrealistic to predict all possible eventualities, but exercising a breadth of scenarios with

non-standard actions will reduce the risk of applying outdated, stereotypical responses to novel situations. In order to update Cold War era war games and fully address effects-based operations, HBR must be included to go beyond the stereotypical SOP threats of the last war.

3.2 TECHNICAL ARCHITECTURE OF OPERATIONAL MODELS

A generic operational model consists of several layers:

- 1) A computer system providing the infrastructure upon which to run a model.

The computer typically handles messages, update data bases, take input and produce output, log events, keep the clock and internal logic, etc. The database is shared by all parties in the simulation and the backbone (see Figure 3-1), takes care of the data exchange, regardless whether it involves a single computer, a computer network or computers communicating over the internet (in which case the database is probably copied). To make models interoperable, High Level Architecture (HLA) is used. HLA allows participants in a simulation to share the simulated environment.

- 2) A set of tools to set up a case.

Entities need to be defined and introduced; they get equipment, are told what to do, and are capable of performing certain activities. Weather and terrain are defined, etc. This layer of the model is highly dependent on the operation that is simulated. The environments for air, sea and land operations are thoroughly different and within these realms, cases differ also.

- 3) A set of sub-models, each taking care of how a certain aspect of the mission is performed.

Sub-models deal with the functionality of equipment and the world (weapons, tools, platforms, determining line of sight, shelter by obstacles, etc.) and the functionality of humans (the behavior). Examples are vision, communication, physical strain, decision making, SA, sleep deprivation, fear, and many others. The more functions are included, the more versatile the simulation is (and the more complex, calculation intensive, error prone). Some of these sub-models are bound more to human functions than to operations and should be transferable between simulations. HF can sometimes be implemented as component models or a selection of component models of varying degrees of fidelity within a self contained synthetic environment. Models that have the level of detail that is justified by the operational context are selected for the entity representation. Alternatively, the HF architecture may reside outside the synthetic environment and may interact with it much as a human player might as shown in Figure 3-1. This approach is likely to be more flexible and powerful than the monolithic approach but it incurs the additional networking costs.

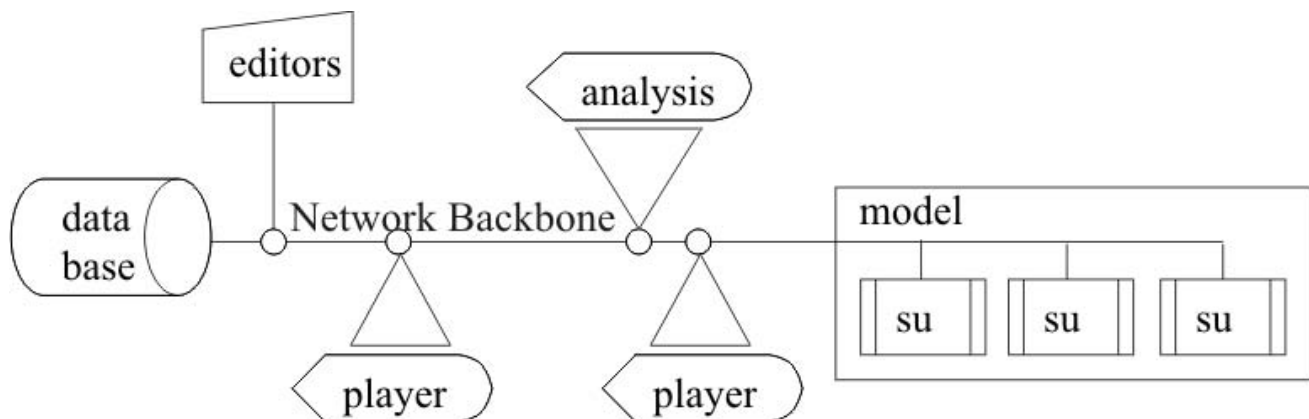


Figure 3-1: A Schematic of a Generic Modeling and Simulation Environment.

A common database, manual editors, displays with selective data for players, an analyst and a composite model are combined to form the M&S environment. The sub-models perform operational or HF functions. The editor is an input-output device and the analyst and players are terminals. The backbone is a communications link, managed by a data traffic manager.

Simulations may be controlled by a single analyst who incorporates all parties in the simulation and makes all the decisions as in many older combat models. Alternatively, players may play against each other, while the analyst is observing. One or more of the players may be represented by a computer generated player. The remarks on the advantages and disadvantages of involving players in the simulation, made by SAS-017, are still true: it is difficult to keep trained staff available for simulation studies. Systematic studies demand that the staff repeatedly solves the same problem, which is only possible by using new staff each time, but players are particularly important when the simulation model is less intelligent than desirable, for instance, when the purpose of the study is tactics development or when a player is also a student training in the simulation.

3.3 REQUIREMENTS FOR THE ANALYSIS ENVIRONMENT

Every operational study needs a modeling environment, composed of:

- 1) The larger setting (often called scenario);
- 2) The direct environment governing the planning and execution of actions (analysis environment); and
- 3) The effects under study.

A quest for a benchmark environment is not uncommon. How attractive this might look, the idea that a particular scenario or analysis environment could serve as a benchmark must be rejected. A generic analysis environment will in many cases lack the discrimination power the study requires because highly relevant events could be sparse. Rather, the analysis environment is an arbitrary compromise between operational relevance and sufficient involvement of the factors under study. In contrast to the analysis environment, the scenario is often fixed and selected from a shortlist. Some nations have agreed-upon high level scenarios from which to choose. Note that the word scenario has different meanings in different nations. For some, it is high level; for others, it is the action in the analysis environment.

The selected HF should not be determined exclusively by the analysis environment as shown in Figure 3-2. The scenario may depend on HF, but this is rarely explicitly recognized. The importance of HF on actions will in general increase when the entities involved are at the individual level. The scenario should specify factors that affect all personnel in the same fashion because such factors will have a significant influence on the outcome. Since these factors are not often specified, the inclusion of HF in the scenario should be advocated.

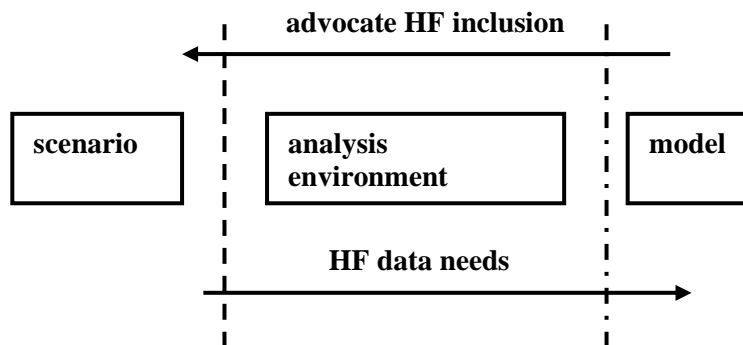


Figure 3-2: An Operational Model in an Analysis Environment.

The model asks for HF input data. If the scenario and analysis environment do not specify such data, the specification must be encouraged.

3.4 HIERARCHY OF AIMS

An EBO is carried out to fulfill an aim that is part of a greater plan. In military organizations, the aim is assigned in the context of a mission that allows the smaller unit to act constructively, even if the original plans have to be changed. Figure 3-3 demonstrates how the commander at level 0 splits his aim (his part of the mission) into tasks and roles for the entities under his command. The entity then makes its own detailed plan. The actual execution at level 0 is done by the entities at level -1, applying drills, rules and reasoning in an interdependent way. In this scheme the individual activities of the subordinate entities (level -1) are turned into level 0 team performance through communication of information and through synchronization of actions. This concept is a major enabler for scalability of models, keeping the complexity of tasking under control: for each level in the operation the scope is limited to one level down and two up. Scalability just means that this construct shifts along the hierarchical ladder. In reality it is more complicated since the HF at the various levels are not the same.

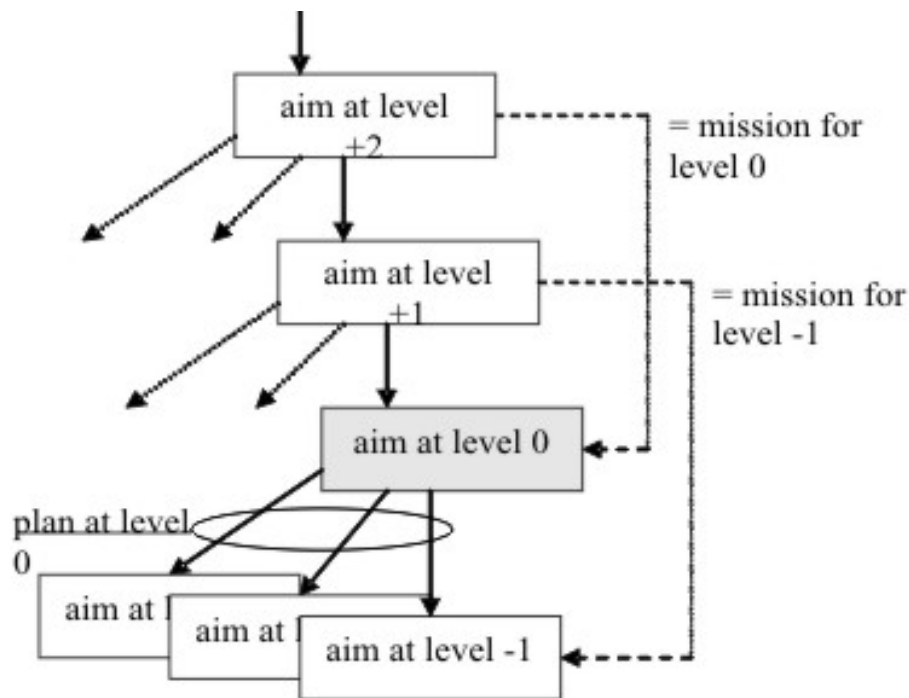


Figure 3-3: Goal Driven Behavior of the Organization as a Hierarchy of Sub-Goals (Aims).

The aim of the level 0 unit is achieved by tasking the subordinate entities at the level -1. The context for the aim of the level 0 unit (its mission) is the aim at level +2. As a consequence, even the lowest level is tasked with an aim in effects based doctrine.

Aims are achieved by performing trained activities. Military tend to discern numerous types of activities that are more or less alike but differ in purpose or context that may be handled parametrically or by extensive lists of vignettes.

3.5 OPERATIONAL PERFORMANCE

The question “What is operational performance?” is of key importance in group tasks. As a concept, operational performance seems to be an ordinal measure of performance that can be compared between operations, maybe done with different equipment, different methods, or different training. No such unique measure exists. In the effects based context, the only success factor is achieving the operational aim. Military exercises are evaluated by judges and they typically state their evaluation in terms such as “It took longer than usual to execute the exercise, but it was completed before dawn and there was no loss of aim.” This statement clarifies that the operation succeeded because the aim was met, despite taking longer than usual, but inference is that the duration was still acceptable. The aim may have been specified in more detail regarding quantity (number of weapons confiscated), quality (suspects missed at the road block), or level (rank of officer arrested in a humanitarian crime case). An operation to collect weapons may be successful if the specified number of weapons has been turned in, but only modern rifles are counted and only those in functional condition. There are just two outcomes of an operation: succeed or fail.

Both success and failure are associated with costs, such as how long it took, personnel requirement, the losses expressed in blue, green and red casualties, damage done, resources used, readiness after the action and others.

MODELING EFFECTS BASED OPERATIONS

If the entity has no other tasks than the one just completed, it is difficult to say what the actual cost of these losses is. Damage will be repaired, casualties treated, and readiness restored. However, if the follow-up plans involve deployment of materiel that is still damaged, if the entity is weakened by casualties and still tired, if ammunition is not replenished, the losses may prevent successful further action. It is the mission (the aim at two levels higher echelon in Figure 3-3) that determines the actual cost, ranging from little for a forgiving mission to high for a critical mission.

Consequently, in a simulation, operational performance cannot be defined without involving the higher levels in the scenario. In order to avoid excessive complexity when assessing performance at a level, the aims of the higher levels are specified and fixed such that the scenario does not accept dynamics above a certain level. This may seem so common in training or modeling that it may not even be noticed any more, but is not obvious in real operations. Stepping up immediately to high levels (calling for air support, political influence on field decisions) is becoming more common because of the availability of information in current tactics.

The metric is summarized in Table 3-1.

Table 3-1: Definitions in the Operational Context

Model Element	Character	Input	Output
Human Resource	Human competencies used for performing tasks.	Specified in the HF literature. May be moderated by training, instrumental leveraging or environment.	Moderated human resources.
Task	A familiar, basic or aggregate operational activity performed by an entity.	Is generated by the plan, quantifying the task.	The use of human and operational resources must be specified and the progress towards completion.
State variable	A human internal parameter that is representative for the state of an HF subsystem.	Human perceptual, mental or physiological processes as well as human traits.	The values feed into PSFs. States may reduce available human resources.
PSF	Operation independent function describing the way a task is done given the development of particular states of the entity in the course of time.	The task determines which PSFs are stimulated and how strong the stimulation is. PSFs may take states as input, but sometimes predicts performance without intervention of states.	PSFs produce quantitative task performance.
Task performance	Quality of task execution related to best possible task execution.	PSFs or given by the operational context as traits (training, fear, morale, etc.).	Relative achievement. Time and other cost factors are passed to the assessment of operational performance.
Effect	Function describing the operational consequences of the combined tasks of the entities involved.	Task performance, in the framework of the plan.	The effect created by the combined tasks and the combined cost.
Operational performance	Function comparing the achieved effects to the intent.	The achieved effects and the intent.	Success or failure of the intent and the cost in the perspective of further actions as required by the mission.

3.6 HANDLING INTELLIGENCE IN THE MODELING STRUCTURE

The challenge of the operational model is to handle those effects that are important in the field. Most models have a scientific bias because the model was created in a specific scientific setting. In Chapter 2.8, it was observed that modeling architectures tend to focus on a specific side of the problem. However, for the operation it is just as important if a soldier cannot do a task, if she/he cannot decide how to do a task, or if she/he is just not present to do a task, justifying the integrated approach. Intelligence is needed but in balance with operational features and physical or physiological capabilities. These aspects are also mutually dependent. The interaction between physical exertion and mental deterioration is militarily interesting because many campaigns have been won or lost over sustainability. An operational model should thus have a bit of both.

Models that are designed for operations involving tactics will include at least a couple of entities, each of which requires independent representations of a state of knowledge and decision making authority to be useful. The decisions need not always be sophisticated but must resolve problems such as tactical strategies for local reconnaissance, finding alternative routes, avoiding or seeking enemy contact, telling parties apart, etc. Some entities need more intelligence to make plans or resolve tactical problems. Given the computational burden of simulation of cognition on top of all other features, one might consider players to deliver the intelligence. For development of tactics and other high level cognitive challenges, it is doubtful if models could replace human players with the current state of the art. However, if each entity in a simulation needs a human controller for the cognitive representation, the number of staff involved is prohibitive. Operational models such as CAEn exploited expert teams and the GBR found it difficult to keep these teams available. In high level models, aiming at training of command staff, large teams are not uncommon. The general awareness that the manning of these teams becomes prohibitive is one of the drivers behind the need for good behavior representation. Figure 3-4 is a simple demonstration using elementary probabilities that show the likelihood that a simulation can be run as the size of the team of key personnel increases. It is clear that once the number of key personnel required exceeds about a dozen, the odds of being able to run a simulation decrease to 50:50.

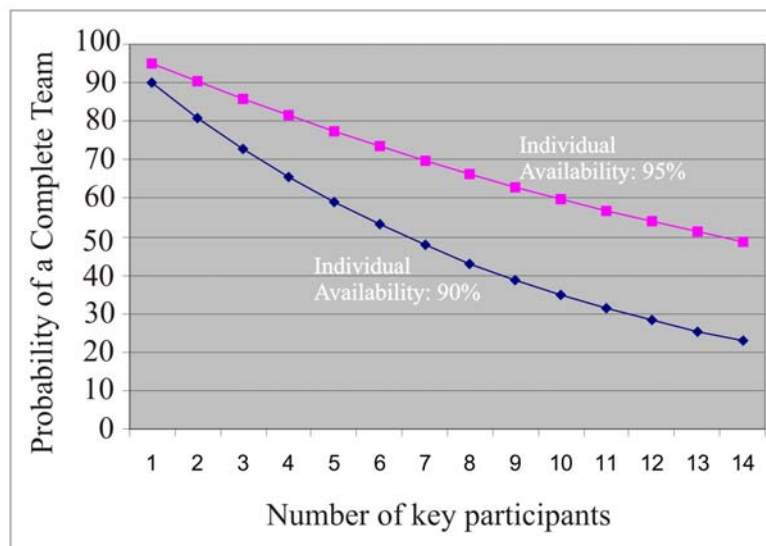


Figure 3-4: Likelihood that a Simulation can be Run Based on Availability of Key Personnel.

An intermediate solution would be to have one blue and one red player who provide the intelligence and decision making for all entities of their respective units. While this approach can be useful in some applications, as soon as differences in knowledge states within the unit become important, such as a radio communications failure or differing perceptions, the outcome would be too optimistic. If the player does the thinking for all entities, he and every entity is better informed than he could be in reality.

A more flexible approach would use grades of computerized intelligence for different entities, depending on their level, using human players when the required intelligence exceeds model or computing capabilities. One could assign the entities the intelligence that the application needs – in the ultimate case human intelligence. That would also be a useful architecture for training purposes. For the study of lower level military problems or small unit training, synthetic teammates play an increasing role. The dual purpose is to save trained staff, while avoiding the disturbance of the learning process by ill-performed roles in the team.

3.7 PERFORMANCE AND LEARNING FROM EXPERIENCE IN EBO

In EBO, the aim of the operation is to obtain a desired effect, and operational performance is the comparison between aim (desired effect) and outcome (achieved effect). This is at the heart of the conceptual model shown in Figure 3-5. If a unit has perfect knowledge of the environment, it will be easier to predict what will happen as soon as the unit takes action and it would likely be able to achieve the desired effect. Units with incomplete knowledge or understanding of a situation will likely experience failures and adjust their plan before they succeed. The discrepancy is noted and the analysis adjusted to match reality: they learn. Learning is probably the only way to discover what the opponent is doing because adversary intentions are observable only through their behavior.

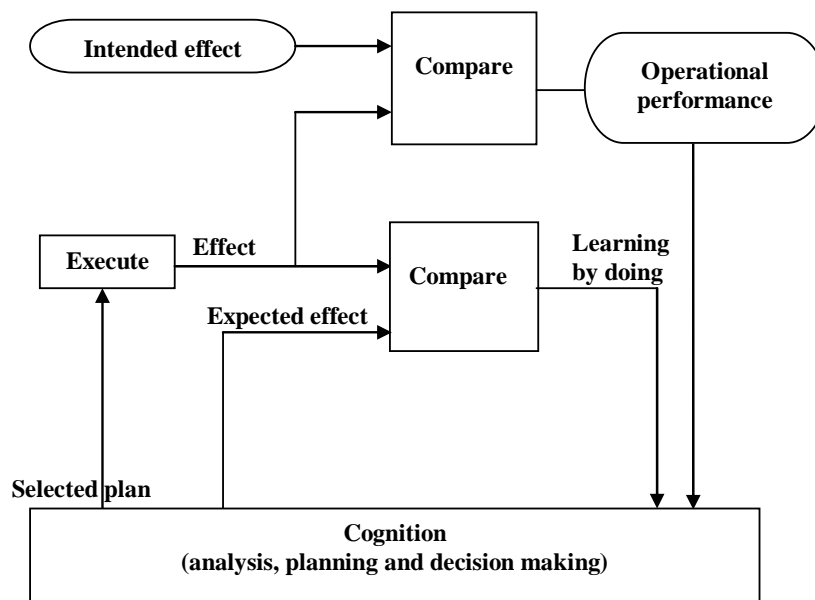


Figure 3-5: Goal Directed Behavior.

In effects based modeling the operational performance is related to the intended effect (part of the mission). The execution may have involved costs. If the expectation is not matching the real effect, the analysis is adjusted through the feedback loop, while learning by doing. The operational performance sets the target for the intended effect.

Some models have explicit learning algorithms, becoming richer as more experiences are collected. Observations of outcomes arising from actions (the learning by doing in Figure 3-5) are converted to a higher level of knowledge.

3.8 LEARNED BEHAVIOR

Militaries train continually, and training status is a key factor in performance. Training results in behaviors that have been classified into three basic types: automatic (immediate, drill based), rule-based (for more complex decisions, takes a while), and reasoned (time consuming, with detailed evaluations). Military training aims at quick and adequate responses through experiences that support the decision making process. This provides new rules to replace frequent reasoning and new drills to replace frequently applied rules. In the scheme of Figure 3-6 training is reducing the time, effort, and error associated with known problems. In particular, training can counteract the effects of stress on cognitive capacity.

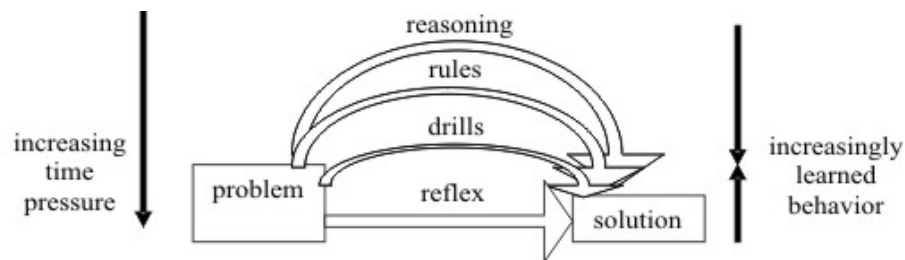


Figure 3-6: Modification of Behavior Through Training.

The natural behavior changes towards learned behavior due to training. Drivers behind this process are quicker reactions in all situations with fewer errors than would be observed from an untrained organization that only reasons under uncertainty with time pressure.

Learned behavior is easier to include in model algorithms than natural behavior. Nevertheless, remainders of natural behavior are still governing semi-automatic low-level processes such as search behavior, attention, cognitive functions or cognitive-emotional interactions or, at the other end of the spectrum, consideration of the unexpected. Modeling the effects of training should include the increased skill level, as well as the reduced taxing of cognitive capacity.

It is said that all but war is simulation. Thus, soldiers' training must be sufficiently realistic to achieve the curriculum objectives, including interacting with other teams, units, and forces in a natural manner, avoiding contrived or simplistic actions of other participants in the scenario. Representation of other entities in training simulations is another application of HBRs that can be developed to varying levels of resolution as required.

3.9 SEQUENCE OF EVENTS

In an engagement, the will of the parties is tested to the limit. In many combat models, the engagement is characteristically a Lanchester (1916) attrition "shoot-out." That is not realistic in today's wars. Threat is exerted to convince the other party to give up its will, not necessarily to kill people and destroy materiel. The ultimate test is in the use of weapons until one of the parties exceeds its acceptable risk, acceptable loss or runs out of resources. If a party gives up, a new situation results that may call for re-planning. In terms of model outcome, a bifurcation in the COA is created. The same happens with any event, abruptly changing

chances of success: a casualty, a physical barrier, detection of a new entity, etc. Each branch has its chances in terms of winning or losing and has expenses in the form of casualties, consumables, time and other. Several events in succession create a chance tree, in which each branch is a consequence of the preceding actions. The branching includes all possible courses of actions and outcomes. One important feature of human behavior is that humans continuously make estimates of risk and change their behavior accordingly.

The Special Ops team, tasked to resolve the hostage situation, is being backed by Lt Anderson's men. In more detail this is what might happen. They are approaching the location where, suspected, the hostage was taken. They are only generally aware of the insurgents since they have not seen them and the same is true for the other party. By chance, an insurgent on watch sees a glimpse of one of the SF. This dramatically changes the likelihood of the course of events. The insurgents hide and wait for the soldiers to come nearer, increasing their chances to eliminate some of the soldiers. Waiting too long would again cause a dramatic change because if the insurgents would also be detected, their chances would reverse. The situation took another turn when the hostage shouted and everyone involved took shelter, preparing for action.

Good military thinking allows for tactical changes during and after each engagement. Too many engagements may render the scenario tree too complex and a limit to the number of bifurcations must be set. Mastering and confining the problem is a major consideration both for the time and effort the study takes as well as for assessing the validity of the results.

There are two basic techniques to deal with the probabilistic outcomes. One is to systematically run all branches and build the stochastic of all possible outcomes. The stochastic allows for descriptive statistics and for sensitivity analysis. In this analysis, one parameter is changed at a time and the outcomes are compared. This method is transparent in the causes of outcomes. The most frequently used simulation technique has been the Monte Carlo method in which the model uses random effects (for visual detection, hits, moving the wrong direction, etc.), and each run is thus different. Many runs of the same operation produce outcomes that can be analyzed statistically.



Part Two – Discussion of Key Human Factors

PART TWO – DISCUSSION OF KEY HUMAN FACTORS



Chapter 4 – PERCEPTION

Lieutenant Anderson appreciates that the information from his helmet-mounted display does not interfere with his ability to perceive and act on information coming directly from his immediately surrounding environment. It is amazing, he thinks, very natural, the way his whole communication system works together. Information is sent not only on a video display but also in auditory and tactile forms. That tactile alert, Lieutenant Anderson thinks, is just like someone tapping my shoulder to get my attention.

4.1 INTRODUCTION

The ability for human operators to be consciously aware of their external environment is typically associated with processes of *perception*. Roth (1986, p. 81) defines perception as the “means by which information acquired from the environment via the sense organs is transformed into experiences of objects, events, sounds, tastes, etc.” The physiological and psychological mechanisms associated with perceptual processing are varied and complex. From a simplistic perspective, the ability to perceive begins with the stimulation of basic physiological systems associated with each sensory modality and proceeds to the higher-order brain processes that integrate and interpret the output of these physiological systems. Fundamental research in the fields of physiology, neurophysiology, neuropsychology, and cognitive psychology dating as far back as the late 1800s has provided significant insight into perceptual processes.

The range and volume of research on various aspects of perception are enormous as is the relative contribution of computational modeling within the field and are beyond the scope of the current document. There are numerous individual monographs and series of books reviewing perception and perceptual processes. Carterette and Friedman (1978) edited a collection of seven books containing articles from many workers in the field on various aspects of perception. Boff, Kaufman, and Thomas (1986a) have edited two volumes that deal with many aspects of perception; additional detailed information on various aspects of sensation and perception may be found in Boff and Lincoln (1988). In all these cases, adapting the information contained in these volumes into a suitable perceptual model for HBR may be difficult, although some features may be extracted and generalized.

For HBR applications, perceptual models provide the means through which the HBR is “aware” and able to perceive facets of the environment (real or virtual) in a manner similar to its human counterpart and forms an integral component of SA. Conceptually, all of the information about objects or entities in a simulation can be made available to an HBR all the time; however, a human in a similar real or virtual environment would not have access to the same information to the same degree. Therefore, the modeling of perception must provide a filter that reduces the information available to an HBR for further processing, similar to a human operator’s inability to attend to every stimuli in his/her environment. Filtering the available information reduces the likelihood that the HBR will perform in an implausible manner – something that could occur if it was “all-knowing.”

The degree to which perceptual processes are modeled varies considerably across modeling approaches and can range from purely response-driven implementations such as found in Semi-automated Forces (SAFs) to highly detailed models of the processes associated with visual perception such as Rybak et al.’s (1998) attention-guided visual perception and recognition model. In a military context where operators are typically faced with the requirement to rapidly and accurately respond to events within their environment,

the capabilities of the perceptual system play a critical role in determining the degree of success achievable within these conditions. To develop effective HBRs, the integration of perceptual processes is required in a reasonable and cost-effective manner. The following sections provide a framework for identifying the various components of perception and perceptual processes and identifying their respective relevance for integration within HBRs.

4.2 BASIC CONCEPTS

4.2.1 Sensation versus Perception

The distinction between *sensation* and *perception* has significant bearing in both psychological research and the implementation of models of perception within HBRs. The study and modeling of *sensation*, or sensory processes, is primarily focused on the initial contact between the perceptual system and the relevant stimuli within the external environment. As such, theories and models of sensory processing are typically restricted to the interactions between the physiological mechanisms (i.e., sense organs) that respond directly to external stimulation and the nature of the external stimuli that give rise to the various aspects of sensation. In this sense, the study of sensory processing is more concerned with the manner in which aspects of stimuli such as brightness, loudness, pitch, texture, and color are perceived via the sense organs.

In contrast, the study of *perception* and perceptual processes is concerned with the explicit experience of stimuli, or objects, and object relationships. In a more general sense, the study of perception is concerned with how representations of the external environment are developed and maintained and ultimately, the accuracy of those representations. At a higher level, the integration of various aspects of sensation across modalities associated with a given stimulus is also of interest and plays a key role in the development of a rich perceptual environment.

For virtual agents, the degree to which various aspects of sensation and perception should be integrated within a model is context dependent. Modeling requirements should be determined *a priori* and be driven by the ultimate goal of the modeling effort. For example, the requirement to develop a high-fidelity model of color perception may be irrelevant in the context of a radar operator performing visual search tasks within a monochrome radar display but may be critical for the development of a virtual soldier using an advanced helmet-mounted display where information is delineated via color coding.

4.2.2 The Senses

Within the domain of perception, the sensory system is commonly divided into the five following components or modalities:

- Vision;
- Audition (Hearing);
- Olfaction (Smell);
- Taste; and
- Touch.

Each of these modalities possesses both separate and integrated systems that are associated with their operation. Of note is the significant body of research associated with the visual and auditory systems that far

out-reaches that associated with the other senses. This disparity is also representative of the level of advancement in the HBR domain with respect to models of vision and audition in contrast to models of smell, taste and touch. The state of research is not surprising, given that the vast majority of research conducted in military modeling and simulation is driven by an interest in developing complex systems where the human operator responds primarily to visual and auditory stimulation. However, as interest grows in the development of display alternatives such as haptic (or tactile, that is, touch or force-feedback-based) systems, HBR practitioners will have to consider the development of alternative sensory representations.

4.2.3 Locus of Information

The accuracy in which stimuli are perceived is moderated in part by the manner in which information is associated with a given stimulus. There are two sources of information that can be used to perceive stimuli accurately:

- a) Currently available sensory input; and
- b) Relevant past experience or knowledge.

Bottom-up or *direct perception* is used to refer to processes mediated solely by sensory input, whereas *top-down*, or *constructive perception* refers to processes mediated by past knowledge, experience, and context. While researchers still debate these two theoretical approaches, the distinction between direct and constructive perception is relevant for the HBR practitioner by identifying the varying degrees to which the simulation environment provides specific stimuli and to which the HBR stores information that can be used to moderate perception.

With theories of direct perception, Gibson (1966; Gibson, 1979) claimed that there is a significant amount of information inherent in sensory stimuli that exceeds the simple physical construction of the related object and includes meaning, all the potential uses (or affordances) of an object, and the like. Gibson's approach rejects the conventional notion that the meaning associated with a stimulus is inherently related to past experience, context, or knowledge stored in long-term memory. Applying this approach to the development of an HBR shifts the locus of meaning from internal memory stores of the HBR itself to external data elements, or ontologies, of stimuli in the environment. This approach may be useful to reduce the computational demands associated with representing elements of knowledge within HBRs and offloading knowledge to an ontology of virtual objects that contains all possible uses and meaning for each object into the virtual environment. In multi-agent environments, this approach may drastically reduce the computational overhead associated with the operation of multiple, complex and knowledge-driven HBRs, as well as reduce developmental overhead via the re-use of object ontologies. Silverman et al.'s (2003) Performance Moderator Function Server (PMFServ) and the popular SimsTM computer game¹ are examples of affordance-based approaches to reduce HBR overhead. Each of these examples off-loads the possible uses or actions associated with objects to objective identifiers within the virtual environment, thereby removing the requirement to represent the association of semantic or procedural knowledge to specific objects within the HBR.

In contrast to direct perception, theories of constructive perception suggest that perception is not directly provided via stimulus input but is a result of the integration of sensory mechanisms and higher-order cognitive processes such as memory and experience. Perception is then influenced by past experience and hypothesis about the world which is not always correct and therefore prone to error (Palmer, 1975). In the context of HBR development, a constructive approach implies that knowledge and meaning about external stimuli must

¹ <http://thesims.ea.com/>

PERCEPTION

be embedded within some knowledge structure within the HBR itself and not within the external stimuli. Accurate perception of the environment is then determined by the accurate assignment of the relevant prior knowledge to existing stimuli; errors can then be associated with inappropriate retrieval and association of prior knowledge.

For the HBR practitioner, the selection of a direct or constructive approach to modeling perception determines the degree to which memory and past experience must be accounted for within the HBR or should be accounted for within the environment. In applications where experience, training, knowledge, and context are of interest, approaches grounded in constructive perception are best suited and drive the requirement for a sufficiently robust representation of memory, experience, etc., within the HBR. Approaches that apply theories of direct perception may reduce the computational cost of modeling individual memory stores but will be less effective at demonstrating the impacts of experience, etc., on performance. Compromises between the two approaches may satisfy both requirements by offloading non-critical knowledge to the environment while maintaining higher-fidelity models for critical operations.

4.2.4 Attention

The concept of attention is most often used to refer to selectivity of processing and describes the ability to choose, out of a myriad of stimuli, a subset to process in further detail, ultimately resulting in the perception and response to that stimulus. In operational contexts, the examination of selective, or focused, attention lends insights into how best to optimize the display of information to facilitate the selection process and allows researchers to determine the fate of unattended stimuli (e.g., does processing occur at some level and if so, does it disrupt ongoing processing?) and of extraneous stimuli (i.e., clutter or information irrelevant to the current task). In addition, the wealth of stimuli in the environment often exceeds the capacity of the human information processing system, thus preventing operators from analyzing the complete content of the environment. The ability for operators to respond to unexpected stimuli that demand attentional resources must also be considered. Effective models of attention must then provide the ability to model selectivity of processing, place constraints on the information processing system in a realistic fashion, and provide a means through which unattended stimuli can capture attention.

Further, the nature of complex systems and environments is such that operators are often required to attend, or focus, on multiple stimuli both within and across processing modalities (e.g., visual, auditory, tactile) and differing physical locations within a given environment. It is therefore of interest to examine the extent to which information processing can be divided or shared between two or more stimuli. Studies of divided attention and multi-tasking then lend insights into processing limitations, capacity, and the underlying mechanisms responsible for information processing.

The evolution of the field of attention is at the foundation of modern theories of human workload. Indeed, approaches to the assessment (via both objective and subjective measures) of operator performance within complex systems is derived from foundational research in the fields of focused and divided attention (e.g., Kahneman, 1973). The workload algorithms themselves are more properly considered as characterizing a type of performance moderator and are discussed in Chapter 6 on sources of variability. Nevertheless, it is worth mentioning here that an important characteristic of mental workload is linked directly to the concept of limited capacity perceptual modalities or channels. The resource capacity of a particular modality defines the capacity of that modality to process associated stimuli without incurring decrements in performance due to channel overload. For example, in the visual, auditory, cognitive, psychomotor (VACP) concept of mental workload (McCracken & Aldrich, 1984), V and A are considered input channels that operate in parallel, while C is the central or higher processing of the information once it has been perceived

and P refers to the motor action of a speech, hand, or leg movement. Similarly, Wickens (2002) has elaborated considerably on this simple VACP model with multiple resources theory that posits between- and within-channel conflicts; however, the basic notion is the same: perceptual input comes in via distinct channels, to be integrated at some point “upstream” from sensation.

4.3 MODELS OF PERCEPTION

The following section provides a brief overview of each sensory modality and its respective modeling approaches and provides some recommendations as appropriate approaches for integration within HBRs. To provide the analyst with a framework from which to develop appropriate perceptual models, it is useful to provide a representation of the stimuli, the sensory system, and perceptual systems to identify components that should be considered during the modeling process as shown in Figure 4-1.

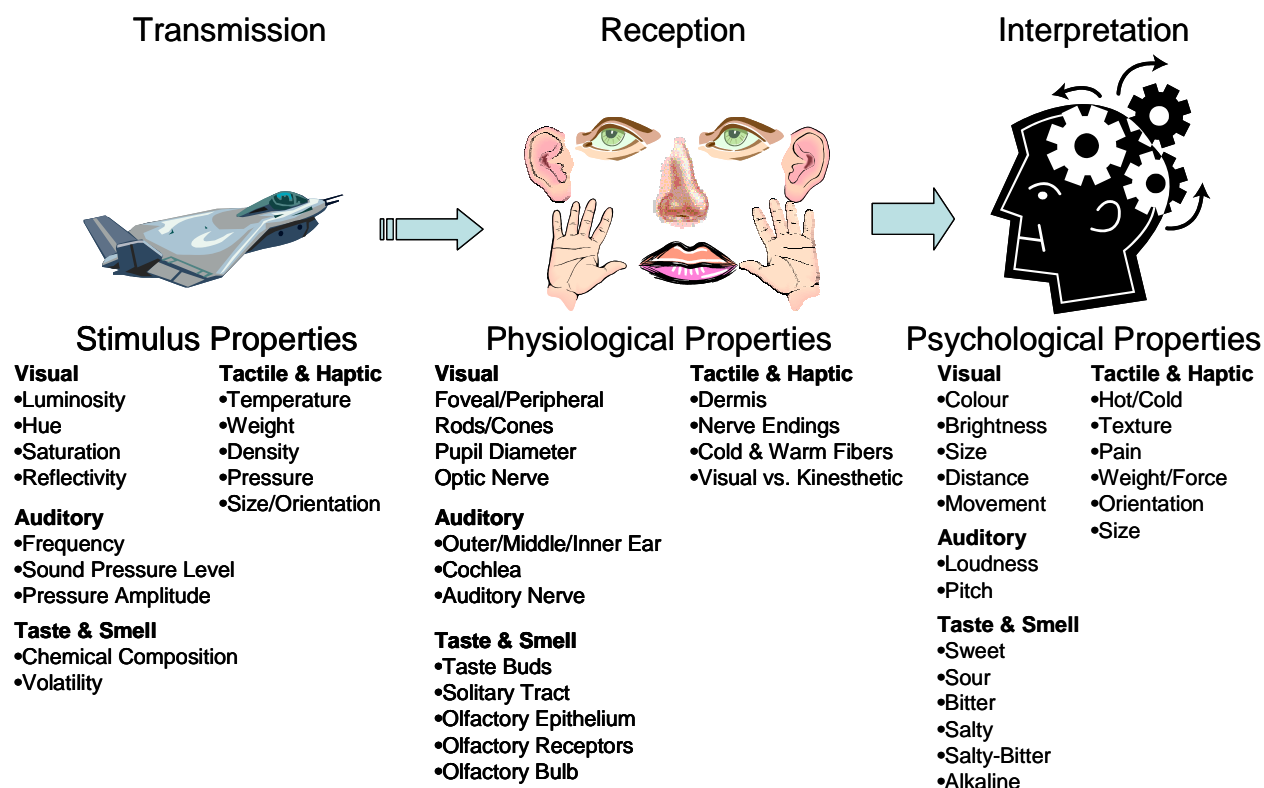


Figure 4-1: Considerations for Perceptual Modeling – the Links between the Physical Environment, the Sensations and Perception.

4.3.1 Stimulus Properties

Stimulus properties are composed of the interactions between physical objects, environmental effects (such as ambient light), the medium of transmission (e.g., water or air), and intervening factors such as interactions with other stimuli (e.g., two sounds generating additive interference). Within virtual environments, the attributes

typically associated with stimuli are either not present or have been vastly simplified (e.g., the lack of a medium through which to transmit sound or visual perception defined by a straight line-of-sight calculation as seen in many SAF systems such as the OneSAF TestBed). For the HBR modeler, it is important to determine the degree to which stimulus properties are available within a given virtual environment in order to determine the limitations or constraints imposed on perceptual modeling.

Virtual environments also simplify the challenge of determining stimulus properties such as size, location, and identity by providing such information inherently within the coding of the objects themselves. This has significant implications for the HBR practitioner since the requirement to infer stimulus information through perceptual processes can be ignored and instead information can be extracted directly from the objects themselves through affordances. As discussed previously, the use of affordances has both limitations and advantages that must be considered by the modeling practitioner.

4.3.2 Physiological Properties

Computational models of the physiological systems associated with sensation and perception have typically been applied to modeling the lower-level properties of the sense organs. Modeling efforts in the development of computational models of visual physiology (e.g., Freed, Smith & Sterling, 1992; Greenberg, Velte, Humayun, Scarlatis & De Juan, 1999; Pattanaik, Fairchild, Ferwerda & Greenberg, 1998; Winslow, 1989); mechanisms of the outer, middle, and inner ear (e.g., Duncan & Grant, 1997; Paillard, Mabilieu, Morissette & Soumagne, 1992; Treurniet, 1996), models of touch, pain, and haptic sensation (e.g., Chen & Sun, 2006; Dandekar, Raju & Srinivasan, 2003; Prince, Campbell, Picton & Turner, 2004), and smell (Linster & Cleland, 2004). Physiological models in these fields provide predictions of physiological detection thresholds and dynamic ranges (minimum stimulus levels, frequency range of hearing, saturation levels, etc.) and the adaptability of a sense to changes in stimulus properties (illumination, sound pressure, touch, etc.).

Computational models of sense physiology provide the front-end to the HBR framework and define the initial limitations of the overall system at receiving input from external stimuli prior to conscious awareness and action. Variability associated with individual capacities, injury, and other performance moderators may directly modify the physiological systems and manipulate performance (see the Moderator section for further discussion).

The integration of physiological models within an overall HBR architecture is dependant on two factors:

- 1) The level of information pertaining to stimulus characteristics available within a given simulation.
- 2) The degree that physiological constraints on information processing affect performance.

As noted previously, virtual environments typically do not provide sufficient depth of information pertaining to characteristics of stimuli within the environment. For example, it is unlikely to obtain realistic models of sound propagation, pressure levels, and frequency within a virtual environment that would be applicable to stimulating a detailed model of the cochlear responses of the middle ear. The availability of such information drives the selection of the appropriate physiological model and will likely be the limiting factor in model selection.

The second issue signals the necessity of defining the model requirements for a given application and may be in contrast or drive changes in the availability of information within an environment. The HBR practitioner must be diligent in selecting an appropriate level of representation that addresses a specific research or analytical interest prior to incorporating a model within an overall HBR framework. For example, if the

mission scenario and associated simulation environment support examination of issues related to cockpit display brightness or night vision goggle sensitivity, then the perceptual model must be at a level of resolution that matches those issues. On the other hand, if the question to be investigated via the simulation has to do with commander decision making as a function of number of incoming messages, then the output of a detailed perceptual model may be overwhelmed by other mission and scenario factors. In addition, the practitioner must ensure that the virtual environment has an adequate set of characteristics that will support an overall research goal (e.g., detailed environment representation of sound characteristics to evaluate auditory detection thresholds). Again, it is a matter of matching the HBR to the mission and scenario requirements and to the simulation characteristics.

4.3.3 Psychological Properties

Modeling of post-receptive perceptual processes such as attention, object recognition, speech and language perception is perhaps the best developed aspect of perceptual modeling by the HBR community. Of note is the significant focus in the community toward the development of visual and auditory models of post-receptive perceptual processing, with some recent interest in the tactile sense but essentially none in the development of models of taste or smell. The focus of research is not surprising, given that the vast majority of model development is conducted in support of assessing the performance of complex systems where the human operator responds to primarily visual and auditory environmental stimuli, and only recently is the development of haptic (or tactile) alternatives emerging.

In the context of current approaches to perceptual modeling, three distinct methods are available for integration within HBRs. The first method does not aim to describe or predict performance with respect to performance time or accuracy per se but by recognizing that information is indeed initially perceived on distinct sensory channels (e.g., visual, auditory, haptic), aims to model channel “usage,” so to speak. In this way, for example, the loading on the visual channel can be tracked for a given mission, scenario, equipment, and display set. If system design alternatives are being considered, visual channel usage can be compared to auditory channel usage; if task allocation across crew members is being considered, one person’s visual load can be compared to another’s, etc. McCracken and Aldrich (1984) developed a paper and pencil version of the VACP model to evaluate task allocation in the U.S. Army’s planned one-seat helicopter. Since then VACP has been put in a computational form as has Wickens’ Multiple Resource Theory and variations thereof (see Mitchell, 2000 for a discussion). Implementations of the channelized workload algorithms have typically arisen within task-network based toolsets such as IMPRINT and the Integrated Performance Modelling Environment (IPME). Extensions to pure loading-based approaches have given rise to scheduler-based approaches that allow modelers to derive an understanding of task performance that is unlike that provided by state/resource-based workload algorithms (Farmer, 2000; Hendy, Liao & Milgram, 1997). Specifically, it models the direct impact on task performance of the high workload generated by multiple tasks due to resource overload or structural interference effects (e.g., attempting to foveate on two disparate locations simultaneously).

The second is based on a probabilistic and data-driven implementation through the development of micro-models integrated with a variety of HBR-based modeling tools such as GOMS (e.g. John & Kieras, 1996; Kieras, 1997), IPME, IMPRINT, or MicroSaint, and evolved directly from psychophysics data on task performance (Card, Moran & Newell, 1983). Micro-models provide estimates of task performance such as task completion times for a variety of perceptual tasks (e.g., eye movements, reading rate, listening, ARL, 2005). Note, perceptual tasks in and of themselves do not predominate in the list of micro-models; rather, motor or output models are the best represented in the list, with some care taken to separate out perceptual times from motor response times. However, many of these models grossly simplify or make general claims as

PERCEPTION

to the time-course for perceptual activities and do not provide high-fidelity models of perceptual processes per se. In addition, there is a significant lack of micro-models pertaining to taste, smell, touch and haptic perception, and little interaction exists between perceptual micro-models and the front-end physiological models discussed in the previous section.

The third approach to perceptual modeling is grounded in the computational cognitive architectures, and arose through the desire to integrate cognitive architectures with external stimuli. Sense modalities are typically represented as modules within cognitive architectures, manage the transmission of the related stimuli for further processing within the core of the architecture, and therefore incorporate aspects of bottom-up processing. The impact of experience, knowledge and context can also be modeled within the architecture to mediate perceptual processing. As an example, (ACT-R: Anderson et al., 2004; Anderson & Lebiere, 1998) architecture includes a visual buffer and module that essentially segment portions of visual perception, both of which are separate from the intentional module and goal buffer. Inherent in ACT-R is the approximate 50 ms time cycle for each “mental” step and various stochastic processes that vary as a function of random noise as well as the effects of experience, knowledge, context, etc. In this way, performance time and accuracy accrue “naturally” through the perceptual process.

Of note is that the approaches currently applied to modeling perceptual processing typically do not model the synthesis and integration of stimulus properties into a single perceptual object. Rather, stimuli properties are assumed to have already been synthesized into the appropriate perceptual object that is available for further processing. As such, errors in recognition and identification are not typically based on errors associated with the integration process but are associated with aspects of memory retrieval or classification.

Chapter 5 – COGNITION IN HUMAN BEHAVIOR REPRESENTATION

When Colonel Schmidt and Lieutenant Anderson's equipment was first being explored in the concept stage, HBR modelers performed a series of M&S-based activities side-by-side with the military experts and designers responsible for developing new equipment concepts intended for use by a future multi-national coalition. Task-level models representing human performance at the granularity of minutes were used to predict which mission functions would benefit from the introduction of new technology capabilities. For example, using task data for a range of Soldier tasks – mission planning, squad navigation, and casualty care – the usefulness of individual Soldier global positioning systems (GPS) accuracies was estimated using metrics such as time to perform a mission, message load, and probability of mission success. These studies were largely based on SOPs and required few human data beyond typical task completion times and error rates. Other HBR models of military personnel in various occupations were used to determine the “cognitive fit” of display concepts such as helmet-integrated on-demand information displays at a granularity of less than a few seconds. These studies required support from the human sciences, introducing psychologically plausible models of perception, decision making, and planning that were used to test concepts in a variety of scenarios. These HBR operators were further stressed by the introduction into the simulation of bright sunlight and hot environments, requiring additional moderator models of visual sensation and thermal physiology. On the basis of these early HBR performance predictions, certain technologies were ruled out and for the technologies that showed promise, specific design recommendations were made to improve operational performance while minimizing operator workload.

5.1 INTRODUCTION

The NATO M&S Master Plan (NATO, 1998) notes that current computational models and simulations of operator behavior do not adequately represent human performance, neither at the individual nor the small group level. While some progress has been made, widespread improvement does not seem to have occurred over the past several years. That is, we do not currently create constructive operator simulations in the range of complexity and capability that are desired for many existing applications, such as training and rehearsal. Nevertheless, there is a wealth of knowledge from psychology and cognitive science that we can draw upon to rectify these shortfalls in some measure.

Most CGFs produced today are tied closely to doctrine and while doctrine is often modeled, it is seldom observed in a pure form in practice. Predictability, based on textbook doctrine, may be desired for preliminary instruction or training basic skill development, but it is inadequate for advanced training in decision making, situation assessment, experimentation with tactics, or mission rehearsal. Current CGF modeling relies largely on rule-based systems usually represented as production systems or frames (contextual groupings of condition-action rules, or schemas). Others instantiate the rules as networks of logically linked tasks and goals, typically oriented toward physical performance outputs rather than processing information. Approaches include intelligent agents and network models that are largely AI-based while some are better grounded in cognitive psychology. Most, if not all, CGFs in use require human controllers to intervene during the simulation. It is believed that prudent introduction of models from cognitive science will extend the capabilities of the current CGF technology, making them more robust and adaptable as well as introducing desirable, plausible variability.

COGNITION IN HUMAN BEHAVIOR REPRESENTATION

Cognitive reasoning, inference (of OPFOR intent or possible future states), SA, decision making, problem solving and reactive planning are not modeled well if at all. Current models are brittle (minor deviations from the design assumptions create unrealistic behavior), display simplistic responses that are not believable, and lack adequate cognitive abilities, tending to make them inappropriate for unscripted scenarios. This approach to human behavioral representation is too crude for many applications of interest, producing unrealistic, predictable or brittle responses. Reliance on comprehensive knowledge bases to improve robustness to unanticipated events is unlikely to be productive because of the high cost, even in the unlikely event that all contingencies could be anticipated. Again, models of learning and reasoning based on psychologically plausible processes are beginning to address these shortfalls.

There is a need for models at various levels of fidelity and resolution. Much insight can be gained by simple models exploring specific aspects of performance and behavior, incorporating little in the way of complex human behaviors and instead focusing on performance according to a script. The tools we have today are largely competent to handle this, needing only more refinement and usability.

For more advanced systems where the focus is on cognitive aspects, such as training or C3I studies, more complex models are required. Military analysts and trainers need predictive, computational human models that are affordable in development as well as execution. Similarly, force development as well as tactics and strategy analysts need some measure of human capabilities in their analysis to study the sensitivity of proposed solutions to human traits and states.

The “perceived world” is the cognitive state of awareness of one’s environment that in a military context is equivalent to SA, (Endsley’s (1993; 1995) Level 1 and 2 SA), while subsequent mental reasoning gives rise to an understanding of the meaning of new information, particularly projection to the future world (Endsley’s Level 3 SA). Such new information may be the perception of adversary behavior or the growing awareness that a chosen COA is inappropriate. The perceived world provides the context for understanding and presents numerous opportunities for disorientation, decision making errors, fratricide, etc., well known to the military.

Figure 5-1 lays out a conceptualization of how HBRs fit into military simulation. Perception and knowledge of objects in the synthetic environment lead to a difference between the perceived and desired state (either of the world or the entity) that result in a decision to act on the world. The action will probably have an effect on the entity states as well as the world, for example fatigue or sleep deprivation, resulting in both moderated behavior (slowing down). In the effects based perspective, the behavior creates an effect on the adversary, who is part of the problem context. The adversary has a complementary effect on the HBR, which is the essence of an engagement.

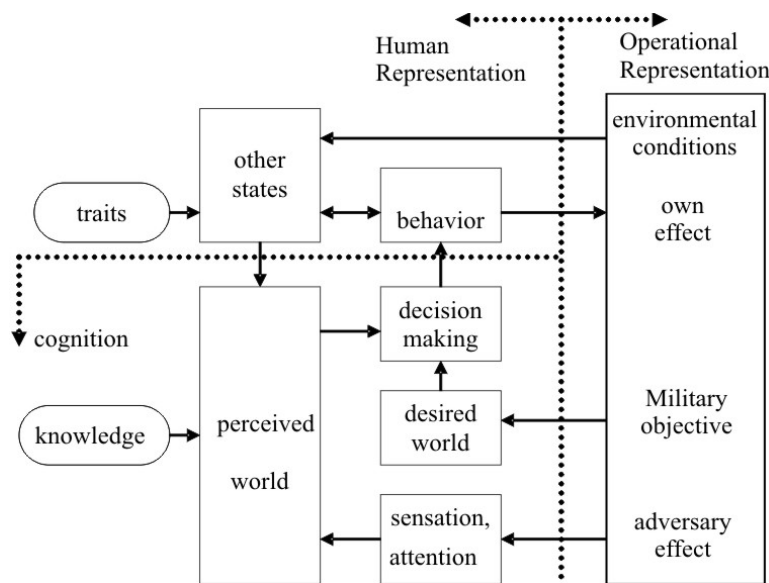


Figure 5-1: Scheme for Implementing HBR in Operational Studies.

Cognition interacts with the physiological human representation and the physical, operational representation of the environment. Behavior is driven by decisions, trying to make the perceived world match the desired world. Behavior is also driven by internal urges, stemming from stressors, however. Behavior affects the operational effect that is achieved, lifting it to the level of the problem context, in reality a battle of the will, if not a fight.

5.2 COGNITIVE SCIENCE CONTRIBUTIONS TO HBR

Cognitive science has the goal of understanding human cognition, so it is crucial that a theory of mental representation must be concerned about how people think. By contrast, AI, as a branch of engineering, is about effective or optimized problem solving, learning and language, regardless of how people think. Many, such as Hawkins et al. (2003) and The Technical Cooperation Program (TTCP) HUM TP2 Strategic Review (TTCP, 2000), note that accurate and believable representations of human behavior will be essential to realizing cost-effective simulation based acquisition, training, and rehearsal. In order that HBRs be credible, they should achieve levels of performance similar to that of people under similar conditions and should make errors similar to human counterparts. This suggests that, while AI approaches may find useful applications in HBR, validated cognitive science approaches are preferable over AI approaches if they are also computationally practicable.

Cognitive science has made many advances in the understanding of cognition over the past 50 years through the development of computational models that can represent numerous aspects of cognition. Although cognitive science lacks a unified theory that explains the full range of psychological phenomena, (Thagard, 2005, p. 134) there are many plausible cognitive models that may be useful in CGF.

A cognitive theory postulates a set of representational structures and a set of processes that operate on these structures. A model is an instantiation of the cognitive theory in a specific context, often in the form of a computer program consisting of data structures and algorithms. To evaluate a theory and model, a computer program is often used to generate results that can be checked against human results to determine whether the program gets similar answers and generates similar errors (Thagard, 2005, p. 13). Testing models as computer programs is an important step since it helps to show that postulated representation and processes are

COGNITION IN HUMAN BEHAVIOR REPRESENTATION

computationally realizable. This is important since many algorithms that seem reasonable at first glance do not scale up to large problems on real computers.

The effectiveness of an approach to mental representation depends on how it accounts for three important high level thinking activities (Thagard, 2005, pp. 16-17):

- 1) Problem solving: how people accomplish goals. This has four aspects to explain:
 - 1.1) Sense making;
 - 1.2) Planning (figuring out how to get from one state to another);
 - 1.3) Decision making (selecting from among a variety of choices to achieve a desired goal); and
 - 1.4) Explanation (reasoning and postulating why something happened).
- 2) Learning (acquiring facts, relationships, and skills).
- 3) Use of language (including the ability to understand, produce utterances, and learn a language).

Hawkins et al. (2003, p. 100) noted that there are few practical examples of unified performance models, principally because few modelers seriously attempt full spectrum modeling and few military organizations are prepared to invest the time and money to capture the extensive domain knowledge it requires. The cognitive psychology-based models such as ACT-R, SOAR, and COGNET/iGen demonstrate that it is possible to bring formal psychological concepts at least part way into operational HBR models and retain broader, formal cognitive and perceptual models without sacrificing performance.

The remainder of this chapter is divided into two sections that explore some frameworks for cognitive process models and constructs that may prove relevant for HBR. The first discusses modeling of the higher level phenomena characterizing simulations of personnel and their actions for defense simulation. The second section describes models from psychology that may be useful in developing the higher level phenomena representations. In many cases, these models have only been used in research to explore specific phenomena and their usefulness in HBR is yet to be proved. However, if it is believed that the means to developing robust, cost-effective CGF is through the introduction of human sciences in the form of HBR, then models incorporating the following concepts may well be a step in that direction.

5.3 HIGH-LEVEL HBR COGNITIVE CAPABILITIES

5.3.1 Situation Awareness

Lieutenant Anderson appreciates that the information from his helmet-mounted display does not interfere with his ability to perceive and act on information coming directly from his immediately surrounding environment. This is especially the case as the platoon travels through one particular neighborhood en route to the medical center. Although the robotic assets had given the all-clear, Lieutenant Anderson has a direct visual sighting of a crowd that has gathered and he is not sure of their intent.

SA is considered to be a state of knowledge of the world in context and the implications for the future. SA is occasionally discussed as something that is modeled rather than an emergent property of a model; however, SA should not be considered to be a modeling construct. SA is a way of representing the current state of

knowledge, whether it is within a model or an individual, to explain the rationale behind a chosen COA or expectations of future events to others.

Defining SA for military applications as the static spatial awareness of friendly and enemy troop positions is too simplistic and ignores many important militarily relevant details. Endsley's (1995) definition is usually accepted as being more complete yet remaining succinct:

SA is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.

Activities with high levels of complexity and detail (as shown later in Figure 9-2) such as mission rehearsal, staff and procedural training, SOP and doctrine development will require more comprehensive models that capture and use SA in a more human-like manner. In other cases where the cognitive processes are not central to the problem being addressed, more mechanistic, AI approaches may suffice. Prediction of operational outcomes and mission rehearsal will also depend on the level of representation resolution, with perception and decision making becoming increasingly important as the simulation focuses on individuals or small groups.

Credible SA will require adequate models of perception to explain and predict the effects of the cognitive processes involved in each of Endsley's three stages of SA as depicted in Figure 5-2. This is different from but overlaps current AI approaches to CGF that tend to rely on a complete knowledge of the environment and precise extrapolation of events. For example, it has been observed in practice that during times of stress and uncertainty, there is frequently a failure to consider an adequate number of facts to ensure accurate SA, including failure to verify assumptions in making decisions, failure to weight information on its quality, failure to interpret information (Level 2) and failure to make predictions (Level 3: Fallesen, 1993).

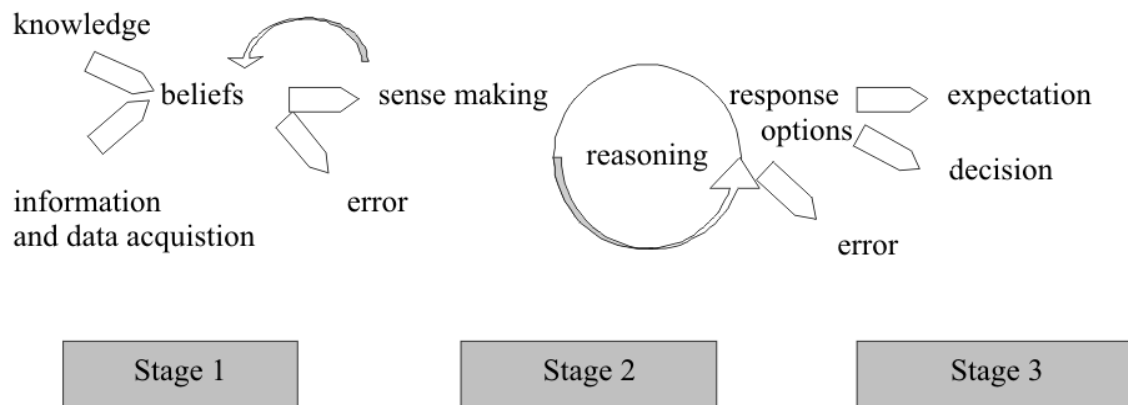


Figure 5-2: Mapping Cognitive Concepts and Processes on the Stages of SA (after Endsley, 1995).
The stages information state (1), understanding (2) and projection of future events (3) are the consequences of lower level cognitive processes that may lead to errors.

Perceptual and attention models should reflect human capabilities, limiting HBR to information a human operator would have access to rather than providing complete knowledge of the environment or allowing precise extrapolation of events. Such models may result in missed or misperceived environmental cues, with a consequent error in judgment based on correct reasoning with faulty data. To make the resulting performance

plausible, purely stochastic approaches are unlikely to be an adequate representation of such processes for most applications, underlying the need for models from the human sciences in the relevant domains.

Beliefs in the state of the world that evolve from the SA process (Figure 5-2) could depend on some form of reasoning either about what will happen next or how current events evolved. Alternatively, pre-programmed or instance-based memory models may come into play to predict future events as representations of expert knowledge or experiential learning in decision making. Planning models and exploration (mental simulation) to select appropriate responses should reflect human biases as well as capabilities rather than exhaustively searching the problem space for an optimal solution. Pew and Mavor (1998, pp. 199-202) predict that future HBRs will use a mixture of expert-system/production rules, case-based reasoning, and belief networks to capture many SA phenomena.

5.3.2 Problem Solving and Decision Making

Colonel Schmidt is thinking that despite all his experience and training, today's events have presented unexpected challenges. He finds himself continually assessing options and considering risks when directing his command.

Problem solving and decision making are concepts that describe the processes of memory retrieval, reasoning (usually to satisfy some goal), belief about the state of the world (SA) and choice (selection from a number of alternative solutions or actions) to form an opinion or a COA. These are key elements of models intended to support analysis in materiel acquisition, SOP and doctrine development, as well as to some extent training and mission rehearsal (see Figure 9-2). Representing human problem solving and decision making becomes increasingly important as HBR applications move from the concept exploration and design evaluation into simulations for safety assessments, predicting operational outcomes, training and rehearsal as there is a need to tie cause and effect within the simulations. Pew and Mavor (1998) suggest that Decision Theory for HBR problem solving and decision making be divided into two major topics: the choice among plausible alternatives and the modification of beliefs based on evidence (reasoning).

It has been reported that current model architectures that attempt to do problem solving and decision making for military simulations are too stereotypical, predictable, rigid, and doctrine limited and so fail to provide realistic characterization of the variability, flexibility, and adaptability exhibited by individuals; variability, flexibility and adaptability are essential for effective decision making in military environment. Also, for most HBR, the decision making process is too uniform, homogeneous and invariable, failing to incorporate effects of stressors, mental and physical states. HBR decision making should reflect individual differences such as individual traits (aggressiveness, impulsiveness, risk aversion, etc.) and should incorporate biases or judgment errors derived from incorrect perceptions. Further, problem solving models in military simulations should be capable of reasoning under uncertainty that results from partially observable environments to reduce the likelihood of brittle or predictable outcomes.

Many decision making theories do not consider operator state factors such as fatigue and stress. Models based on these theories need to be extended to be sensitive to common military stressors when incorporated in HBRs. Models of decision making used for HBR should be capable of plausible judgmental errors, reflecting speed-accuracy trade-off, base-rate neglect or confidence threshold effects as moderators (Chapter 6) to explain some of the observed range of performance.

Individual differences are important to providing a comprehensive study of the locus of effect of the problem solving and decision making processes over a number of operators. For instance, biases and heuristic

approaches could be introduced by integrating instance-based learning with decision making (DM) models to produce a synthesis that provides adaptive planning based on experience with previous episodes similar to the current experience (Pew & Mavor, 1998, p. 171).

5.3.3 Planning

He queries the network and his display re-activates instantaneously to show the message confirming that the crowd is not a threat, that they are gathered for the opening of a new school. The display fades and he decides on a detour so as to avoid any appearance of aggression, leading the platoon on toward the medical center by an alternate route that had been identified for such contingencies.

Planning is critical to successful operations, playing a key role in tactical decision making at all levels. Capturing the essence of the planning process in HBRs is thought to be essential to producing realistic behaviors at the tactical level in activities such as SOP development, staff training, and mission rehearsal but also at the organizational level when dealing with manning, policy, or doctrine development.

Formal planning processes will differ between nations or even within nations, depending on the service, but the underlying human planning capabilities should have similarities as the formal planning process is a learned behavior. It has been observed that, in practice, formal planning is seldom used (Fallesen, 1993), particularly under high time pressure situations; instead, heuristic or RPDM-like processes seem to be the norm:

If the ultimate objective of future efforts in human behavior representation is to model doctrinally correct planning behavior, this should be an adequate starting point for a requirements definition effort. If, however, future efforts in human behavior representation are to be aimed at modeling actual tactical planning, a serious knowledge engineering effort is needed to build a behavioral database describing planning during combat decision making. This conclusion follows from the observations of many researchers that the actual planning behavior observed and reported differs markedly from that prescribed by doctrine. (Pew & Mavor, 1998, p. 209)

Thus, planning as an activity goes back to how people think and reason as well as how we choose among alternatives, and HBR models requiring planning should reflect this fact rather than defaulting to a formal planning process.

Usually, planning is for a sequence of “allowable” activities that will move the system from an initial state to a desired goal state (or hypothetical states for contingency planning). The allowable activities must be defined by the modeler as an abstraction of the modeler’s understanding about how the world works. It is possible that “disallowed” activities could lead to new and better ways of doing things, relaxing the agent’s beliefs about what is doable or not. This suggests the need for experimentation or “mental” simulation within the HBR itself.

Planning can be represented as a sequential search (by various methods) through a set of rules (DND, 2002) or the U.S. 5-step process (Mission analysis/METT-T, battlefield assessment, COA generation, COA assessment, COA selection, and CAO refinement) but these procedural approaches are time consuming and computationally expensive. Fallesen (1993) observes that formal planning, even at the organizational level, is seldom used explicitly and is probably not how people generally plan as individuals.

This suggests that HBR models of both doctrinal and non-doctrinal approaches to planning will rely on detailed and extensive knowledge bases of service-specific domain information, dealing with highly complex

aspects, far removed from first-principle planning typical of early AI attempts where logical planning systems encountered problems with uncertainty or conflicting data, leading to newer, more robust formal logics or simpler heuristic approaches. Nevertheless, rules can be applied to reason forward from a state to a goal, or backward, from a goal toward the current state (or a combination of the two), and HBR planning models should attempt to capture the doctrinal planning process to some degree, using it as a guide rather than a rule-based system.

5.4 LOW-LEVEL COGNITIVE PROCESSES

5.4.1 Memory

Memory processes, both capabilities and limitations, are mechanisms that can be developed into HBR models to better reflect how people acquire skill, build mental representations of external information, form generalizations as well as make errors through misinterpretation, retrieval interference, confusion or forgetting. Pew and Mavor (1998) divide memory into three broad categories for HBR applications, each with some finer divisions:

- 1) Episodic, general and implicit;
- 2) Short term working memory; and
- 3) Long term memory and retrieval.

Pew and Mavor feel that representing episodic and generic storage are the most important of the three for HBR in military simulation.

5.4.1.1 Episodic Memory

... he recognizes that these behaviors are often associated with positive events in this culture and the atmosphere seems different from that when attacks were imminent.

Episodic memory refers to memory for events. That is, memories that can be traced back to a particular time and place (e.g., What did you have for dinner yesterday?). In very primitive terms, episodic and semantic memory can be differentiated by considering the former to refer to things you remember, while the latter to things you know.

Episodic memory is context dependent, involving the rememberer's personal experience in particular situations with specific cues. The Minerva2 approach is a model of episodic memory that has been used to reproduce a number of psychological phenomena. Generic episodic memory refers to general knowledge, not necessarily linked to circumstances, specific experiences, or specific contextual clues. Implicit episodic memory refers to retrieval of general knowledge from recent experience, even if the specific experience is not recalled. This may represent by ordering memory traces in terms of recency and strength, or number of co-occurrences so that more recent experiences can have an influence and dominate highly learned experiences or heuristics. Pew and Mavor note that encoding episodic memories into a more general rule-based, even over the short term such as an engagement rather over extended training sessions, is of relevance to military simulations, perhaps reflecting an interaction or a tighter coupling between episodic and implicit memory.

The models of episodic memory that exist have been developed almost exclusively for research purposes, explaining human performance in laboratory tasks. As such, it is difficult to make recommendations of one

model over another with any real certainty. Perhaps the important characteristic of the various models is their manner in which information is retrieved from memory. Retrieval in most episodic memory models collapse across the entire contents of memory in response to a probe. The ACT-R model is an exception, generally retrieving single items from memory in response to a probe. A major distinguishing characteristic of current models is whether the contents of memory are collapsed during encoding (e.g., holographic models such as TODAM, Murdock, 1982) or retrieval (e.g. instance models such as Minerva2: Hintzman, 1984). While the two styles of model are approximately equivalent in their ability to account for laboratory data using only a few training items, instance models provide a means of allowing memory to hold enormous amounts of data without the same computational expense. Hence, for a large number of traces in memory, instance models such as Minerva2 are an attractive option.

Of potential importance to those constructing simulated operators is the close association between episodic and semantic memory despite the separation the two have enjoyed in the theoretical literature. For example, familiar words are more difficult to recognize as “old” items in a recognition task than less familiar items. Also, words are recognized more quickly in a recognition task when they are primed by a semantically related word than an unrelated word. The interplay between episodic and semantic memory is well established in the psychological literature, and the formal explanations of the relationship are still a matter of debate. Two models that have tackled the impact that semantic memory has on episodic memory are the Retrieving Effectively from Memory (REM, (Shiffrin & Steyvers, 1997) and the Bind Cue Decide Model of Episodic Memory (BCDMEM, (Dennis & Humphreys, 2001) models. Both models use a vector representation for items, and both possess a representation for the human’s prior experience with the stimuli. As such, the models provide a starting point for modelers seeking to incorporate aspects of semantic memory into the episodic memories of simulated operators.

5.4.1.2 Short Term Memory

The system prompts him regarding exactly what information he has reviewed so far and highlights important aspects of the remaining information that should be reviewed to ensure getting the total picture. Colonel Schmidt is reminded of several important details he had forgotten during the interruption but quickly refocuses on the problem without missing information or losing time.

Short term memory (STM), sometimes referred to as working memory, refers to the temporary (several seconds to a few minutes) storage of information. The research on STM has historically focused on people’s ability to recall a list of items (most often, words) in order, immediately after they are presented. Models of short-term, working memory are thought to be important for military HBR simulations, particularly models of explicit retrieval or retrieval from generic memory, when encoding or retrieval of information is central to the application area. Learning is essential for intelligent systems that adapt, but current military simulations make little or no use of learning. A short term solution could be to modify current rule-based systems to incorporate episodic- or case-based learning processes to develop more robust systems. Neural network learning mechanisms that create symbolic rules are also thought to be of value.

There are several simple, yet powerful models of short-term memory. What differentiates them is the manner in which they encode order information for items that are recovered during recall. None of the models in the literature are very expensive computationally. Hence, if needed, they could be implemented in a simulated operator without sacrificing real time or faster-than-real time performance.

5.4.1.3 Semantic Memory

The delivery of the supplies to the medical center is well underway when Colonel Schmidt receives an urgent message about a possible hostage situation in a remote section of the city. He immediately realizes that this implies a number of steps to prepare for his new tasking.

The study of semantics represents the examination of how “meaning” is represented by the human. There are two general types of semantics that may be of interest to the person constructing a simulated operator: lexical semantics refers to the meanings of individual words, while compositional semantics refers to the meaning conveyed by words grouped to form sentences.

Early work on lexical semantics came in the form of semantic networks (Collins & Quillian, 1969). The networks, although informative, were hand-wired and so had limited ability to provide any explanation as to how the connections among words or concepts were created through exposure to language. More recently, however, unsupervised models of lexical semantics have been developed that are able to automatically develop semantic representations for words. Since the first such models came out around 10 years ago, several models have been developed that work on similar principles, namely, that words that are semantically related tend to co-occur in documents.

Of the models in existence, the most developed is Latent Semantic Analysis (LSA, Landauer & Dumais, 1997). LSA forms vector representations of the meanings of words by processing the content of a term by document matrix, tabulating the frequency with which each word occurred in each of thousands of documents/context. Thus far, LSA has been shown to perform well on tasks such as the TOEFL synonyms test and marking undergraduate essays. Since lexical semantics models provide representations for individual words, they have limited utility as models for understanding sentences. Lexical semantics models may be suited to providing flexibility in processing individual words or concepts. Depending on the size of the corpus used to train models such as LSA, the speed with which they operate can be close to real time.

Compositional semantic models attempt to represent the role that syntax plays in the construction of a meaning. The general notion is that a propositional representation of the sentence captures the relationship between concepts while preserving the roles (agent, patient, location, etc.) played by the words in a sentence. The most common approach to constructing the propositional representation is to apply a semantic parser that, using rules derived from text contained in a training set, label each word in a sentence with the role it plays in conveying the meaning. There are several difficulties with the approach:

- 1) The accuracy of the parser depends greatly on the quality of the pre-labeled training corpus;
- 2) The similarity of the content of the training and test corpora; and
- 3) The granularity of the labeling scheme used for words.

Dennis (2005) developed the Syntagmatic-Paradigmatic (SP) Model to address these shortfalls in compositional semantics. The SP model creates proposition-like units without using labeled training data; it derives them directly from a corpus of language in a process model of memory. The approach is attractive because it allows for the derivation of propositional structure without having to depend on a labeled training corpus. Also, because the representation of words in the model is a distributed pattern across several “roles,” a given word can be simultaneously assigned to several roles. Importantly, without a labeled training corpus, the model is capable of deriving propositional structure from languages for which there is no labeled training corpus. Dennis has shown that the SP model correctly predicts subject performance on tasks where syntactic structure is manipulated.

The loosely structured and automatic manner in which the SP model is able to derive propositional structure from language makes the SP model a good candidate model for natural language understanding in simulated operators. One criticism that has been made of the SP model is that, as an “instance model” (memory traces are stored separately), the time it takes to get an output from the model will increase with the amount of information in memory. Dennis has proposed a solution to this problem using a Locality Sensitive Hashing technique that returns a satisfying match that meets a tolerance instead of searching for exact matches in a massive memory system.

5.4.1.4 Prospective Memory

Colonel Schmidt makes a mental note to commend Lieutenant Anderson’s platoon on their professional response under difficult circumstances.

One’s memory to do something in the future is called prospective memory. Interest in prospective memory has increased in recent years, but there is little theoretical work done in the field. There are two competing theories of prospective memory that try to explain what is commonly referred to as “event-based” prospective memory or a memory to do something that is triggered or brought into awareness when a target event occurs in the environment, for example, remembering to mail a letter when driving past a post office. Event-based prospective memory is contrasted with time-based prospective memory where memory to do something is brought into awareness at a particular time.

The major factor differentiating the event-based theories is the extent to which attentional resources must be used up to perform the task. The Preparatory Attentional Memory (PAM: Smith & Bayen, 2004) proposes that attention is allocated to monitor the environment for the target and opportunities to do the intended actions. By contrast, the reflexive-association perspective (McDaniel & Einstein, 2000) proposes that a target event in the environment serves as a cue that elicits the automatic retrieval of the intended actions.

Both perspectives come with their own set of validation data from the laboratory. Hence, it is difficult to recommend one alternative over the other and it is likely that context plays a role in determining which phenomenon is appropriate. Specifically, if it is clear that having to remember to do something later interferes with the task at hand, implementing a prospective memory component based on the PAM approach in a simulated operator is a more fitting approach than one based on reflexive association.

5.4.1.5 Categorization

...he is interrupted with a message about a power outage in a remote neighborhood. This may indicate insurgent activity or be just another infrastructure glitch. Command staff report no other indicators of hostile activity on the situation displays, so he decides on the latter for now...

There are currently four different kinds of categorization tasks being studied in the psychological literature: rule-based tasks, prototype distortion tasks, information integration tasks, and weather prediction tasks. Neurological evidence suggests that different neural pathways are used, depending on the task – a notion consistent with the current situation that no single model can account for the data collected in the various tasks. The categorization literature is rich with models and paradigms, but no single model of categorization has been able to perform well on all four of these task categories. The different types of categorization that an operator must perform are likely context dependent, arguing for inclusion of a number of different models for different tasks. Most of the categorization models under study are computationally inexpensive, making inclusion of several approaches viable without sacrificing real time or faster than real time execution.

Prototype theories assume that for each category a single prototype is formed. The prototype is thought to represent the central tendency of a category, such as the arithmetic mean or the mode of each of the features. During categorization, stimuli are classified as belonging to the class with the closest matching prototype.

In exemplar models, all training stimuli and their associated category are explicitly remembered. During categorization, novel stimuli are compared to all known exemplars. Category membership can either be obtained from the best matching exemplar or the prominent response in a collection of K exemplars or can be assigned to the category for which the sum of the similarities is greatest. With respect to exemplar models of category learning, two of the most prominent models include Nosofsky's (1986) generalized context model and Kruschke's (1992) Alcov model.

Decision Bound theory assumes that subjects partition the input space into regions that respond to each category. When labeling novel stimuli, the category is determined from the region within which the stimuli fall. Examples of Decision Bound models include Ashby and Waldron's (1999) striatal pattern classifier, Anderson's (1991) rational model, and Love, Medin, and Gureckis' (2004) SUSTAIN (Supervised and Unsupervised STRatified Adaptive Incremental Network) model.

Although many models exist for modeling human performance on prototype distortion and information-integration tasks, modeling the decision making processes associated with rule-based tasks is far more complex. The most prominent model of human categorization that includes rule-based learning is the COVIS model (Competition between verbal and implicit systems, Ashby, Alfonso-Reese, Turken & Waldon, 1998).

Although COVIS provides an explanation of the different systems involved in different categorization tasks, mechanisms to account for complex rule-based categorization have not been explored. As stated, such a system would need to integrate several memory systems (procedural memory, working memory, semantic memory, and episodic memory), utilizing the same processes involved in other high-level perceptual tasks such as planning and analogy making. Prominent models of such high-level problem solving include SOAR (Laird, Newell & Rosenbloom, 1987; Laird, Rosenbloom & Newell, 1986; Newell, 1990), ACT-R (Anderson, 1991; Anderson, 1993; Anderson & Lebiere, 1998) and models by the Fluid Analogies Research Group (Hofstadter, 1995).

5.4.1.6 Learning

...The crowd seems very excited, although from training and experience, he recognizes that these behaviors are often associated with positive events in this culture and the atmosphere seems different from that when attacks were imminent.

A feature thought necessary for HBR robustness is adaptation; that is, the old concept of rigid rule bases or strict process is inadequate for many applications. Learning is the acquisition or modification of rules, facts or relationships (Thagard, 2005, pp. 48-50). Some knowledge may be innate, but much may also be obtained through inductive generalization, extrapolating from specific examples or by other deductive reasoning processes (such as chunking in SOAR or composition in ACT-R) where the deductive process creates a new rule linking the preconditions to the goal directly. Alternatively, existing rules may be modified to be more specific to handle special cases. A rule can gain in strength each time it is used successfully (much like adding traces of experiences in Minera2 memory models) so that unsuccessful rules fade (and could be pruned or "forgotten"). While there is considerable work in machine learning, using mechanisms that mimic human learning may result in behaviors that are closer to those observed in practice and for similar reasons.

SOAR and ACT-R have rules for creating rules from reasoning about problems that build on Knowledge Acquisition/Knowledge Engineering already in place. ACT-R's declarative memory with facts and semantic relations among facts is a useful feature and has empirical support. ACT-R's gain-cost utility function for selecting production as well as pattern-matching associative retrieval processes are useful, employing Bayesian updating and learning models to provide adaptive estimates for the values of each rule.

Episodic, exemplar, case-based learning models may be more useful than Bayesian models since analysts seldom have access to the joint probabilities and cues, or the data that are available may be fuzzy, unknown, or even contradictory. Neural networks may be very useful for approximating learning in noisy systems, such as perception, learning the following relationship coefficients, categorical, probabilistic relationships, etc. ART (Adaptive Resonance Theory) is an example of the use of neural nets in human models. Pew and Mavor (1998, pp. 145-147) provide aspects of ART that may be useful including the automatic development of productions (if-then, condition-action rules) from the learned weights using ARTMAP.

The above approaches are principally cognitive science representations of psychological processes, but there is room for classical parameter tuning of performance models where the acquisition of skill may not be the objective of the HBR modeling. Further, learning should not be considered only in the context of human-like learning. In order to make models less costly to create, they should have some mechanisms to "learn" at an accelerated rate to reduce reliance on manual creation of rules or processes. A learning mechanism should permit reducing knowledge to rules and rules to skills, with commensurate reductions in demands, possibly employing heuristics to achieve this end. There needs to be a rational choice of modeling techniques for HBR development to meet the application objectives.

5.4.2 Reasoning and Thinking

Should the situation degrade, immediate action could save lives, prompting Colonel Schmidt to deploy a fresh, fast response Special Ops Team to lead the rescue assigning Lieutenant Anderson's platoon to a supporting role, securing the perimeter from strong points.

Reasoning and thinking comprise much of the machinery for problem solving, planning, and replanning. People have evolved to function in a partially open system where uncertainty and even deception might be considered the norm in a competitive environment. Constructive simulation systems and even much of virtual simulation are potentially open to constructive agents operating in those synthetic environments; that is, most of the information can be made available to all agents. Human agents in such systems may not be equipped to deal with this volume of information because of the skills and heuristics developed to survive in the real world. If we want HBRs to act in a manner similar to human agents in these systems, we should have similar mechanisms of perception, reasoning and judgment or choice that mimic human mechanisms, considering a corresponding subset of the available information.

There are a number of AI approaches to reasoning, none of which is likely to be humanlike but could possibly be adapted by constraining them to specific reasoning styles or strategies that depend on tasks, context (including situation-specific conditions such as time pressure and consequence of errors) all moderated by human traits such as acceptance criteria, tendencies (consideration of all aspects vs. consideration of the first that looks possible), etc. The AI literature (Nilsson, 1998; Russell & Norvig, 2002) has a number of reasoning algorithms such as depth first, breadth first, and iterative deepening that might be applied to reflect current thinking on how people think and reason.

There is considerable evidence to support that people do not reason rationally using formal logic or even considering all the relevant information available in a situation (Thagard, 2005, pp. 34-35) and that formal

logic is only distantly related to human reasoning. Further, people consistently display biases that are not rational, such as systematically greater aversion to losses than attraction to gains. Instead, human reasoning and decision making tend to be content and context dependent rather than formally valid, although arguments using formal logic can be produced by people with sufficient effort. The use of logic as a norm for human rationality has been replaced in large measure by an ecological conception of rationality, based on adaptation to the environment. The most developed model of this kind is ACT-R, which is a comprehensive model that suggests how various aspects of cognition might be integrated.

It is clear that not all thinking processes are alike and it has been found necessary to distinguish between thinking that is rapid, parallel, automatic, and effortless but lacks flexibility and, in contrast, thinking that is slower, serial, strategic, effortful and flexible, in the sense that it can be adapted “online” to changing task demands. Halford et al. (2006) summarize a number of observations, particularly those of Evans, suggesting that a human reasoning model should be represented as a dual process. These types of thinking have been variously described as implicit/explicit, associative/rule-based, System 1/System 2, subsymbolic/symbolic, unconscious/conscious, evolutionarily early/ evolutionarily late, not associated/ associated with language and symbolic processes, and they also involve patterns of activation across different brain regions. These two systems would be combined to represent a broader range of phenomena, relying on complementary strengths to overcome weaknesses yet maintaining psychological plausibility. In this approach, System 1 is an automatic, low demand process that is associative (subsymbolic), fast, and not related to language; System 2 is a deliberate, effortful process that is rule-based (symbolic), largely serial, and related to language. This is similar in part to the philosophy adopted by ACT-R.

It may be difficult to distinguish between the systems in practice because the power and flexibility of System 2 methods can “represent” much of the associative, System 1 methods, but attributing System 1 phenomena to System 2 is likely to be inappropriate. Representation of phenomena should be through the simplest mechanism that captures the essence of the problem. In some instances, phenomena may actually be a combination of both types of systems working in parallel, interacting and overriding one another under certain conditions, but individual differences on System 1 may be quite different from individual differences on System 2 (Halford et al., 2006, p. 39). Because System 2 is capacity limited, System 1 may be useful in condition of high processing load, providing heuristic solutions that will be correct most of the time, and the interaction of the two systems seem to reflect many of the phenomena of human reasoning process outcomes, with System 1 pre-processing information that may then be used by System 2.

Thagard (2005, p. 51) states that of all the computational-representational approaches, rule-based systems currently have the most psychological applications. Much of human knowledge is naturally described in terms of rules, and many kinds of thinking such as planning can be modeled by rule-based systems. Rules can be applied to reason forward from a state to a goal or backward, from a goal toward the current state (or a combination of the two).

Yet most production rule systems fail to adequately represent human reasoning because:

- 1) They are restricted to simple forward chaining and do not support inductive or abductive, diagnostic reasoning;
- 2) They do not consider uncertainty in events, situations, information, and the rules linking them; and
- 3) They do not make use of memory but simply reflect the current instantaneous event state.

More advanced production rule systems may be able to overcome these limitations. Many if not most rule-based/production-system expert systems use a condition-action rule set; few if any build an implication

rule set (condition-implication) first that then supports the response choice (implication-action). This latter condition-implication-action better supports more advanced reasoning capabilities while the former reflects more skill-based responsiveness that can be developed as by SOAR chunking. Ignoring the requirement for strict matches with the antecedent, the condition-action approach provides an opportunity to create errors of commission or omission because of missed or incorrectly perceived information.

It is thought that much of our social knowledge can be represented as typical sequences of events or can be represented by their typical characteristics. Representing knowledge of concepts has been implemented as frames, schemas, and scripts. Concepts are knowledge representations that capture the typical characteristics of objects or events rather than reflecting strict definitions. Concepts can go beyond rules to organize knowledge into hierarchies that can impart additional knowledge through inheritance rather than explicit representation, which is a powerful tool for inferencing and to describe relationships, indicating that things are part of other things. Used with rules, concept architectures provide a mechanism for representing considerable knowledge, focusing on what rules to apply in certain situations. Using concepts requires two crucial steps: matching and inference. A specific situation is assessed and matched against concepts. The concepts that match, whether partially or completely, can cause other concepts to come into play through an activation spreading process. Those concepts that become activated and seem to match the current situation are selected and the system makes inferences about the situation by inheritance based on these concepts.

Analogy has been acknowledged as an important component of intelligence and thus should be a fundamental process to be represented in natural reasoning. Analogy is effective at modeling similar tasks with a common approach but it is particularly important when capturing transfer between domains – taking knowledge of one domain into another. Analogical thinking, also called case-based reasoning (CBR) in the AI field when applied to single-domain problem solving, involves adapting something known to a new situation because of some similarity between the two. An analogy is a structural correspondence between two cognitive representations, a base/source and a target. Analogs are like concepts in that they bundle together packages of information, but they are like logical statements in that they contain specific instances or situations; analogical schemas go further to include more general information. When working in very familiar domain, general knowledge can be captured in rules and concepts; analogical reasoning becomes useful when you have some experience in a domain but little general knowledge, when rules and conceptual knowledge are scarce.

Typically analogical reasoning follows a process where, when faced by a new problem, you remember a similar problem for which you have a solution and you adapt the old solution to the new case. The process requires procedures for memory retrieval, comparison, and then adaptation. Thagard and Holyoak (Thagard, 2005, p. 80) maintain that practical analogical reasoning occurs in three parallel assessment processes: similarity, structure, and purpose. Similarity occurs at a superficial level; structure is a relationship or correspondence at a lower level; purpose can be made part of the probe to determine similarity, but some believe that purpose warrants a higher weighting in analogical reasoning. While analogies can be expedient, they do not always present the best solution and may lack deep, relevant similarity necessary for the new problem. If a problem is new, then no other analogue may exist and this process may be misleading (p 83: “In military planning, generals often fight the last war, using outmoded analogs.”). COA decisions are often made analogically. Analogues can be used to both support and contradict solutions; people often fixate on one set that supports their particular objective and ignore other relevant analogues that could provide alternative solutions or even contradict the selected solution.

Analogical thinking involves three types of learning. The first is simply the storage of cases based on previous experiences. The second is storage of a subsequent case solved using analogous situations, etc. The third type of learning is the development of an analogical schema that abstracts the common elements from the source

and target problems and creates a simplified concept, although a concept that is specific since it is based only on two cases, but one that might contain rules that although rough, may provide a partial solution to some future problem.

A number of computational models of analogy have been developed, many based around ANN such as ACME, COPYCAT, STAR, LISA (Hummel & Holyoak, 2005), but others such as Dennis' (2005) Syntagmatic-Paradigmatic (SP) model of Natural Language Processing (NLP) and analogical reasoning are based on Minerva2 instance-based model of memory.

The traditional assumption that human reasoning is inherently rational and logical was severely challenged in the twentieth century. While there are deductive inference rules that are applied in natural reasoning (e.g., *modus ponens*), there is little evidence that natural reasoning normally follows rules of deductive inference. People's conditional reasoning is not monotonic, and formally valid arguments are sometimes rejected if additional premises cast doubt on the truth of the conclusion in the real world. Interpretation of logical connectives also departs from standard logic. Common natural interpretations of the conditional do not conform to the truth functional definition, the counter-factual cases being seen as irrelevant. Interpretation varies according to context so "if then" might be interpreted as "if" or as "iff" (if and only if), depending on context, and "or" is sometimes interpreted inclusively and sometimes exclusively. Reasoning is also heavily influenced by content, one of the most notable examples being that performance in the Wason Selection Task¹ is much better when the task is framed in terms of permission rules rather than abstract rules. Reasoning also depends to a considerable extent on knowledge of the world or of the domain, so retrieval from memory is likely to be a more prominent process in reasoning than application of logical rules.

This suggests that natural reasoning is more akin to induction rather than to deduction because reasoning seems to be concerned more with predicting what is likely or plausible in the world, given a particular set of circumstances, than about formally valid inferences. Consequently, soundness rather than deductive validity is the main criterion that constrains human thinking. This is consistent with the philosophy that the function of thinking is to aid adaptation to the environment; in essence, humans are surviving machines rather than computing machines and HBR reasoning systems should reflect this phenomenon.

Theories of reasoning based on heuristics have partly replaced theories of reasoning based on logical rules of inferences. Heuristics include availability and representativeness, "take-the-best" strategies, or the configural and atmosphere heuristics in categorical syllogisms. Economic heuristics, such as being more averse to losses than attracted to gains, have also been shown to be important. Probabilistic heuristics represent premises as approximate probabilities, so conditionals should be interpreted as conditional probabilities. Thus, $p \rightarrow q^2$ can be interpreted as $P(q | p)^3$ is high. Pragmatic reasoning schemas are of general validity and are induced from life experience. Examples include permission and obligation. Although pragmatic reasoning schemas are often interpreted as deontic, there seems to be no reason why any schema induced from life experience could not be used as an analog for reasoning if it matches the structure of the task. Thus, pragmatic reasoning schemas should probably be seen as a sophisticated case of analogical reasoning.

Neural networks (connectionist models) have many psychologically plausible features; they learn, they generalize automatically, and they degrade gracefully as input degrades. They do not require top-down

¹ See <http://coglab.wadsworth.com/experiments/WasonSelection.shtml> for further information.

² $p \rightarrow q$ may be read as: condition "p" implies result "q."

³ $P(q | p)$ may be read as the conditional "probability of result q, given the occurrence of condition p exists."

control, and they form prototypic categories automatically. These properties capture some of the fundamentals of cognition. However, they have been criticized for lacking properties such as compositionality and systematicity that are essential to thought and language. This has led to the development of symbolic connectionist models that are more successful in capturing the properties of higher cognition. The distinction between classical and symbolic connectionist models corresponds in many respects to the distinction between System-1 and System-2 processes, and neural net modeling gives considerable insight into these two levels of process. Some of the most promising contemporary models of thinking are a hybrid of classical neural networks and symbolic or rule-based nets.

Mental Model theory (Markman & Gentner, 2001, pp. 228-235) is an influential although somewhat controversial approach to human reasoning. Mental Models are a means of formalizing knowledge but in an incomplete manner that seems to produce effects that reflect human deduction. The criticisms are that it entails the logic that it rejects, adopting a logic based on semantics rather than syntax, and it is not sufficiently well specified to be testable. Nevertheless, Mental Models provide a mechanism that reproduces some of the observed non-rational effects, particularly in more complex situations, producing illusory inferences. Mental simulation is a way of using Mental Models, often involving imagery, with qualitative assessments that are closely tied to motor actions.

Mental Models do not typically represent the full set of logical possibilities consistent with the premise. In Mental Model theory, the complexity of reasoning depends on the number of alternative models that must be constructed to validate the conclusion. It has been proposed that sophisticated reasoners will not accept inferences that follow from the first model constructed but will seek alternative models; only inferences that are consistent with all models are then held to be logically valid. Naïve reasoners fail to consider alternative models and consequently are prone to concluding that inferences that are possible, but not necessary, are logically valid. Part of the model construction process is the attempt to construct counter-examples, that is, possibilities that are consistent with the premises but inconsistent with the conclusion.

Mental Models for deduction can be based on analogy. A Mental Model is iconic, with some structural correspondence between the model and the content of the problem. A structural correspondence is the defining property of analogies; analogy seems to lie at the heart of mental models. Much is known about how analogs are selected or retrieved from memory and Mental Models could be instantiated as instances in a memory system such as Minerva2. This would tie a reasoning framework to a psychologically plausible memory system that could be populated with an initial knowledge state, create plausible decisions and errors, but learn through experience.

5.4.3 Judgment and Choice

...he decides on a detour so as to avoid any appearance of aggression, leading the platoon on towards the medical center by an alternate route that had been identified for such contingencies.

A complication in the decision making process arises when multiple solutions result from the reasoning process. Similarly, when multiple relevant rules occur at a decision point, the number of combinations and paths can explode as the plan is elaborated (such as the number of moves in chess) beyond the capacity to evaluate them all in the time available. People appear to make judgments and choices that depend on context, expectation, personal traits and states, etc., and seem able to make decisions despite uncertainty and conflicting evidence. To resolve this problem, HBRs require some method of selecting among alternative options.

Utility theories are commonly used for complex decision making models, making decisions that maximize long-run average, Expected Value although Maximum Expected Value has not proved to reflect how people

make decisions. Most people appear to be risk averse and prefer to incur extra short-term expense to avoid catastrophic losses that, when summed over a significant period, exceed the expected loss. Expected Utility was proposed to avoid the shortcomings of Expected Value, where “Utility is a nonlinear transformation of a physical measurement or value” (Pew & Mavor, 1998, p. 153). The shape of the utility function represents an individual’s attitude toward risk aversion, but this theory too has experienced many exceptions.

It has been found in practice that when probabilities among options are approximately even, the difference in gains overwhelms the probability differences resulting in a choice for maximum gain; however, when the probabilities change with a small but certain loss versus a probable, significant gain, then probability dominates over the differences in gains. Kahneman and Tversky (1979) call this the “certainty effect.” There are a number of other utility measures that have been developed to address the shortcoming of the basic Expected Utility approach including rank-dependent utility, multi-attribute utility, game theory, random utility, sequential sampling, and adaptive planning.

Signal detection theory (Green & Swets, 1967) is an approach that has been used to describe discrimination among choices, primarily perceptual difference judgments, but difficulties arise in predicting error rates and latencies when it is coupled with sampling over time (Vickers, 1970). This led to the development or resurgence of a number of models of evidence accrual such as random walk (Heath, 1984) and Vickers’ (1970) accumulator models that seemed to capture some of the decision latency dependence on discrimination as well as the error rate while reflecting incidence and magnitude effects in the choice.

Gigerenzer and Goldstein (1996) reject the position that decisions are made by considering all or even a large portion of the total information that is available, at least for conscious reasoning. They argue that humans seldom have the time (due to a competitive environment) or the will to invest the mental capacity to integrate all the information available on a daily basis. Instead, they propose that their Fast-and-Frugal/Take-The-Best approach model is a more appropriate characterization of human decision making. In this model, only the most salient cue that differentiates among options is used to form a judgment. In their research, they have been able to reproduce several decision making phenomena with this theory. Recently, there has been an attempt to unify the “Take-the-best” with “substantively rational” strategies, treating each as a special case (limiting conditions) of the unified model (Lee & Cummins, 2004).

All of these theories can incorporate individual differences into the judgment process including factors such as: risk aversion, pessimistic versus optimistic, aggressive versus passive, rational versus irrational, impulsive versus deliberative, expert versus novices.

5.4.4 Belief and Confidence

...he is interrupted with a message about a power outage in a remote neighborhood. This may indicate insurgent activity or be just another infrastructure glitch. Command staff report no other indicators of hostile activity on the situation displays, so he decides on the latter for now...

Belief in the state of the world and confidence in a decision are important factors for models that must accommodate uncertainty. Belief in this context is the accumulation of evidence by sequential sampling or by reasoning to form a judgment (i.e., make a decision). Confidence is the degree to which the judgment is thought to be certain. Any model that deals with uncertainty must therefore take both belief and confidence into consideration.

Baranski and Petrusic (1998; 2003) have investigated the formation of confidence in decisions, building on a long history of research; a brief overview of the field is presented in the first reference. Baranski and Petrusic

have created a formal model of confidence, the “Doubt-Scaling Model,” that seems to capture more of the empirical results concerning confidence estimates than other representations in models such as Vickers’ Balance-of-Evidence or Audley’s “Runs” models. In many instances, the model of confidence estimation can be separated from but takes input from an accumulation of evidence model so that various approaches can be combined as required.

Baranski and Petrusic have discovered that different processes seem to come into play when forming a judgment of confidence or belief, depending on whether the subject focuses on the speed or the accuracy of judgment. During time stress, estimates of confidence were thought to be made after the decision in a serial cognitive task that allows an accurate diagnosis of decision errors. When a decision is made with an accuracy emphasis, confidence seems to be built in parallel with the accumulation of evidence and judgment. This allows the use of adaptively regulating parameters in the decision making process but yields poorer error detection after a mistake has been made.

Belief and confidence models can be used to moderate the decision making process, causing the reasoner to seek additional information, form judgments too early (ignore information), or explore alternative solutions. Decision times and error rates can thus be moderated by varying parameters within the belief and confidence models. While no formal representation of these individual differences currently exists to link decision making styles or personality traits to performance (Bruyn-Martin, Bandali & Lamoureux, 2006). Sufficient guidance does exist to produce coarse categorization in this regard. These models could be linked to other performance shaping factors such as time stress, fatigue, etc., or context characteristics such as consequences of error and outcome importance, to moderate the reasoning process of an operator model in an attempt to better reproduce human-like behavior.

5.5 OBSERVATIONS

The current state of human sciences knowledge about cognition and computer technology does not support the development of a single, formal, human behavioral model suitable for all aspects of military simulation. Knowledge and formal representation of all the important aspects of human cognition or performance relevant to military simulation do not yet exist. Yet there are cognitive models and theories that we can begin to introduce into current technologies that address these shortfalls and are supported by cognitive science.

A feature thought necessary for robustness is adaptation; that is, the old concept of rigid rule bases or strict process is inadequate. There is a need for the CGF to dynamically redefine its own rules using mechanisms that mimic human learning. Further, it is likely that human problem solving is not strictly rational or even ecologically rational so that more diverse decision making schemes need to be considered in the development of HBRs.

Representation of knowledge as formal propositional logic or if-then rules may be suitable for some applications, however, for more complex applications, particularly with human in the loop participation, more human-like cognitive representations at various stages of processing are required. It has become clear that analogy is a basic process in human reasoning, which is arguably more analogical than logical. This implies that reasoning occurs by mapping components of a problem, such as premises in a deductive inference task, into an integrated representation of the relevant relations. Analogical inference has considerable flexibility and power, but inferences are not necessarily deductively valid and require verification by the HBR, leading to errors of commission or omission.

COGNITION IN HUMAN BEHAVIOR REPRESENTATION

Because thinking employs more than one type of representation, strategy or process, hybrid systems in which a number of processes are harmonized to optimize performance, offer considerable benefit. The higher level phenomena may be represented by means other than these low level models, such as AI models; however, the analyst has to weight the importance of the representation to the application objectives against pragmatic questions such as development cost and time, execution speed, and robustness of the resulting HBR. There needs to be a rational choice of modeling techniques to meet the application objectives, but the modeling approach should afford the opportunity to replace models and capabilities with better or more appropriate versions that can work with the accumulated knowledge representation, as well as to forget faulty knowledge.

A number of observations and suggestions follow for future HBR approaches:

- 1) Development of CGF using HBR needs to take account of the way conceptions of human thought have changed fundamentally in the last few decades. There is a move from being logic-based to being based on heuristics, mental models, verbal reasoning, analogy, and retrieval of information from memory. An ecological conception of rationality, defined by adaptation to the environment, has partly replaced norms based on logical inference rules. This implies a move from the traditional AI rational approaches but does not preclude their use in a human-sciences based context.
- 2) Contemporary cognitive models have gone a long way toward taking account of content and context effects. Future work should try to build on these advances to extend the capabilities of HBR and improve their robustness.
- 3) A consensus is emerging that the brain contains many specialized modules but also that a central integrative function is essential, such as that provided in ACT-R (Anderson et al., 2004).
- 4) Two levels of reasoning have been identified in many contexts, and models should recognize that more than one type of reasoning process occurs. Models that integrate these levels offer the potential for capturing the flexibility and robustness of thinking.
- 5) Metacognitive, or executive processes, are probably important to flexibility of thinking because they enable representations and strategies to be adapted to task demands and to feedback about success and failure. Although theory in this area is less developed than for some other conceptions, there are signs of accelerating work in this area, and it should be closely watched in the future.
- 6) Human reasoning is knowledge based and is more nonmonotonic than monotonic. This implies that induction is a more natural human reasoning process than deduction.
- 7) Memory retrieval of acquired information about the world plays a major role in cognition. Models that include induction and memory retrieval will be more likely to capture human reasoning processes.

5.6 ADDITIONAL READING

The following references may not be included in the document reference section since they may not be cited directly in the text. Nevertheless, these papers represent a wealth of information that may be useful to the reader interested in formal memory models.

Some relevant papers on models of short term working memory include:

- [1] Lewandowsky, S. and Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, 96 (1), 25-57.

- [2] Page, M. P. A. and Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105 (4), 761-781.
- [3] Brown, G. D. A., Preece, T., and Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107, 127-181.
- [4] Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, 36, 73-137.
- [5] Anderson, J. R. and Matessa, M. (1997). A production system theory of serial memory. *Psychological Review*, 104, 728-748.
- [6] Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 268-299.

Some relevant articles on models of semantic memory include:

- [7] Collins, M. and Quillian, M. R. (1969). Retrieval from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.
- [8] Dennis, S. (2005). A memory-based theory of verbal cognition. *Cognitive Science*, 29 (2), 145-193.
- [9] Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- [10] Loftus, E. (1975). A spreading activation theory of semantic memory. *Psychological Review*, 82, 407-428.

The following articles describe computational models of episodic memory:

- [11] Dennis, S. and Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, 108 (2), 452-478.
- [12] Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review*, 89 (6), 627-661.
- [13] Hintzman, D. L. (1984). Minerva-2 – a simulation-model of human-memory. *Behavior Research Methods Instruments & Computers*, 16 (2), 96-101.
- [14] Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89 (6), 609-626.
- [15] Shiffrin, R. M. and Steyvers, M. (1997). Model for recognition memory: REM – retrieving effectively from memory. *Psychonomic Bulletin & Review*, 4 (2), 145-166.

The following articles describe formal models of prospective memory:

- [16] Smith, R. E. and Bayen, U. J. (2004). A multinomial model of prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 756-777.

- [17] Einstein, G. O. and McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 717-726.
- [18] McDaniel, M. A. (1995). Prospective memory: Progress and process. *Psychology of Learning and Motivation*, 33, 191-221.

Chapter 6 – SOURCES OF VARIATION IN HUMAN COGNITION, BEHAVIOR AND PERFORMANCE

Colonel Schmidt is concerned about his attention to detail since he has now been without sleep for 20 hours. The situation update display, being tuned to recognize possible performance decrements, implements an increased level of decision aiding automation...

The platoon has been exposed to the effects of extreme vehicle vibration and now is suffering from the heat of day. In order to ensure their effectiveness on this next mission, the platoon takes on additional water and specially packaged nutrition supplements; their information displays are reconfigured automatically for better readability; and their micro-cooling is enabled...

Even before arriving in-country, Colonel Schmidt and Lieutenant Anderson had trained together even though their home units were in different NATO countries. Their operating styles as NATO partners and their shared understanding about the common mission – including both support and combat missions – grew out of distributed team training with simulations...

6.1 INTRODUCTION

Human performance differs from the performance of physical components in a system in that it changes with well-known effects such as fatigue or other environmental stress and tends to vary within individuals from occasion to occasion in a stochastic manner. Different individuals perform and behave differently in the same context so the full spectrum of human variability involves both inter- and intra-individual variability. The performance of computerized agents that use pre-determined unchanging behaviors is overly predictable and uncharacteristic of human agents since it fails to capture the rich variability that is observed in the full spectrum of human behavior. Regardless whether the purpose of constructive modeling is to assess the performance of future systems, test future tactics and doctrine, or provide a training environment, it is important that the conclusions drawn from any analysis are robust against reasonable variations in the human behaviors and that a training environment provides an appropriate challenge to the trainee. There is a place for deterministic “best” or “worst” behaviors to provide specific insight into key aspects of a system or to support specific educational objectives, but even if these are to be employed for at least part of the time, robustness of the conclusions to realistic variation remains of overriding importance. If a realistic range of behavior variation is to be captured, the influences that determine the pattern must be understood so that all the components can be identified and put together in a single modeling framework.

6.2 MODERATORS

The term moderator is sometimes used in connection with the variation of performance. There is confusion as to the interpretation of this term. Sometimes traits are considered moderators, affecting the performance in a generic way, without a specific underlying mechanism. Others regard state variables as moderators, moderating performance in a scientifically understood way. To avoid confusion, we consider any variable that changes the performance as a moderator but handle more specific terms to explain what we exactly mean.

The sources of variation can be divided into three groups: those that lead to intra-individual variation, those that determine inter-individual variation, and those that affect organizational behavior. The first group comprises the effects of external moderators (stressors) and stochastic variation from occasion to occasion.

The second group comprises the internal moderators and unexplained stochastic between-individual variation. It might be argued that, in principle, all between-individual variation should be explained by a set of internal moderators, and the stochastic component can be eliminated. On current evidence, a complete model of this kind is not yet available and the appropriate model comprises a systematic element and a stochastic element. The third group comprises the determinants of inter-individual interaction within an organizational, team, or group context that are not captured by the first two groups and include the components of organizational culture and possibly organizational and team processes.

It is not the purpose of the present document to detail all the sources of evidence that can be used to develop models of behavior since the field is substantial and beyond the scope of the present study. The objective of the present document is to identify a framework in which the behavioral models can be placed and to provide an indication of how models can be put together using some examples. In principle, operator state can be expressed as a list of variables that are necessary and sufficient to describe the performance in all activities in the analysis. The purpose of this study is also to provide practical guidance on the state variables, in particular on relevant, measurable state variables. State variables cover a wide range of HF, including perceptual, physiological, cognitive, emotional, and motor aspects. State variables may interact, increasing the complexity.

There is a useful distinction that can be made between the representation of behavior and performance. In the discussion of the modeling of the effect of moderators, behavior is defined as the choice of a COA that meets a goal, while performance is the description of how well that COA is executed. Since the representation of these characteristics of human behavior in a model is different, the way in which moderators are incorporated in a model differs between these two elements. The construction of a framework in which the effects of the three groups of moderators on both behavior and performance can be brought together is outlined in the following sections. As a first step, some moderators that are important in military simulations are defined and the general principles of how data can be assembled to describe the effects are outlined in Section 6.2. In Section 6.3 the impact of external moderators on performance and behavior is outlined and methods for representing the effects in models are described. In Section 6.4, the internal moderators are discussed and the problems of representing inter-individual variability are outlined. In Section 6.11 of this chapter, the representation of the organizational moderators is considered.

6.3 MODERATORS IN MILITARY SIMULATIONS

6.3.1 Important Moderators

As stated in the introduction, the moderators can be allocated to three groups: external moderators (stressors), internal moderators (personal attributes), or team and organizational moderators. The first stage in the development of a model of the effects of the moderators is to identify a list appropriate to military simulations. This list has to be justified by consideration of operational importance and relevance to the applications that we need to consider.

The military context is characterized by the need to perform in environments that are more extreme than those regularly encountered in industrial settings and tends to place more extreme demands on the individuals participating in the operation in terms of fatigue and workload. As a result there are a number of environmental stressors that have been studied in detail as part of military research over the last 50 years that are relevant to a wide range of military operations but are less important in the management of related operations in the civil domain.

Analysis of the requirements in military operations suggests that several stressors are almost always important – both operationally and for training applications – and should be modeled, while the others have less frequent application. The proposed list is shown in Table 6-1.

Table 6-1: Environmental Stressors Important for Military Simulations

Key Environmental Stressors	Occasional Environmental Stressors
<ul style="list-style-type: none"> • Sleep deprivation • Rapid time zone shift – circadian effects • Sustained physical demand • Thermal effects (causing thermal strain, dehydration, discomfort) • Visual environment • Task demand – taskload 	<ul style="list-style-type: none"> • Noise (continuous and impulse) • Vibration • Hypoxia (Loss of oxygen in high flying fast jets and work at altitude) • Acceleration: High G for fast jets; alternating G (Push-Pull) for all aircraft • Vestibular effects

The requirements for modeling internal moderators are less well defined than for environmental stressors because many of these variables are mental constructs that cannot be measured directly or objectively in a manner suitable for formal models. There has been less research on the effects of the moderators on performance; thus, there is less certainty about a complete list. Collective performance is influenced by moderators that apply to the team as opposed to the individuals within the team. In this context, training and experience refer to collective training and experience of performance in the team. Table 6-2 shows a proposed minimum list of moderators, based on some background material for both personal and collective aspects that are thought to be important to represent in military simulations.

Table 6-2: Characteristics of Individuals and Groups Important for Military Simulations

Personal Characteristics	Collective Characteristics
<ul style="list-style-type: none"> • Training (both task and physical) • Experience (on task and with equipment) • Age • Personality, coping style and culture • General intelligence • Anthropometry • Fear, Anxiety, Morale 	<ul style="list-style-type: none"> • Training • Experience (with teamwork) • Team composition (Ad-hoc vs. established) • Cohesion • Leadership • Culture and Organization • Language

6.4 CHARACTERIZING THE EFFECTS OF MODERATORS

There are three steps to identifying the effects of moderators on performance and behavior:

- Identify the relationships to be investigated;
- Assemble data and models from appropriate sources; and
- Parameterize the selected relationships.

The large number of relationships that is involved, even in a relatively simple generic model, precludes the development of a simple experimental program from which all results can be drawn. In practice, models of this kind have to be constructed based on a combination of meta-analysis of results assembled from the scientific literature and constructive models where it is feasible. The use of meta-analysis to estimate the values of unknown parameters is not simple. There are frequently inconsistencies in experimental methods, there are variations in the sampled population, and the manner in which the results have been derived may differ from that required to construct the selected model. These are well-known issues in meta-analysis that impose discipline on the analyst, but they can be managed, and with care in interpretation, robust results can be deduced.

The potential complexity can be illuminated by considering a relatively simple example of the potential representation of external moderators on behavior. Figure 6-1 represents a hypothetical, conceptual model of perception.

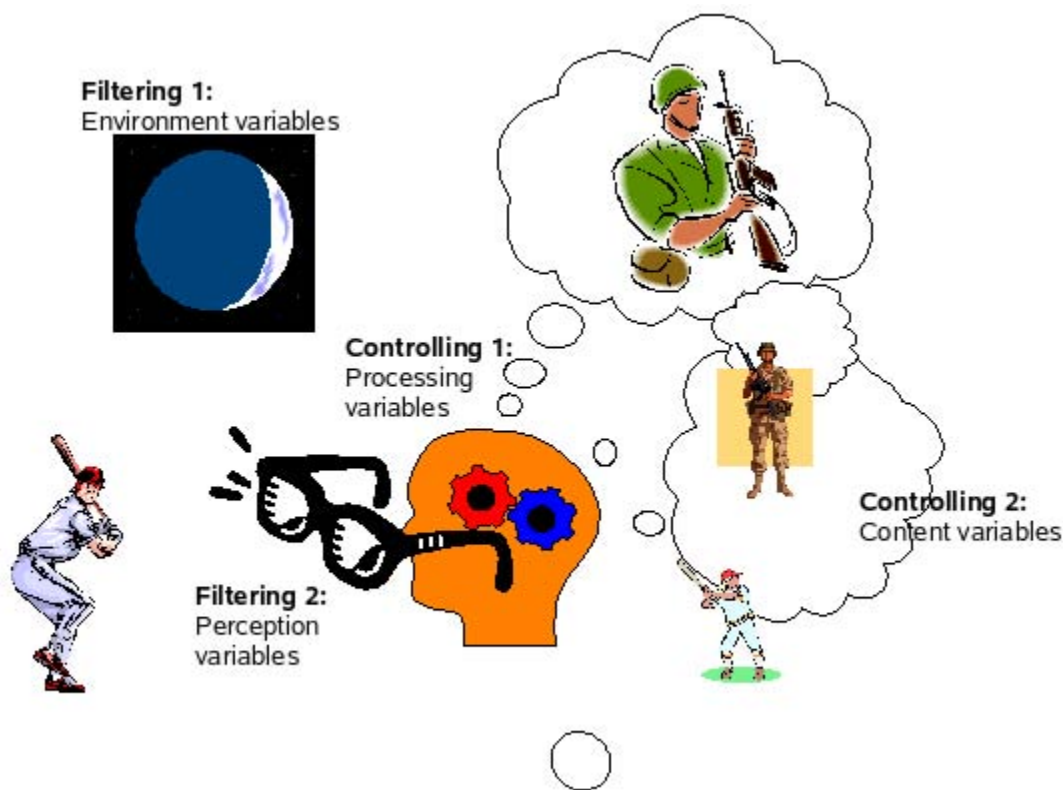


Figure 6-1: Conceptual Model of Perception.

This model assumes that we assess the situation through a perceptive filter before processing the filtered information in a controlled process path. Consequently, each aspect of the process can be moderated. For example, in Figure 6-1, the operator perceives a soldier when looking at a baseball player. It could be explained by:

- Environment variable: Low value representing nighttime;
- Perception variable: Low value representing impaired visual equipment;
- Processing variable: Low value representing slow working memory speed; and
- Content variable: High value representing associative bonding with soldier features and low value representing bonding with baseball player features.

The key points are that the complexity of our moderator model is a function of the processes we wish to represent and the behaviors we wish to capture, as well as that multiple moderator components are likely to be present even in a relatively simple case.

As a potential simplification it should be noted that if the chosen model is represented by more than one relationship, the parameters for the two components do not have to be estimated from the same source if it is assumed that the relationships are generic. For example, suppose that the model is a cascade from moderator A to performance measure C but there is an intervening variable B. We may express the model as the determination of the value of B given the value of A, and the subsequent determination of C, given the value of B. The data we use to understand the conditional relationship of B with A do not need to coincide with the data that are used to understand the conditional relationship of C with B.

This general approach can be applied successfully where there is a relatively rich literature embodying experimental findings. It is more difficult for those areas that have not been the subject of detailed research, or the findings have not been reported in a form that can be used to populate a model. For these areas, there are two possible approaches that could be employed: a more general model might be developed which covers a number of cases for which sufficient data can be assembled; a targeted experimental program can be conducted.

6.5 REPRESENTATION OF EXTERNAL MODERATORS

6.5.1 Performance Metrics

Before the effect of environmental stress can be defined, it is necessary to define the operator activities and specify metrics through which performance can be quantified. In experimental work, a task has customarily been defined as a unit of operator activity with an observable output, and this definition has generally been retained in task network modeling. The metrics of task performance have then been defined usually as the time taken to undertake the specified task and the accuracy with which the task has been executed.

This general approach has been employed in the modeling of human performance by identifying the “degradation factor” associated with the particular stressor and applying it to the time taken to do the task. If suitable data are available, a similar approach has been used to estimate the effect on error, although the latter has proved more difficult in practice.

6.5.2 Operator State as a Predictor of Performance

In this paper, the interaction between a human agent and the environment is assumed to be moderated by the state of an entity (Figure 6-2). Stressors (*properties of the operational environment or tasking that offset state variables from their neutral value*) change the state in a dynamical way, and traits (*characteristics of the entity that remain constant over the operation*) affect the state in a static way. The activities are performed better or worse, depending on the state of the entity. As an example, the activity marching with pack (stressor) will exhaust the entity, but physical training (trait) moderates exhaustion. Exhaustion is then a state variable (changing over time) that degrades the marching performance.

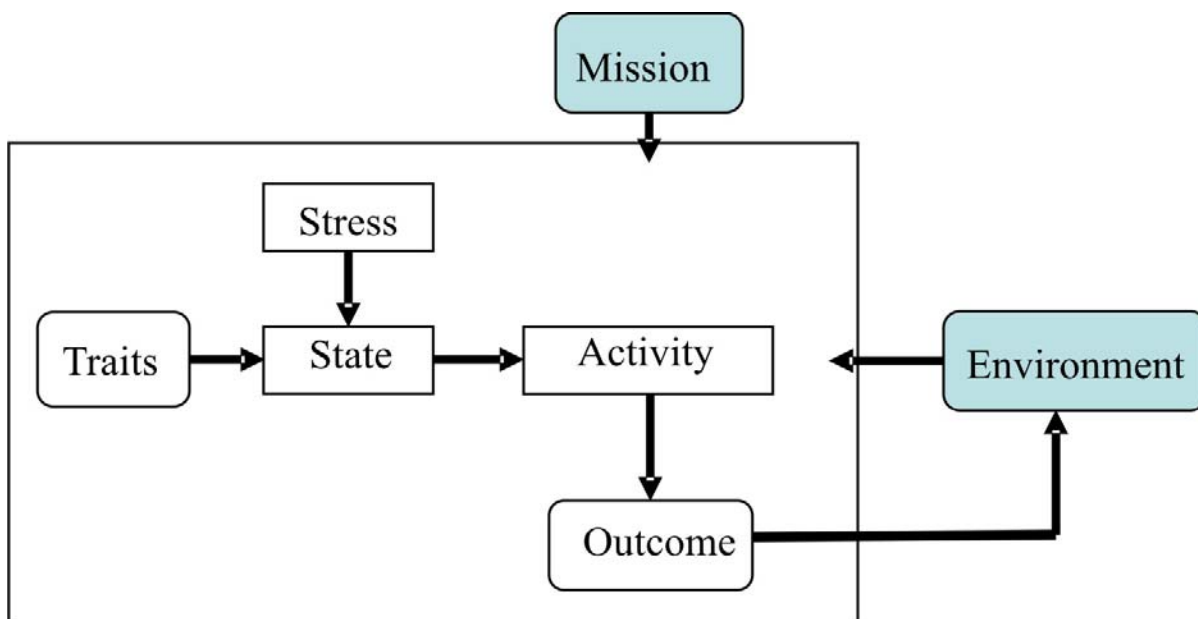


Figure 6-2: The HFM-128 Philosophy Approach to Modeling Operator Behavior.

This approach assumes that the way activities are carried out depends on the state of the actor, resulting from his personal traits and stressors emerging from the activity itself. The state varies over time and the state variables are key to the HF effect on performance.

In practice, the stressors in the primary list can be divided into two groups: those arising directly from the environment such as heat, noise, vibration, and those arising from the context of the task such as training, sleep deprivation, physical exertion or fear. The first group represents a direct change to the environmental conditions, which can be manipulated in a controlled manner in the experimental context. The second group represents stressors that can be managed in a similarly controlled manner, although their performance effects are clearly mediated by a change in the state of the experimental participant.

Any description of the effect of the second group of stressors has at least to recognize the change in the state of the operator implicit in the exposure to the stressful condition. In this case, the sequence of cause and effect between environment change and performance change can be represented in the form shown in Figure 6-3.

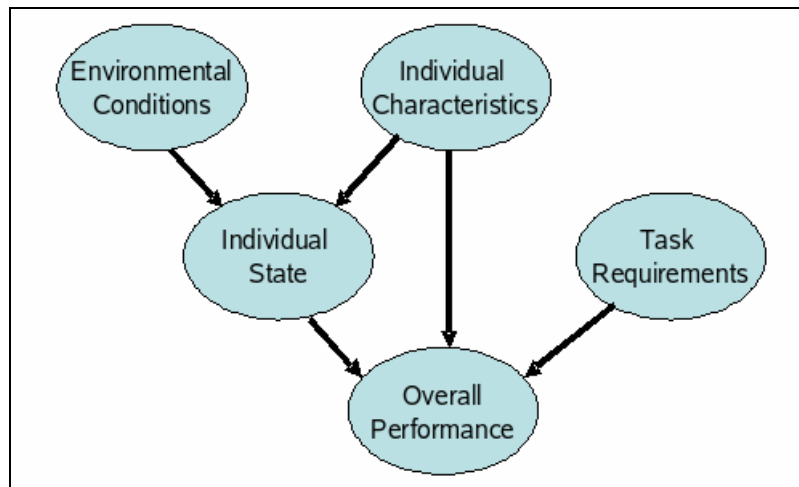


Figure 6-3: Conceptual Flow of Variable Influence from Environment to Performance.

6.5.3 Use of Performance Shaping Functions

If there is a known relationship between state and task performance, we call this a performance shaping function (PSF) and this highlights two issues. The first is the interoperability between HF sub-models and performance models. The output from the first is the input for the second and a perfect match is required to make the model work seamlessly (Figure 6-4, right frame). The second issue is that state has particularly an added value for time-dependent processes. For some state measures, such as body temperature, the state works as an integrator for conditions that vary over time. In addition, many human physiological or psychological responses are homeostatic, meaning that deviations from a standard condition occur and that the condition returns to the standard condition in a controlled way, often as a result of adjusting behavior: getting tired and resting, burning food and eating, getting warm and cooling, getting stressed and relaxing, etc. A model based on states and PSFs explains a great deal of variance in performance that would typically be attributed to random error variance, providing greater understanding of the problem under study and the means to address it.

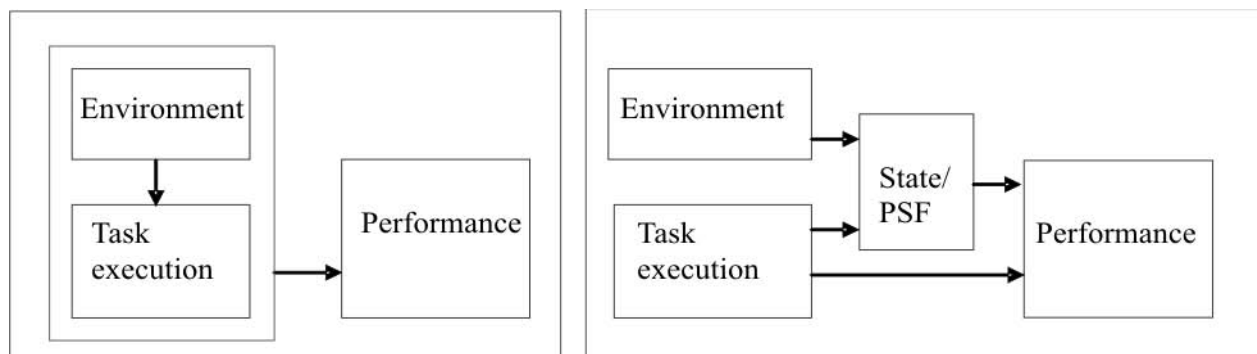


Figure 6-4: Task Execution Leads to Performance.

In the left frame the environment has a modifying effect on performance, which is heuristically established. In the right frame, the modifying effect is made explicit through PSFs. These functions are based on known relationships between states and performance.

Without the use of PSF, heuristic relationships must be established between task, environmental context and performance (Figure 6-4, right frame). The performance on a task may, for instance, be described both for fatigues and chemical, biological, radiological, and nuclear (CBRN) protective clothing in a 2-way table (environment by clothing). Since the number of variables involved is potentially large, multidimensional tables may result and it is then more efficient to use states.

For example, body core temperature is regarded as a good indicator of thermal state since it is a consequence of external conditions, work rate, and clothing. Core temperature appears to have an effect on cognition. Slightly raised temperatures increase arousal level, improving attention. Cooler or warmer humans tend to perform less well. Core temperature is apparently an intermediate parameter (PSF) between thermal conditions and mental performance. Core temperature is also a measurable parameter suitable for evaluation. In the psychological literature, the effect of the thermal environment on cognition has been studied for a long time by looking at the environmental temperature, disregarding the fact that the body only slowly follows the environmental temperature. Consequently, measurements have been taken at equal environmental temperature but at various core temperatures, with inconclusive results. Only after applying core temperature as a PSF the performance effect of temperature was uncovered.

To clarify ideas, the model of the impact of sleep loss on task performance, developed over the last 15 years (Belyavin & Spencer, 2004) will be sketched. From analysis of the nature of performance degradation as a consequence of sleep loss fatigue, it has been concluded that performance is mediated by an individual state – currently identified as Alertness (Bunting & Wilson, 1997). This measure is constructed as the sum of two components, one based on time of day, representing circadian rhythm, and the other based on time since sleep. Since the Alertness state is influenced by the circadian effect, which can itself be modified through shift patterns or rapid time zone changes (Montgomery & Spencer, 1996), there is not a direct link between sleep deprivation and performance change, and it is necessary to consider the intervening state measure. In addition, since a key component of the model is time since sleep, a model of the effect of sleep on alertness has to be included, and this is provided by the “S” process (Daan & Beersma, 1984). Finally, if complex patterns of sleep and wakefulness are to be considered, a model of the sleep process itself may be required since sleep duration is important to the overall model.

Analysis of the relationship between the Alertness measure and performance of a range of simple tasks indicates that Alertness is a direct predictor of task performance. There is no indication that separate consideration of the two components of the model provides a stronger relationship. To determine the effect on specific tasks, the distinct effects of Alertness on perception, cognition, and motor action have to be identified and quantified. The complete sequence of models represents a more detailed breakdown of the right-hand frame of Figure 6-4 and is displayed in Figure 6-5.

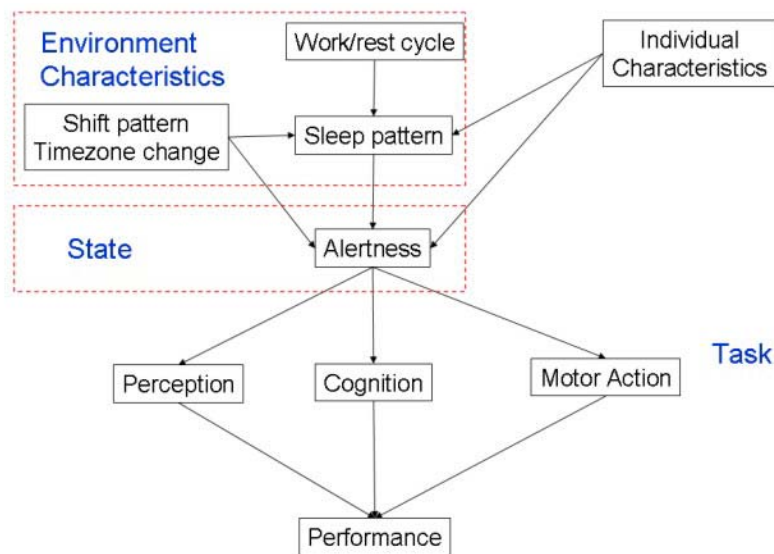


Figure 6-5: Model Chain for Sleep Deprivation and Circadian Rhythm.

In practice, the same causal sequence from environment to task performance almost certainly applies to all the environmental stressors in the first group. In a review of the literature undertaken for the IPME project, Bradley and Robertson (1998) concluded that a complex representation of the impact of thermal stress on task performance involving thermal strain, thermal comfort and dehydration was consistent with the available evidence. Changes in all three of these measures follow environmental changes with some delay. If a pattern of physical exertion and rest is imposed on top of temperature changes, then both operator traits (body size and percentage body fat) and clothing characteristics are involved in the calculation of all three state measures, implying a complex dynamic relationship between environmental stress and task performance.

In the psychological literature, a similar approach to the analysis of the effect of stress on performance has been employed, although there is a tendency to identify a single state measure – arousal – rather than a multiplicity of state dimensions as the previous outline suggests (Hockey, 1986). Whether a single measure can be employed to cover all states remains to be tested rigorously. What the scientific evidence indicates is that a sound predictive model of human performance under stress should be considered as a two-stage model: first a model of operator state and then a model relating state to task performance. Both of these stages almost certainly involve more than one model component and both may be moderated by individual characteristics.

6.6 MODELING THE EFFECT OF STATE ON TASK PERFORMANCE

In Sections 3.1-3.2, the general problem of modeling task performance in terms of environmental stressors has been outlined. It has been concluded that the model should comprise a two-stage process: projection of operator state from the environment or identification of interference effects, followed by projection of task performance from the intervening state. If the second stage of the model is to be implemented satisfactorily, a method of relating performance of specific tasks to operator state has to be devised. There are two main approaches that have been employed:

- Engineering modeling (e.g., used in IPME and IMPRINT); and
- Constructive modeling (e.g., used in ACT-R and Soar).

6.6.1 Engineering Approach

Classically, psychologists have classified tasks for the purposes of identifying performance changes through the processes that the tasks are assumed to employ. Typical classifications recognize verbal or spatial processes, visual or auditory input, and so on (Wickens, 1984). Engineering models of human performance that have attempted to take account of stressors have all made use of some simplified form of task taxonomy to model the effect of stress on performance.

Having identified the appropriate taxonomy, the performance degradation associated with a change in state measure for each taxon can be derived and the model completed. An example of this approach is provided by the representation of degradation in the IMPRINT tool (Allender, Kelley, Archer & Adkins, 1997). It has not yet proved possible to construct a sufficiently well populated matrix of performance degradations, task processes, and potential state measures to confirm either a specific task taxonomy or a specific complete set of state measures.

6.6.2 Constructive Modeling

The constructive approach is intrinsically more complex than the engineering approach in that it depends on a low level model of task performance built in an appropriate framework. It has the clear advantage in that it is based on a formal model of the processes involved in task execution and can be used to project beyond the data involved in construction (e.g., Gunzelmann, Gluck, Van Dongen, O'Connor & Dinges, 2005).

6.7 MULTIPLE STRESSORS

The chain between cause and effect for a single stressor has been outlined in Sections 6.1 and 6.2. The question arises as to whether the changes to a particular state measure induced by one environmental stress are independent of the changes induced by another at the same time. Sparseness of information prevents a categorical answer to this question at present. However, in a review of the literature on combined stressors undertaken for the IPME project in 1998, Robertson and Bradley concluded that within the limits of current evidence, stressors could be treated as independent. The main focus of the study was on sleep loss fatigue, physical fatigue, and thermal effects, although a more limited review of altitude effects and the impact of noise and vibration was undertaken.

6.8 IMPLEMENTATION ISSUES

The implementation of the model outlined in Figure 6-4 and Figure 6-5 involves a number of distinct elements that may vary with model requirements and the stressors represented. The architecture that is used must support at least the following elements:

- Representation of the variation of the stressor variables;
- Representation of the variation of the state variables; and
- Implementation of the models that describe the relationship between stressors and states and between states and performance.

Constructive models that describe the evolution of states in response to variations in environmental stressors can involve the solution of complex systems of differential equations. For example the prediction of thermal state is frequently undertaken using whole body thermal models that include solution of blood flow dynamics,

coupled diffusion equations and clothing properties (e.g., Stolwijk & Hardy, 1977; Werner & Webb, 1993; Wissler, 1985). Where complex models are to be re-used for the purposes of HBR, re-implementation of the model from scratch in a new framework is not an economical solution. Alternative approaches are required either through the use of libraries or the implementation of client server architectures. The latter approach has been used in IPME with some success, although exchanging information between the HBR framework and the thermal client can become a significant overhead. Although fewer constructive models of the relationship between state and performance have been developed to date, some have been constructed in ACT-R (Gluck, Gunzelmann, Harris, Scheutz & Kershner, 2007), and the same issue applies to the relationship between state and performance.

For engineering models based on relatively simple regression relationships between variables, direct incorporation in the HBR framework is more likely to be a practical proposition. It is practicable to implement such models for procedural tasks where the metrics of performance can be defined as time to complete a specific procedure and whether the procedure was completed successfully. These metrics are well suited to engineering models that employ task frames, such as the SAFs or task networks, such as MicroSaint, IMPRINT and IPME.

6.9 MODELING TASK DEMAND

Sections 6.1 to 6.4 describe the modeling process for the effects of environmental stressors on task performance and behavior. Task demand – workload – has generally been treated as a distinct stressor in its own right since it is a function of what an operator is doing, and a different form of model has been adopted. It is well established that trying to execute too many tasks in too short a time frame tends to lead to performance degradation. From the modeling point of view, the key question is whether the model describes the *state* of overload without explicit representation of the *effect* of overload or the impact of overload on performance is modeled.

There are established methodologies for modeling workload in systems based on multiple resource theory (Wickens, 1984) notably the Visual, Auditory, Cognitive, and Psychomotor (VACP) methodology (McCracken & Aldrich, 1984) and the W/Index system (North & Riley, 1989). These two approaches use a system of task ratings and a method for determining conflicting combinations that provides an estimate of operator workload as a current state. The objective of the use of state is to enable a system designer to pin-point periods of high workload and investigate modifications to the system that mitigate the level of workload for the operator. Alternate models of workload have been developed for IPME): the Prediction of Operator Performance (POP: Belyavin & Farmer, 2006; Farmer, 2000) and Information Processing (IP) models (Hendy & Farrell, 1997) that predict the change in performance as a consequence of task load.

There are three basic components of any workload model: the overall task model (task network or other form), the ratings applied to individual components within the model, and the calculus for combining workload ratings for different tasks. Overall predictions of workload can be made that appear correct, yet two or more model components may be incorrect in different ways that cancel out. In addition, if the workload model attempts to predict changes in performance and behavior in response to “high” workload, there is a negative feedback element present as well. Validation of the open loop components of a model from closed loop observations is always difficult if negative feedback is present.

One way to provide some basis for the validation is to use ancillary evidence, such as performance prediction, to test whether the basic model is sound. This is feasible for those workload models that predict the impact on performance. If it is possible to establish the soundness of the basic model, the predicted workload associated

with a single task can then be used to provide validation of the allocation of ratings. Finally the calculus for combining ratings from different tasks can be validated by comparing observed and predicted ratings for dual-task cases. Care has to be taken in assessing these predictions since there is a behavioral component in the response to high workload in that tasks may be deferred or ignored as a consequence of the stress and the choices made will affect the level of workload experienced. A simple example of the validation of a complex workload model is provided by Belyavin and Farmer (2006).

6.10 REPRESENTATION OF INTERNAL MODERATORS

6.10.1 Systematic Component

The systematic components of internal moderators can be represented in the same way as external moderators. An individual characteristic can be defined, such as the level of experience, and the impact on task performance can be described either as a direct effect on performance of a range of tasks or can be mediated by an internal state such as the background knowledge of a task. For either case the approach is very similar to that used for the external moderators and the problems involved in model development are the same. The primary problem in the development of a satisfactory model for the internal moderators is the identification of a sufficient set of moderators to span the space required for military simulations. A satisfactory set of external moderators can be identified by demanding that the range of environments relevant to the current analysis has been identified. For the internal moderators, the question arises as to which moderators may have an effect in a specific context and from this to determine what segments of the population should be represented in the model.

The literature on ways of describing and measuring personality is diverse and provides a large number of candidate frameworks for defining traits. The literature on the relationship between these traits and performance or behavior is equally diverse and the range of observed effects is not easy to bring together in a single framework that would encompass all observations in a way that lends itself to the construction of models. A number of groups have developed methods of representing personality as a series of traits coupled to measures of state that can include emotions. Notable examples are the Personality-based Architecture for Cognition (PAC, (Read et al., 2006; Zachary, LeMentec, Miller, Read & Thomas-Meyers, 2005) and the work of Silverman and colleagues (Silverman et al., 2003; Silverman, Johns, Cornwell & O'Brien, 2006; Silverman et al., 2001). As an illustration of possible approaches, an approach to modeling coping styles is outlined below.

6.10.1.1 Illustrative Example

As an illustrative example, consider potential models of different individual coping styles.

6.10.1.1.1 Empirical Evidence

Under the influence of stress, people change their behavior and mental state. Given a stressful situation, there will be more than one way to handle it. Several psychological studies have explored these ways of coping. Lazarus and Folkman (1984) distinguished between problem-focused (PFC) and emotion-focused (EFC) styles of coping when a person experiences stress (assessed with the "Ways of Coping Assessment Questionnaire") This was criticized as too simple by for example (Carver, Scheier & Weintraub, 1989) who explored other ways of coping as well while developing the COPE inventory where different aspects of PFC and EFC, as well as disengagement styles was assessed.

Based on Lazarus, further empirical research has been done on for example the relationship between sleep and coping styles (Sadeh, Keinan & Daon, 2004) and self-efficacy and coping styles (Jex, Bliese, Buzzell & Primeau, 2001). Day and Livingstone (2001) compared different coping styles on chronic and acute stressors among military personnel to investigate their effects on health.

Another example of where coping styles have been investigated among military personnel is Svensson et al.'s (1993) study on fighter pilots. Here EMC coping was associated with "tension," "effort," and "adrenaline activity," while PFC was expressed in "commitment," and "activation" (energy mobilization for the task). Both coping styles were used increasingly as a consequence of greater challenge.

6.10.1.1.2 Modeling

Lazarus and Folkman (1984) talk about a general concept of stress and coping styles. This means, in a computational model, that all stressor variables (temperature, role confusion, mental workload, etc.) could be summed up and passed through a coping style module (possibly submodule of a trait module) that in turn triggers higher activity in other parts of the architecture (e.g., appraisal module for EFC and cognitive processing regarding the task for PFC) such as shown in Figure 6-6. Such a model would distinguish between "EFC people," and "PFC people."

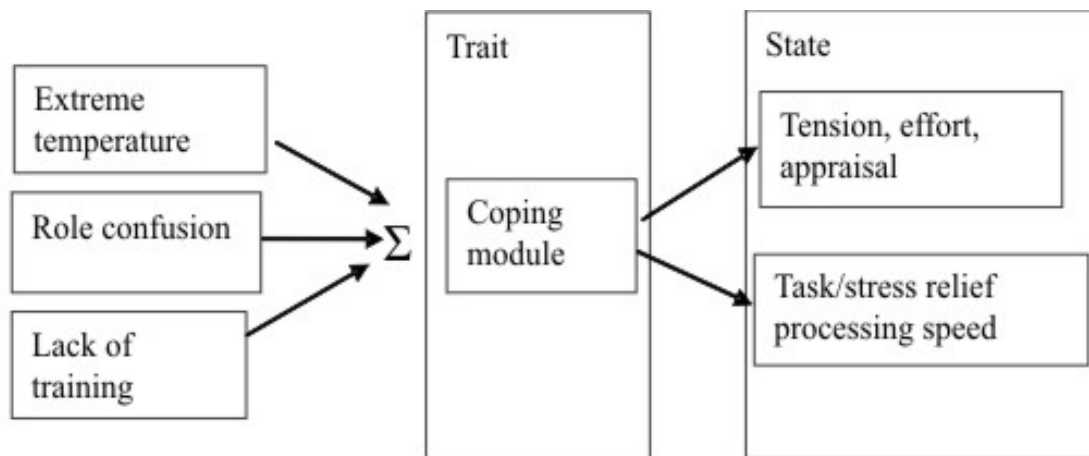


Figure 6-6: Illustrative Approach to Coping Style.

Finally, one modeling attempt that takes coping style into account already is PMFServ (Silverman et al., 2006). For the stress model, PMFServ distinguishes between "event stress," "effective fatigue," and "time pressure" (Gillis & Hursh, 1999) but sums these up in an integrated stress value. This integrated stress value is then input for Janis and Mann's coping patterns (Janis & Mann, 1977). Figure 6-7 displays the relationship between the accumulated stress and decision effectiveness. It follows the pattern of the Yerkes-Dodson arousal curve in that at both low and high levels of stimulation, decision effectiveness is reduced and is optimal in a relatively narrow range in the centre.

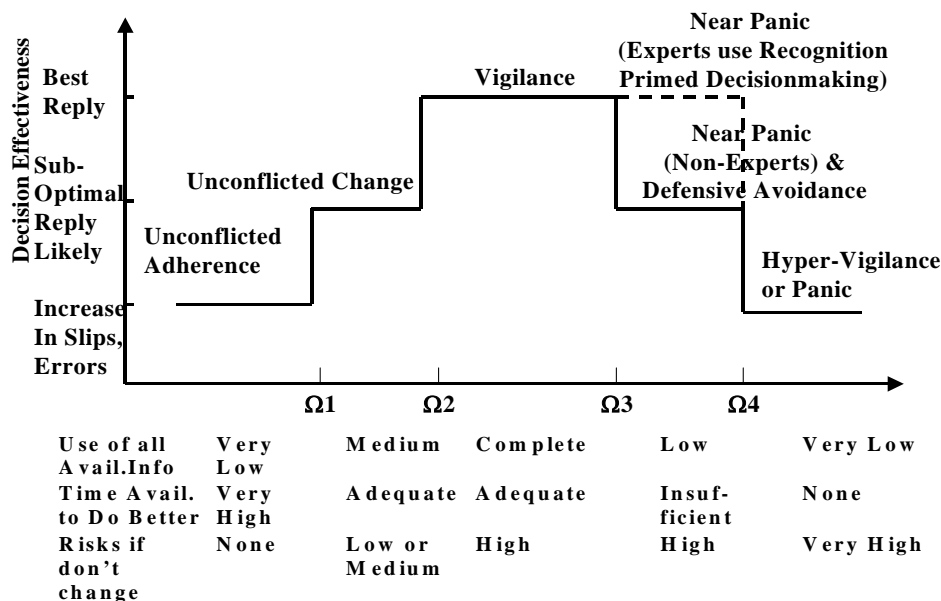


Figure 6-7: State Variable Decision Effectiveness and Performance.

While Lazarus and Folkman (1984) distinguished between EFC and PFC coping styles, Janis and Mann (1977) distinguish between decision strategies under different levels of stress:

Our current implementation closely follows Janis and Mann's model. In level 1 (Unconflicted Adherence) the agent does not update its perceptions about the world and continues doing whatever it was doing in the last decision cycle. In level 2 (Unconflicted Change) the agent does not update its perceptions about the world but uses those outdated perceptions to formulate its present COA nonetheless. In level 3 (Vigilant) the agent updates its perceptions and makes a decision based on which action offers the highest utility value. In level 4 (Defensive Avoidance) the agent updates some of its perceptions, but fails to update its perceptions concerning those objects that cause it the most negative Event Stress. In level 5 (Panic) the agent either cowers in place or flees. The stress thresholds at which agents shift between coping styles can be set on a per-agent basis, allowing for individual differences in reaction to stress. (Gillis & Hursh, 1999)

6.10.2 Non-Systematic Component

The second problem with internal moderators arises from inter-individual differences. If we wish to know the effect of these differences on the variability of the model findings, a model of the incidence of the relevant characteristics in the target population has to be constructed. The obvious assumption is that the set of characteristics is distributed independently. When the number of characteristics is small (<5), it may be possible to argue that independence is adequate. If the number increases, the assumption of independence is likely to be much more difficult to defend, and any model assuming independence will provide misleading models of the sampling properties of the population.

An example is provided by the variation of individual physical dimensions in populations – the study of Anthropometrics. Surveys of body size and shape frequently report more than 100 measures of body size

(e.g., Bolton & al., 1973). These are incorporated in man-models such as that implemented in the SAFEWORK framework. To investigate the match of the model man to a simulated workstation, it is frequently necessary to be able to represent the sampling distribution of the key physical man measures. Assuming independence for these measures would provide a misleading estimate of the fraction of the population that can use the workstation comfortably and it is necessary to model the inter-correlations of the variables in a realistic manner.

Constructing the joint distribution of a set of characteristics for the target population is a formidable undertaking, and while it is feasible for physical measures such as body size, there are currently few data available in the general literature for the less directly measurable quantities such as personality or aptitudes to support the approach. Although it is difficult to construct the sampling model for inter-individual characteristics, it is important that the distinction between inter- and intra- individual variability is correctly addressed in any model that is constructed.

6.11 TEAM AND ORGANIZATIONAL MODERATORS

The representation of team, group, and organizational performance is more complex than representing individual performance due to the interactions between the individuals in the group. In what follows, the word “collective” is used to specify the generic collection of individuals that form the team, group or organization. The behavior of a collective will be determined by the degree to which there are organized and inter-dependent roles and the procedures laid down that govern the interactions. A military team is normally clearly structured with defined goals and roles for all the team members and defined patterns of interaction in that orders are passed down a hierarchy and peer-peer interactions are managed. At the other extreme, a crowd has no defined roles, may lack a common goal, and lacks defined patterns of interaction. In the first part of this section, approaches to describing moderator in structured teams are discussed.

Larger entities (teams, squads) may have additional states relating to the collective. Team cohesion and team diversity do not exist at the individual level. On the other hand, load carrying capacity or mental capacity take on different meanings at the team level since individual states are shared in an optimizing process: stronger and larger team members get more to carry.

6.11.1 Teams

Behavior can be observed and analyzed at the individual, team and organizational level. We focus here on the team level factors. Teams are defined by Dyer (1984) as *at least two people, who are working together toward a common goal, where each person has been assigned a specific role or function to perform and where completion of the goal requires some dependency among group members*. Interactions between team members incorporates coordinated activity, communication as brought about by the organization and structure of the team necessary to achieve the team’s objectives and social and emotional processes.

The literature on psycho-social factors that are important to team work is diverse and originates from different settings. Most studies focus on antecedent – consequent relationships between certain psycho-social factors or task performance factors. However, in several studies, attempts were made to integrate various psycho-social factors into a more general team model, with a common basic structure. The structure can be reduced to an input -> throughput -> output model, in which moral/cohesion influences the throughput or is envisioned as output. Group cohesion can be considered the major psycho-social team factor, with leadership, team diversity, and social support as important co-factors.

6.11.2 Cohesion

Oliver, Harman, Hoover, Hayes, and Pandhi (1999) conclude in their meta-analysis that group cohesion has a very significant relationship with performance. Group performance is even more influenced by cohesion than individual performance. Cohesion has an influence on performance but also the other way around. Shared experiences of a team are the first requirement for the emergence of group cohesion.

6.11.3 Leadership

Leadership is not naturally given, as long believed but can be learned. The crux of this theory of situational leadership is that a leader is able to apply the appropriate behaviors depending on the demands of the situation (Hersey & Blanchard, 1982). With respect to cohesion, participative leadership seems to be the most appropriate form of leadership to enhance cohesion. Cohesion is enhanced when leaders are focused on team cohesion, even if the team is already cohesive.

6.11.4 Team Diversity

Team members vary in age, gender, rank, ethnic background, and other visible characteristics, but also in knowledge, skills, and personality. In general, teamwork is enhanced by similarity in background and other characteristics (Ingraham & Manning, 1981) and an increase in diversity can decrease cohesion. However, team diversity may improve some team characteristics such as sensitivity or creativity, providing alternative perspectives and incorporating different experiences into the team decision making process.

6.11.5 Social Support

Mutual social support is one of the major characteristics for cohesive units. According to Manning and Fullerton (1988), social support is one of the prime predictors for wellbeing and health with soldiers. Social support has similarity with cohesion, but where cohesion is strongly related to productivity of the group, social support enhances harmony and interpersonal relationships. Social support buffers the effects of stresses such as harsh and difficult environmental conditions on stress reactions and mental breakdown.

6.11.6 Interactions

Although many psycho-social factors have been identified as modifiers of performance, the above four factors stand out. Cohesion reflects the interaction between leadership, team age and social support, expressing the resulting binding force.

If the distinction between task and psycho-social processes were to be made, Figure 6-8 would be a graphical illustration of the psycho-social processes and how these relate to task processes.

HBR modelers provided representations of teammates at all echelons, from commanders to squad members, customized according to the training objective. Simulation of coalition partners from other countries included modifications to HBR parameters such as cultural differences that lead to a likelihood to make certain types of decisions and variations in tactics. In some cases, it even included changes to fundamental aspects of cognition such as the factors affecting recall from memory and what aspects are considered important in situational assessment and decision making, all leading to different behaviors that broaden the team's perspective. HBR for enemy and neutral parties has been a long term challenge for the HBR community using simple rule sets, but incorporation of more flexible models that learn and reason, leading to behaviors that the analyst had never considered,

provides broader experiences and richer simulations for training, including actions that seem out-of-the-box of conventional military tactics and doctrine.

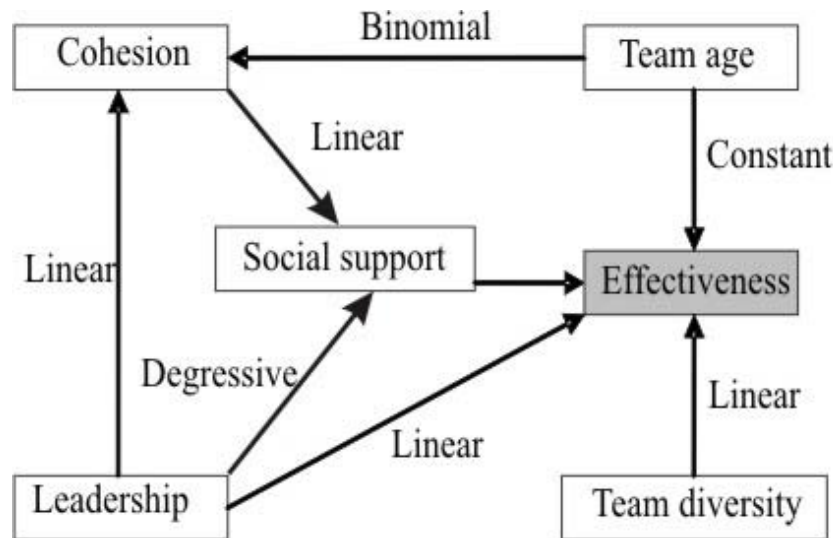


Figure 6-8: Team Characteristics, Leading to Team Effectiveness (after Bruin, Verwijis & Vliet, 2007).

6.11.7 Structured Teams

A useful simplification is to divide the activities in the team into *Taskwork* and *Teamwork*. *Taskwork* is defined as any activity undertaken by an individual that does *not* involve interaction with other members of the collective, while *Teamwork* is defined as any activity that explicitly involves interaction with other members of the collective. If this model is followed, all individual skilled behavior relating to a specific role is classed as *Taskwork* and all the interactions are grouped under *Teamwork*. The key to defining moderators for collective performance lies in defining moderators for *Teamwork*.

Teamwork is the collection of behaviors that defines how the group interacts and encompasses inter-individual communication, assistance, and monitoring. These behaviors drive how the collective performs and the way they are executed is assumed to derive from some level of team state. For example a team might be in a state in which every interaction is managed by the team leader. Peer-to-peer interactions occur only on the leader's orders and all monitoring is undertaken by the leader. Clearly, *Taskwork* is influenced by moderators in the same way as individual performance is moderated as outlined in Section 6.2 to 6.4 since it comprises individual performance and behavior. Discussion of team moderators must therefore include all the individual moderators. The external moderators are determined by the nature of the environment which is the same as for individual performance. The key question in team moderators is therefore whether there are additional internal moderators at the team level.

Research into military team characteristics using questionnaires indicates that there are a number of dimensions along which teams vary that can be classified under the general heading cohesion. These can be broken out into social cohesion and vertical and horizontal task cohesion (Siebold, 1999). Other studies indicate the importance of leadership as a source of variation, although it is not clear whether leadership is independent of cohesion (Essens et al., 2005). Cohesion is a set of measures that is subject to change with training and is therefore comparable with a state measure.

Culture is frequently identified as an influence on team and organizational behavior. It is likely that some aspect of culture is an enduring trait that influences the current state of cohesion, given a combination of other external and internal moderators. Similarly, there is evidence that the geographical distribution of the team – distributed or co-located – has an impact on team behaviors, at least under some circumstances (Cooke, 2003). Regardless of the nature of the moderators, their effect on team performance must be determined by the teamwork behaviors that they induce. Thus, the problem of identifying moderator effects can be reduced to one of identifying the effects of moderators on a specific repertoire of teamwork behaviors. From direct observations of teams there are four critical observable behaviors: communication, monitoring, feedback and backing up. The pattern and volume of these behaviors is the key driver of team performance. A putative summary of the candidate moderators that have been identified and teamwork behaviors is displayed in Figure 6-9.

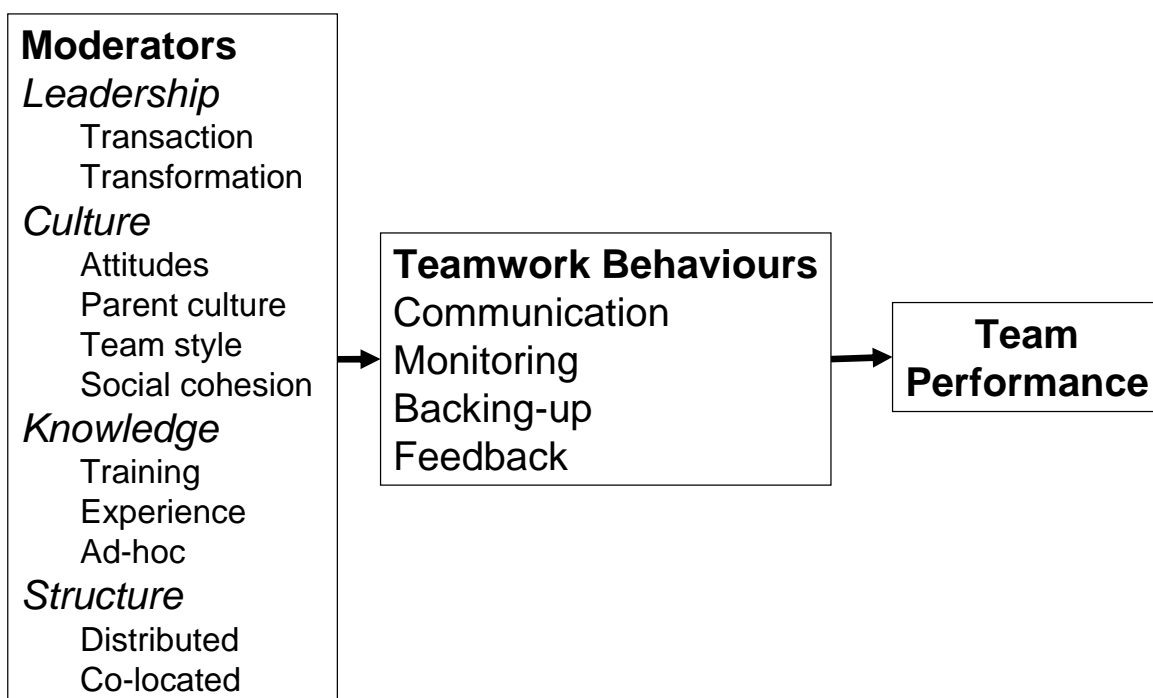


Figure 6-9: Relationship between Moderators, Teamwork Behaviors and Team Performance.

The key to collective performance is the contribution of the teamwork behaviors in Figure 6-9 to team performance. If there is no interdependence in the task, there is no need for teamwork and performance can be identified entirely in terms of the performance of the individuals that make up the team on their tasks. It should be noted that teamwork is an overhead in that it contributes to team performance indirectly and it causes workload for the team members. Too much teamwork is as bad as too little. To aid the process of understanding, teamwork behaviors can be analyzed in terms of their overall contribution to workload and the pattern of teamwork in terms of who is interacting with whom with what behavior. Good teams have appropriate teamwork workload and a good pattern of teamwork interaction while bad teams have too much or too little teamwork of the wrong kind. The effect of collective moderators on both teamwork workload and pattern is the subject of active research and definitive statements of an exhaustive set of moderators for structured team performance is not yet available.

6.11.8 Unstructured Groups

The modeling of unstructured groups has a number of applications in the military context. First there is a need for the representation of group or crowd behavior and the interaction with military personnel in peace support and peace enforcement operations. Second, there is the need to describe the evacuation of groups from confined contexts or flow through confined environments either in disaster relief or as a consequence of some intervention. The primary distinction between structured groups and unstructured groups is that in the former, there is a set of distinct roles each of which involves different tasks and patterns of interaction with the remainder of the group while in the unstructured case, the interaction pattern is more homogeneous.

In the limited practical examples of crowd models for the first application that have been developed, it has proved necessary to define a crowd state that determines both the pattern of interaction within the crowd and the behavior of the individuals. For example Nguyen, McKenzie, and Petty (2005) use crowd aggression as a state variable to moderate crowd behavior. Since the crowd is composed of individuals, at least part of their behavior will be moderated by the stressors that influence individual behavior. It is not clear whether the individual and collective moderators both have to be represented or whether the interactions between individuals dominate collective behavior and the individual moderators can be neglected. The understanding of moderators of crowd behavior is in its infancy, and a definitive list of both states and moderators has yet to be established.

Modeling the flow of crowds through restricted thoroughfares and the escape of groups from confined spaces has a longer history than the modeling of crowd behavior in the peace support or enforcement context. The main applications of pedestrian flow modeling are in the design of transport interchanges for railway stations or airports and in town planning. In the majority of cases a uniform rule structure is applied to each agent to represent both individual behavior and interactions with other agents, although the parameters within the rules can be moderated through the choice of characteristics of groups of the agents. Modeling human flow in evacuation is a smaller market than general pedestrian flows, but the principles involved are similar in that relatively uniform rule structures can be used to model the behavior, although these can be moderated through the selection of group characteristics. A key element that differentiates evacuation modeling from other cases is that the initial conditions – assumptions about how the crowd is arranged at the start of the simulation – are an important element in determining the simulation results. Examples of evacuation models are EXODUS (Filippidis, Galea, Gwynne & Lawrence, 2004), Myriad/SIMULEX (Olsson & Regan, 1998), and STEPS (RSSB, 2004).

6.11.9 Crowds

As military operations become increasingly oriented toward asymmetric operations, it is desirable to provide realistic and accurate real time simulation of crowd-level human behavior. However, current SAF and individual behavior modeling techniques were never designed to support the simulation of entities that is scalable to aggregate levels of group or crowd behavior. As such, there is movement within the community to attempt to provide realistic models of crowd behavior for interactive environments.

The phenomenon of crowds covers a wide range of behaviors and situations. A riot is characterized by the violent or aggressive behavior shown by (some) crowd members. Crowd research has advanced a lot in the last 15 years as numerous observational studies have been done (Adang, 1998; McPhail, 1991) that would result in similar behavior (Allport, 1924; Miller & Dollard, 1941) provided in empirical evidence by falsifying prevailing intuitive crowd theories. These theories focused on explaining riots that were based on, for example, a group mind that takes over behavioral control (LeBon, 1895) or on the idea that similar dispositions of crowd members cause similar behavior. Although these intuitive theories were contradicted,

the observational studies do not give insight into what happens between and with individuals in a crowd and why certain behaviors arise. To be able to model the behavior of an individual in a crowd, we need to take a closer look at all the different levels of influence that play a role in determining crowd behavior.

A crowd is a group phenomenon, which implies an emphasis on the social setting in terms of behavioral influence. The social environment mainly takes effect by perceived behaviors of other individuals. A social network of an individual defines the relationship between him and other entities more than his individual properties. Such a relationship between individuals can be characterized by the existence of a relationship, the intensity of influence, and the frequency of interaction between individuals. As in a crowd people tend to stay in the neighborhood of friends (Aveni, 1977; Neal, 1994), the description in terms of relationships is more relevant as the distinction between friends and others is more dominant. Furthermore, studies relating physical factors with statistical data on aggressive behavior found a relation between factors such as weather, noise, and scent (Krahe, 2001). Weather extremes tend to ameliorate aggression, presumably because the energy to maintain crowd behavior is lacking, while arousal, noise and scent work as enhancing moderators on an already aggressive mood (Geen, 2001; Rotten, Barry, Milligan & Fitzpatrick, 1979).

Individuals in a crowd show a dynamic repertoire of behaviors. As human behavior is always goal-directed (Kendrick, Neuberg & Cialdini, 2005), the range of behaviors that can be expressed in a crowd serves specific goals. As it is possible to show the same behavior for different reasons, as well as having several available behaviors serving a certain motivation, it is important to understand what drives the behavior. Abstract goals are therefore important as the motivation behind a possible behavior.

Recent computational techniques being developed for the purposes of simulating crowd and group behaviors are fundamentally different from single entity-level behavior modeling. Typical approaches include those that model aggregate crowd behavior using fluid flow and network models to those that represent entity-level behavior bounded by physical laws (e.g., particle systems). In a crowd, avoiding collisions is a basic mechanism in which human density influences the walking behavior. Human density is a necessary characteristic of crowds since it involves the co-presence of other individuals. These techniques have been applied to simulating crowd behavior for the computer-gaming and movie industry, training of military personnel or policemen, crowd motion simulations to support architectural design for everyday use and for emergency evacuation conditions, simulations of physical aspect of crowd dynamics and lastly sociological and behavioral simulation. Of note is that crowd modeling techniques are not concerned with the internal cognitive processes of individual entities but focus on the simulation of the outward observable behavior of the aggregate system. Internal drivers are sometimes used to generate this behavior without simulating real cognitive processes. For the purposes of the current report, these issues are not further addressed.

6.11.10 Culture

There is evidence to suggest that subtle differences in the organizational and national cultures of contributing nations' armed services can have an impact upon the overall operational effectiveness of the multinational force. It seems necessary to consider, integrate, and guide armed forces on the inter-cultural issues and factors that surround multinational inter-working, particularly at the operational level of command.

Throughout the past century, many definitions of culture have been developed of which the following is a compilation:

Culture is defined through the existence of underlying ideas, conceptions or thinking patterns that are shared by all members of a group, that have been developed over time by those members while

solving problems, that govern how new problems are solved, and that are taught to new members as correct behavior and conceptions.

When we speak of cultural diversity in teams, we are actually referring to three types of cultural diversity:

- 1) Cultural diversity within a team: Diversity within teams has been associated with both advantages and disadvantages. Compared to homogeneous groups, heterogeneous groups tend to make more creative and higher quality decisions (Adler, 1990; Kai, Bridgewater & Spencer, 2001; Triandis, Kurowski & Gelfand, 1994). Heterogeneous groups are less socially integrated, have a greater potential for conflict (Berry & Kalin, 1995), and experience more stress (Triandis et al., 1994) than homogeneous groups.
- 2) Cultural diversity between teams: In order to succeed, the operational commander will have to optimize the forces assigned and create a true integrated team; will have to accept the implied task of training and development in spite of the argument that it should not be part of his/her mandate; and must keep in mind that success might very well rest with the weakest contingent. Thus, the commander can ill afford to leave out or to marginalize any contingent.
- 3) Cultural diversity between a team and the social environment: Marginalization, sometimes to the extreme, happens when Western culture considers itself to be far more technologically advanced than another culture. This type of behavior can simply be defined as the refusal to associate with, the isolation or the rejection of an element of a group for whatever reasons.

In 1980, Hofstede presented his theory and extensive research in different nations on culture. Up to this day, his work still inspires many researchers. Despite criticism, his five factors model of culture value dimensions remains the most robust and influential model on culture, also among newer theories. Initially, Hofstede (1980) identified four cultural value dimensions. Later work (Hofstede, 1991) with the Chinese Culture Connection added a fifth dimension based on a study of Asian cultures, a region largely excluded from Hofstede's earlier work. These five dimensions are:

- 1) Power Distance: the degree of inequality among people that the populace of a country considers as normal.
- 2) Uncertainty Avoidance: the degree to which people in a country prefer structured over unstructured situations. Structured situations are those in which there are clear rules as to how one should behave.
- 3) Masculinity-Femininity: the degree to which values such as assertiveness, performance, success, and competition (which in nearly all societies are associated with the role of men) prevail over values such as the quality of life, maintaining warm personal relationships, service, care for the weak, and solidarity (which in nearly all societies are more associated with the role of women).
- 4) Individualism-Collectivism: describes whether one's identity is defined by personal choices and achievements or by the character of the collective groups to which one is more or less permanently attached.
- 5) Long-term – short-term: orientation based on the values stressed in the teachings of Confucius.

6.12 MODERATOR SUMMARY

The previous sections of this chapter have outlined the general framework in which moderators can be described and represented as part of HBR. Table 6-3 provides a high level summary of models that have been developed to represent the effects of some moderators in HBR and provides an outline of sources that describe the underlying models.

Table 6-3: Summary of Moderator Models

Moderator	State(s)	Performance or Behavior	Work and Model	References
<p>Sleep Deprivation</p> <p>Circadian Shift (Dynamic/Static)</p>	<p>Alertness (QinetiQ)</p>	<p>Performance</p>	<p>Range of work:</p> <p>QinetiQ – SAFE/IPME (1985-2006)</p> <p>SAIC SAFTE</p> <p>AFRL ACT-R model of fatigue and cognitive performance</p> <p>General literature</p> <p>General work covers the relationship between Sleep/loss and Circadian rhythm and Alertness. A number of models have been constructed, with SAFTE, SAFE particularly noteworthy.</p> <p>Further work covers the relationship between Alertness and task performance. Literature supports the notion that cognitive tasks are affected by Sleep Loss and Time of day but physical tasks are not.</p> <p>QinetiQ work views the steps from sleep loss to performance as a two stage relationship, and models them separately using different methods. Some results are described in the reference. The QinetiQ model also covers both “entrained” local circadian rhythm and a model of how the circadian oscillator adapts to changes in “zeitgebers.”</p> <p>AFRL are examining how to model task performance as a consequence of fatigue using ACT-R.</p>	<p>Belyavin A, Spencer MB. Modeling performance and alertness: the QinetiQ approach. <i>Aviat Space Environ Med.</i> 2004 Mar;75(3 Suppl):A93-103; discussion 104-6.</p> <p>Hursh SR, Redmond DP, Johnson ML, Thorne DR, Belenky G, Balkin TJ, Storm WF, Miller JC, Eddy DR. Fatigue models for applied research in warfighting. <i>Aviat Space Environ Med.</i> 2004 Mar;75(3 Suppl):A44-53; discussion A54-60.</p> <p>Allender L, Salvi L, Promisel D. Evaluation of Human Performance under Diverse Conditions via Modeling Technology. <i>Proceedings of Workshop on Emerging Technologies in Human Engineering Testing and Evaluation, NATO RSG 24. June 1997.</i></p> <p>Gluck, K., G. Gunzelmann, et al. (2007). Combinatorics Meets Processing Power: Large-Scale Computational Resources for BRIMS. <i>Behavioral Representation In Modeling and Simulation (BRIMS)</i>, Norfolk, VA, <i>Simulation Interoperability Standards Organization (SISO)</i>.</p>

SOURCES OF VARIATION IN HUMAN COGNITION, BEHAVIOR AND PERFORMANCE

Moderator	State(s)	Performance or Behavior	Work and Model	References
			<p>SAFTE is implemented in IUSS.</p> <p>The QinetiQ work has been used with IPME where the intervening state is specifically represented.</p> <p>IMPRINT uses sleep loss directly and steps to performance without representation of the intervening state.</p>	
Sleep Deprivation, etc.	Alertness	Behavior	Can extrapolate some performance measures although detailed investigation of behavior not known.	Gluck K.A., Pew R.W. (Eds) Modeling Human Behavior with Integrated Cognitive Architectures: Comparison, Evaluation, and Validation <i>Lawrence Erlbaum Associates Inc, US; (Aug 2005)</i>
Thermal Environment	Body temperature; Dehydration	Performance	<p>Number of models for predicting the state variables from a combination of environmental stressors (temperature, humidity, wind), clothing insulation, load carriage, exercise rate.</p> <p>These range from USARIEM Heat Strain model (P2NBC2), to full whole body thermal models.</p> <p>P2NBC2 has been used in IUSS – although it is marked as obsolescent.</p> <p>QinetiQ whole body thermal models have been used for load carriage analysis in conjunction with IPME.</p> <p>Some work has been done on relationship with cognitive task performance – although conclusions are not well reproduced. More work has been conducted in the heat than in the cold.</p>	<p>Stolwijk, J. A. J. and J. D. Hardy (1977). Chapter 4. Control of body temperature. <i>Handbook of Physiology – Reactions to Environmental Agents. D. H. K. Lee. Bethesda, American Physiological Society.</i></p> <p>Wissler, E. H. (1985). Mathematical simulation of human thermal behavior using whole body model. Heat and mass transfer in medicine and biology. <i>A. Shitzer and R. Eberhart. New York, Plenum Press: 325–373</i></p> <p>Bruyn L& Lamoureux T. Literature Review: Cognitive Effects of Thermal Strain. <i>DRDC No. CR-2004-191</i></p> <p>Belyavin AJ. Modeling the effect of thermal stress on human performance. <i>Military, Government and Aerospace Simulation. Vol 32,3. Ed: Michael J. Chinni. April 2000</i></p>

SOURCES OF VARIATION IN HUMAN COGNITION, BEHAVIOR AND PERFORMANCE



Moderator	State(s)	Performance or Behavior	Work and Model	References
			<p>Lengthy literature on thermal effects and task performance. Some attempt has been made to extract common themes from the performance literature and make sense of disparate data, although there is not agreement on all the relationships.</p> <ul style="list-style-type: none"> a) Dehydration appears to promote cognitive errors, although the results appear mixed, and not all authors agree. b) Increasing body temperature below 39(+?) appears to promote speed, without altering accuracy, although there may be secondary effects relating to thermal discomfort or anxiety and again not all authors agree. c) Increasing discomfort or anxiety slows down tasks and increases errors, although not all authors agree. <p>Possible explanation of this pattern is the combination of a number of different effects has not been dissected cleanly. Very difficult to induce dehydration without, possibly, producing additional thermal or fatigue effects.</p>	<p>Oksa J, Rintamäki H, & Rissanen S (1997). Dose dependent effects of cooling and rewarming on muscular performance. <i>Proceedings of The International Symposium on Thermal Physiology, Copenhagen, 8-12 July, 1997, Nielsen B & Nielsen R (eds), pp. 223-225.</i></p> <p>Hodgdon JA, Hesslink RL, Hackney AC, Vickers RR & Hilbert RP (1991). Norwegian military field exercises in the arctic: cognitive and physical performance. <i>Arctic Med. Res., 50:Suppl. 6:132-6.</i></p> <p>Gopinathan PM, Pichan G & Sharma VM (1988). Role of dehydration in heat stress-induced variations in mental performance. <i>Archives of Environmental Health, 43(1):15-7.</i></p> <p>Sharma VM, Sridharan K, Pichan G & Panwar MR (1986). Influence of heat-stress induced dehydration on mental functions. <i>Ergonomics, 29(6):791-799.</i></p>
Thermal Environment	<p>Body temperature</p> <p>Dehydration</p>	Behavior	<p>Much less complete picture – although errors work may be viewed as an element.</p> <p>Reduced work rates as the heart rate increases; ceasing voluntary work rate at body temperature tolerance.</p> <p>Distraction by discomfort from primary task.</p>	

Moderator	State(s)	Performance or Behavior	Work and Model	References
Visual Environment	N/A	Performance	<p>A complex set of psycho-physical phenomena have been modeled at a wide range of levels of detail.</p> <p>At the level of whole human target detection and identification the BAE Oracle model is one of the more complete attempts, including thermal and optical aids and character recognition on screens. Other models include the Visdet model and simple relationships.</p> <p>The basic principles of many of these models are manipulations of visual lobe.</p> <p>Visual target detection probability per unit of time; identification distance.</p>	Review: Alexander Toet, Piet Bijl, and J. Mathieu Valetton. Test of three visual search and detection models. <i>Optical Engineering</i> -- May 2000 -- Volume 39, Issue 5, pp. 1344-1353
Visual Environment	N/A	Behavior	Visual detections add to Situation Awareness. Recognitions have all sorts of implied behaviors.	
Fear/Anxiety Morale	Anxiety	Performance and Behavior	Range of work on first parachute jumps done in 1960s/1970s using performance tasks. Some other work on aspects fear and anxiety down the years. There is a QinetiQ review. Some work done at Bristol University for QinetiQ supports findings from previous studies. Suggestion that Fear/Anxiety creates “worry work”.	
Task Demand – Workload	N/A	Performance	Variety of models. There are a number of established methodologies for modeling workload in systems based on multiple resource theory, notably: Visual Auditory Cognitive Psychomotor (VACP) methodology and the W/Index system.	McCracken JH, Aldrich TB. (1984). <i>Analysis of selected LHX mission functions: Implications for operator workload and system automation goals</i> . (Technical Note ASI479-024-84). Fort Rucker, AL: Army Research Institute Aviation Research and Development

SOURCES OF VARIATION IN HUMAN COGNITION, BEHAVIOR AND PERFORMANCE



Moderator	State(s)	Performance or Behavior	Work and Model	References
			<p>These two approaches use a system of task ratings and a method for determining conflicting combinations that provides an estimate of operator workload as a current state. The objective of the use of state is to enable a system designer to pin-point periods of high workload and investigate modifications to the system that mitigate the level of workload for the operator.</p> <p>Alternate models of workload have been developed for the Integrated Performance Modeling Environment (IPME), the Prediction of Operator Performance (POP) (Farmer 2000) and Information Processing (IP) models (Hendy and Farrell, 1997) that predict the change in performance as a consequence of task load.</p>	<p>Boettcher K, North V, Riley R (1989). On developing theory-based functions to moderate human performance models in the context of systems analysis. In <i>Proceedings of the 33rd Annual Human Factors Society Meeting</i>, Denver, Colorado. October 1989</p> <p>Hendy KC, Farrell PS (1997). <i>Implementing a Model of Human Information Processing in a Task Network Simulation Environment</i>. Defence and Civil Institute of Environmental Medicine, DCIEM No 97-R-71.</p> <p>Belyavin AJ, Farmer EW. Modeling the workload and performance of psychomotor tasks. <i>Proceedings BRIMS 2006</i>.</p>
Workload	N/A	Behavior	<p>Overload may result in slow or erroneous decision making.</p> <p>Underload may result in inattentiveness and missed signals.</p>	
Noise		Performance	<p>Major impact appears to be communication. QinetiQ work supports the analysis.</p> <p>There is some work suggesting a focusing of attention in the presence of noise cited in Human Engineering compendium.</p> <p>In parallel with noise there is the so called “irrelevant speech” effect which may cause interference with some task performance. Work by Dylan Jones and others at Cardiff University indicates that the irrelevant speech effect can be reproduced using a variety of noise patterns that are not</p>	<p>Allender L, Salvi L, Promisel D. Evaluation of Human Performance under Diverse Conditions via Modeling Technology. <i>Proceedings of Workshop on Emerging Technologies in Human Engineering Testing and Evaluation</i>, NATO RSG 24. June 1997.</p> <p>Hockey, R, Varieties of attentional state: The effects of environment. In: <i>R. Parasuraman and D. R. Davies, Editors, Varieties of Attention</i>, Academic Press, New York (1984), pp. 449–483.</p>

SOURCES OF VARIATION IN HUMAN COGNITION, BEHAVIOR AND PERFORMANCE

Moderator	State(s)	Performance or Behavior	Work and Model	References
			interpretable speech. Some patterns of noise could cause problems with some cognitive tasks, although the strongest effect is on tasks involving memory of a series of stimuli.	Jones DM, Macken W.J. Irrelevant tones produce an irrelevant speech effect: implications for phonological coding in working memory. <i>Journal of experimental psychology. Learning, memory, and cognition</i> . 1993, vol. 19, no. 2, pp. 369-381.
Hypoxia	Blood oxygenation	Performance and behavior	There are no established models of the relationship between environment and state or state and performance. Substantial experimental work has been conducted over the last 70 years into the topic without a formal model having been developed.	
High G	Blood flow to the head and other parts of the body	Performance and behavior	There are no established models of the relationship between environment and state or state and performance. Substantial experimental work has been conducted over the last 70 years into the topic without a formal model having been developed.	
Vestibular Effects (1) Motion Sickness	Motion dose	Performance	The human central nervous system creates the sense of orientation by means of various sensory inputs (vestibular, visual, proprioceptive), and by means of experience, and under normal circumstances this works well. However, in situations with conflicting cues (e.g., visual and vestibular), there is a physiological reaction inducing nausea and vomiting described as motion sickness. This is modeled through motion stimulus input.	Bos JE, Bles W. Theoretical considerations on canal-otolith interaction and an observer model. <i>Biological Cybernetics</i> 86:191-207. 2002. Nooij SAE, Bos JE, Groen EL, Bles W, Ockels WJ. Space sickness on Earth. <i>Microgravity Science and Technology</i> 19:113-117. 2007 Brandt Th. Vertigo, its Multisensory Syndromes, <i>Springer, London, 2nd Edition</i> , 1999
Vestibular Effects (2) Disorientation	Unknown	Performance	Accidents due to spatial disorientation (SD) are common events in every air force and, although there has been considerable observation and analysis, there is no formal model of the process.	Previc H, Ercoline, WR. Spatial Disorientation in Aviation. <i>Progress in Astronautics and Aeronautics, Volume 203</i> . 2004.

SOURCES OF VARIATION IN HUMAN COGNITION, BEHAVIOR AND PERFORMANCE



Moderator	State(s)	Performance or Behavior	Work and Model	References
Training	Training level	Performance	<p>Standard quantification of training indicates t^{-A} where t is the amount of training, and A is the power, as the way near trained performance deviates from trained performance. Model extended by Anderson et al to cover both training and skill fade for ACT-R.</p> <p>Some work done by Micro-Analysis and Design on training Performance Shaping Factor using power law.</p>	<p>Anderson, J., Fincham, J., and Douglas, S. Practice and Retention: A Unifying Analysis. <i>J. Exp. Psychol</i>, 25(5), 1120–1136. 1999.</p> <p>U.S. ARI TRD Aid U.S. Army Research Institute. User's manual for predicting task retention. 1999.</p> <p>Anderson, J. R. Cognitive Psychology and its Implications (Sixth Edition), New York, NY: Worth Publisher. 2005.</p> <p>Archer R, Walters B, Yow A, Carolan T, Laughery R, Gillis P. Training as a Performance Shaping Factor in Computer Generated Forces. <i>Proceedings 8th CGF meeting 1999</i>.</p>
Experience	Years in organization; Years in role; years in job	Performance	Wide range of work, but not clear on the relationship with training.	<p>Sturman, M.C. (2001). Time and Performance: A Three-Part Study Examining the Relationships of Job Experience, Organizational Tenure, and Age with Job Performance. <i>Working Paper 01–05. Cornell University: Center for Advanced Human Resource Studies</i></p> <p>Bosman, E.A. & Charness, N. (1996). Age-related differences in skilled performance and skill acquisition. In F. Blanchard-Fields & T.M. Hess (Eds.), <i>Perspectives on Cognitive Change in Adulthood and Aging</i> (pp. 428–453). New York: McGraw-Hill.</p>
General Intelligence	g scale	Performance	Range of results indicating association of g with performance and learning.	

Chapter 7 – MOTOR TASK REQUIREMENTS FOR HUMAN BEHAVIOR REPRESENTATION

The hostage rescue now complete, the soldiers felt waves of exhaustion wash over them. Some of this was due to the intense cognitive concentration that the operation required, but some of it was due to pure physical exertion from motor actions. While the expected mission of the day, the delivery of medical supplies, required continuous motor control for driving and gross motor strength for lifting the boxes of supplies, the hostage rescue required both fine motor control and coordination of arms, hands, eyes – even breathing – for the aiming of weapons and gross motor activity and coordination to walk, run, crouch, and jump while balancing carried equipment and maintaining visual and auditory awareness of the surroundings. Almost beneath notice were the more subtle and nearly automatic motor actions required for speech communications and pressing buttons and typing to send and receive messages electronically. Thankfully, the dismounted soldier equipment had been designed to support, coordinate, and enable this broad range of motor activities through lighter designs and with more intuitive interfaces that supported soldier activities rather than being technologically convenient.

7.1 INTRODUCTION

The emphasis on EBO operations on the battlefield has brought about an increased requirement for the representation of higher level cognition. Even so, no actions occur without motor actions. That is, Soldiers accomplish tasks by moving (e.g., walking, running, lifting), shooting, and communicating (e.g., speaking, or typing, pointing). Motor tasks can be considered the HBR output of a simulation. In the most basic case – a reflex action – no cognition is involved; however, in virtually all other cases, motor tasks are tied to and depend on cognition. For instance, motor actions are required to obtain visual information in the first place via eye movements. Thus, motor task representation at some level is important for modeling both input *to* the human system and action or output *from* the human system. Of course, as has been discussed in other sections of this report, the question being asked – the model, the mission, the simulation environment, data availability, etc. – all influence the way in which motor tasks need to be represented.

Motor tasks at the level considered here are generally invoked as a series or set of movements that involves coordination of muscle groups. Looking at a peripheral target, for example, involves eye rotation, neck rotation, and spine rotation in proportion to the total deflection. By doing so, the limited high resolution visual field (the fovea) is turned toward the target for close inspection. More sophisticated movements involve 3D trajectories such as when the hand moves toward a position or an object around an obstacle. Discrete movements can also be trained or practiced to the point where they become motor “programs” (or “chunked”) such as touch typing, expert piano playing, pressing a familiar sequence on a telephone keypad, or a highly trained soldier operating a weapon. Motor training or practice thus turns explicitly controlled motor tasks into automatic ones, thereby reducing the cognitive effort. Therefore, when representing motor tasks for models or either expert or novice performance, the representation must consider the level of learning or practice and the degree of automaticity of the task.

One implication of the degree of automaticity is whether the motor task can be performed concurrently with other tasks. Of course, many motor tasks such as eye movements, speech, chewing, maintaining a hand grip, and walking on a smooth surface do not offer much interference with other tasks. Conversely, when the same part of the body is invoked, the conflict is obvious: one cannot play the piano and type on a computer keyboard at the same time. Also, motor tasks that require focused attention and cognitive control will interfere

with other motor tasks and other cognitive tasks as well. The caveat must be added, however, that the notion of “concurrent” depends on the level of representation. On the order of minutes or even seconds, “concurrent” has one meaning and at the level of milliseconds, it has another. Some further discussion of this is included in the rest of this chapter.

To organize this chapter, it is logical, even obvious, to break out computational approaches to representing motor tasks according to the body part involved. Moreover, there are several domains of research and application upon which we can draw and for the most part, they are in good agreement about the list of body parts, but, of course they vary according to the specified purpose and need. For example, see Pew and Mavor (1998) as reproduced in Table 7-1 or Ritter et al. (2002) for recent charts showing how motor actions are represented in cognitive architectures and task modeling tools. Also, to avoid repetition in the subsequent sections, two sources are acknowledged here as original sources of the so-called micro-models that have been included in other HBR tools and architectures: the early Human Operator Simulator (HOS) work of Wherry (1976) and Card, Moran, and Newell’s seminal book (Card et al., 1983), which not only included much data but also formulations for assembling it.

Table 7-1: Integrative Architectures, Motor Column of Table 3.1 in Pew & Mavor (1998)

Architecture	Motor Submodels
ACT-R	Motor processors: manual, vocal, oculomotor
COGNET	Abstract, with provision for user-defined models
EPIC	Motor processors: manual, vocal, oculomotor
HOS	Eye, hand, trunk, foot, etc., micro-models
Micro Saint-based tools	Times and accuracies, plus micro-models
MIDAS (and redesigned)	Jack-animated mannequin
Neural network based tools	Sensor/motor integration, oculomotor
OMAR	Default effector models
SAMPLE	Time-delayed procedural actions
Soar	Motor processors: manual, vocal, oculomotor

For this discussion of HBR motor tasks, we will consider the following: with respect to HBR, the list of body parts considered is:

- 1) Eye or oculomotor;
- 2) Speech or vocal;
- 3) Hands or manual; and
- 4) Legs, trunk, whole body or movement.

Additional qualifiers may be added to each category. For example, oculomotor actions may be more or less automatic (saccades) or consciously directed (search). Manual tasks, that is tasks primarily using the hands are considered to be fine motor tasks and may be discrete (typing or shooting) or continuous (driving). Movement and whole body tasks are both considered to be gross motor tasks and may be light (walking and carrying a light load) or heavy (climbing or carrying a heavy load). Another perspective on categorizing motor tasks is to

associate motor actions with specific types of tools, equipment, or situations. Yet another dimension of motor actions is crowd behavior, that is, the movement (thronging, milling about, coming and going) of large numbers of entities. While it is the case that some aspects of crowd behavior include aspects that are more cognitive, affective, motivational, or goal-directed, it is also the case that individuals in a crowd tend to exhibit certain behavior. They may come together in close proximity, yet may try to maintain some personal space or interpersonal distance. Individuals in a crowd will either be moved along with the crowd or in some cases, may try to move against it.

Before turning to a discussion of current computational approaches for representing motor tasks and some other motor-related factors, it will be worthwhile to mention one formal approach to measuring movement, the Methods-Time Measurement (MTM) system. MTM, credited originally to Maynard, Stegemerten, and Schwab (1948) and Clark (1972), is now a proprietary system used principally for manufacturing and industrial engineering applications. Given its user and application base, it is not generally considered a “model” of motor performance but rather a database of movement times. MTM involves detailed inspection of people performing movements via direct observation or video and then a meticulous timing procedure and standardized coding. The routine motions that MTM captures and documents so laboriously are finding a new application in robotics (e.g., Drumwright, Ng-Thow-Hing & Matarić, 2006). The value of MTM to HBR is that it is a potentially rich catalogue of movement times as a function of distance, type of task activity, and type of movement. For example, it breaks out five basic motions of finger-hand-arm movement (i.e., reach, grasp, move an object, position, and release). It offers a further break out of detail (i.e., disengage, apply pressure, and turn), eye movement times (i.e., move and fixate), and 15 full body motion types (see Table 7-2) (Schlick, 2006). There are refinements to the taxonomy on a host of factors including distance, acceleration, whether the motion is repetitive or not, and for some reaching and grasping movements, the features of the object, etc. Ultimately, MTM produces a sum of component motor actions to produce the time required to perform a task (e.g., the time to tighten a bolt = 2.25 s). Thus, it is a potentially rich resource of task data based on an essentially summative model of motor actions. Note, however, that MTM is principally used to describe repetitive, routine actions, not dynamic actions characterized by uncertainty and stress, actions performed in a multi-tasking environment, or even actions that might be carried out using different strategies, although there may be some new developments in these areas.

Table 7-2: MTM: Fifteen Fundamental Motions for Body Movements (After Schlick, 2006)

Body Axis		
With Shift	With Shift	With Inclination
Foot motion Leg motion	Side step Turn body Walk	Bend Arise from bend Stoop Arise from stoop Kneel on one knee Arise from kneel on one Knee Kneel on both knees Arise from kneel on both Knees Sit Stand

7.2 EYE MOVEMENTS OR OCULOMOTOR

The lightweight, flexible head-mounted display that Lieutenant Anderson employed in the 2020 scenario was first evaluated using a highly detailed model of visual perception and eye movements. The time to search a display was calculated as a function of eye saccades and eye fixations. Different display layouts were evaluated for efficiency and effectiveness.

The HBRs used to evaluate display layouts such as in this scenario are at the level of milliseconds and are able to “read” the display inputs in some fashion. Models of eye movement would also be used in HBRs where target search and scan patterns came into play. Eye movement is also a consideration when evaluating vision devices, such as night vision goggles or a helmet-mounted display. Of course, in order for an eye movement model to represent “reading,” in the full sense, it would have to be linked to models of perception as discussed elsewhere in this report.

One approach that serves as a starting point for enabling this level of HBR is to use micro-models of basic eye movements and task-based movements such as reading, largely drawn from the basic psychological literature. The notion of using elemental “micro-models” of performance and assembling them to represent larger task performance can be traced to the Human Operator Simulator (HOS) project (originally reported by Wherry, 1976) and to the Goals, Operators, Methods (GOMS) formulation proposed by Card, Moran, and Newell (1983). In the GOMS approach, reading rate is calculated by summing the time for individual eye movement saccades (~230 ms per) times the number of saccades per word for the total number of words to be read. This calculation would be varied, however, as a function of the skill of the reader and the complexity of the material, at which point the degree of cognitive processing involved may be the overriding factor, not eye movement time per se.

Other eye movement models break out eye and associated head movements separately from reading rate (see Table 7-3). For example time-based models (ARL, 2005, see also IPME and MIDAS)¹ are available for

- Eye movement time (target located in eye field), 100 ms (travel time only);
- Head movement time (target located in head field), 200 ms (travel time only);
- Eye fixation time, 100 ms – 500 ms; and
- Search time, $(T_m + T_f)N$ where N = number of fixations, T_m = movement time, T_f = fixation time.

¹ IPME: Integrated Performance Modelling Environment, <http://www.maad.com/index.pl/ipme>

MIDAS: Man-machine Integration Design and Analysis System, <http://human-factors.arc.nasa.gov/dev/www-midas/>

Table 7-3: Eye and Head Movements (ARL, 2005)

Eye Movement Time (target located in eye field)	100 ms (travel time only)	(Houtmans & Sanders, 1984; Sanders & Houtmans, 1985)
Head Movement Time (target located in head field)	200 ms (travel time only)	(Houtmans & Sanders, 1984; Sanders & Houtmans, 1985)
Eye Fixation Time	100 ms - 500 ms	(Houtmans & Sanders, 1984; Sanders & Houtmans, 1985)
Search Time	$(T_m + T_f)N$ where N = number of fixations, T_m = movement time, T_f = fixation time	
Reading Rate	52 words/min (5 saccades per word) 261 words/min (1 saccade per word) 652 words/min (1 saccade per phrase), Assumes: 5-letter words, 2.5 words or 13 characters per phrase	Card et al., 1983
Prioritization (i.e., of targets)	$.310[n(n+1) / 2]$ Where n = number of targets in a sector. This formula treats prioritization as a unidimensional [worth] pairwise comparison between all possible targets in a sector.	McCarthy and Plocher, personal communication, 1990
Terrain Association	Two stage process: Stage 1. Reduce size of area of uncertainty: Time = 5 sec. (Performed once every time a completely new view is encountered.) Stage 2. Pinpoint own location Time = 2 sec. Per terrain matching attempt. (Four to seven matching attempts required to pinpoint own location. Other job activities are typically interspersed in between terrain matching attempts.)	(Cross, Rugge & Thorndyke, 1982)

These times would be associated with tasks to be modeled; some variability could also be assigned, and the times would serve as model execution times. Eye movements are represented in the GOMS family of models (John & Kieras, 1994) as being able to occur in parallel with other activities (e.g., hand movements) and consist of two steps, initiate and move. The time to initiate will depend on the preceding action and the time to move will depend on distance.

A somewhat different modeling approach is used in the cognitive architectures such as EPIC (Kieras, 2004) and ACT-R (Byrne, 2001; Byrne & Anderson, 1998) wherein the event (EPIC) or the task goal (ACT-R), along with various memory and recall parameters produce times associated with the visual module processing and with the storage into and retrieval from the visual buffer. As a general rule, each step is approximately 50 ms. In EPIC, eye movements can be either voluntary (as in a directed search) or involuntary (as driven by an attention-demanding stimulus). Both types include a start and end marker, with the time to travel governed by distance from the start and end fixation points. EPIC also includes the level of detail that distinguishes the foveal, parafoveal, and peripheral views, called “retinal availability,” such that the items or information within the foveal view can be accessed simultaneously. This level of detail requires a considerable amount of “bookkeeping” on EPIC’s part to track the eyes in x, y, z, space but could be valuable for certain types of applications such as display design. EPIC has embedded parameter values associated with whether the eye movement is a saccade (a jump) or smooth (4 vs. 2 ms), and the refractory period before the movement can be repeated (10 ms) that fine tunes the 50 ms step.

ACT-R leveraged EPIC heavily for the ACT-R/Perceptual Motor (PM) implementation (Byrne, 2001; Byrne & Anderson, 1998). Both include the notion of parallelism in that certain visual motor actions can occur simultaneously with other motor actions (e.g., hands) or cognitive processes; however, ACT-R/P/M assumes that cognitive processing itself is essentially serial, thus placing that limit on the overall model. Also, although the ACT-R/PM leveraged EPIC, it is more a model of visual attention than of visual motor actions per se. It defaults to a 135 ms latency in attentional shifts that occurs asynchronously to other processing and can be used to represent both shifts in attention that roughly correspond to saccades and fixations. If more than one element is present at the fixation point, all of that information will be available to be perceived. Common to both the EPIC and ACT-R/PM architectures is the close link of eye movement to perceptual processing and cognitive processing, such that eye movements are virtually always modeled as a part of the full context. Both architectures are grounded in the large amount of search and reading literature, which can be used as a data source directly, and importantly, EPIC and ACT-R/PM models of these sorts of tasks, particularly in the human-computer interface domain, have been developed and so can be used for time and error estimates or as executable models directly. Note that EPIC and ACT-R are not the only architectures to include eye movement representations but are arguably the best documented.

It is worth mentioning that none of the eye movement models described is actually part of systems that take in “raw” visual input. They are all designed to operate with simulated environments, either represented implicitly within the task or represented explicitly by designating screen locations as containing certain items or by passing simulation-generated messages that describe the visual environment. One final note about eye movements is that just because the eyes move to a location or to an object in the line of sight, in HBR terms, there are other factors affecting whether the item is seen or later remembered (e.g., Darken & Jones, 2007), and these factors are associated with perception, cognition, and performance moderator effects.

7.3 SPEECH OR VOCAL

Armed with this virtual but realistic experience, along with good equipment and information, Colonel Schmidt sees his plan succeed and Lieutenant Anderson leads his platoon into the abandoned

warehouse to coordinate with the Special Ops Team to rescue the hostages without casualties. Fundamental to the success was the ability to communicate directly – not by typed messages – between Schmidt and Anderson. Speech conveys the message not only through the words themselves but also through the tone, pitch, and intensity of the sound.

Embedded within the scenario are references to networked communications and fluid C2 across multiple levels. The means by which this information is transmitted may well be voice. In an entity-level model, this is likely to be represented only as “give command” or “communicate.” That might be modified by being associated with a piece of equipment such as “via radio.” Times to execute speech are shown in Table 7-4. Because speech production is deemed a relatively low cognitive effort activity for native speakers, this aspect of HBR has not been extensively developed. The matter of deciding “what” to say, typified by a negotiation or an interrogation, or memorizing a speech or code, becomes a matter of cognition rather than speech per se. For speech production, EPIC, for example, uses a 100 ms initiation delay and then uses a syllable count for the production time (Kieras, 1997) ACT-R/PM (Byrne & Anderson, 1998) operates in a fashion similar to EPIC. While not exactly the reciprocal of speech, a baseline model for listening time is also shown in Table 7-4.

Table 7-4: Speaking and Listening Rates (ARL, 2005)

Speech rate	3.4 words/sec (large vocabulary), 2.4 words/sec (small vocabulary)	(McCormick, 1970)
Listening	2.4 words/sec	(Miller & Licklider, 1950)

7.4 HANDS OR MANUAL

Much more detail is available regarding motor task representation for hands and fingers, mainly because this work has been motivated by and applied to human-computer interaction and soldier-console interaction. In our keyboard-, mouse-, and button-intensive environment, hand and finger positions can be mission critical. Minimizing the number of keystrokes to execute a command can add precious seconds to a time-stressed decision. Ensuring that the hand movements required to operate a hand-held communication device do not interfere with operating a weapon can be the difference between life and death. HBRs have much to draw from since measurements of this type of motor action abound in the psychological literature. Fitts’ Law, which predicts the time required to rapidly move to a target as a function of the distance to the target and the size of the target (Fitts, 1954) is the starting point from which most current models grew. Decades of time/accuracy trade-off and reaction time data subsequently fed the current approaches for modeling hands or manual activity. Manual task times, either generally or as related to simple choice decisions, are shown in Table 7-5.

MOTOR TASK REQUIREMENTS FOR HUMAN BEHAVIOR REPRESENTATION

Table 7-5: Cognitive Activity Durations (ARL, 2005)
 τ_P is the time for the perceptual recognition, τ_C is the time for the cognitive processing, τ_M is the time for the motor response

Motor Process	$\tau_M = 70 \text{ ms}$	(Card et al., 1983)
Simple Reaction Times – On/Off Response	$\tau_P + \tau_C + \tau_M = 240 \text{ ms}$	(Card et al., 1983)
Simple Reaction Times – Physical Match	$\tau_P + 2\tau_C + \tau_M = 310 \text{ ms}$	(Card et al., 1983)
Simple Reaction Times – Name Match	$\tau_P + 3\tau_C + \tau_M = 380 \text{ ms}$	(Card et al., 1983)
Simple Reaction Times – Class Match	$\tau_P + 4\tau_C + \tau_M = 450 \text{ ms}$	(Card et al., 1983)
Choice Reaction Time	$K * \log_2 (n+1)$ k is a constant representing simple reaction time and is nominally set at 150 ms n is the number of possible alternatives	Hick's Law as discussed in Card et al., 1983

Table 7-6 is a listing of micro-models that apply to common motor hand tasks as associated with movement to a target area or with specific input devices. As with the eye movement times described in the previous section, these times would be used to create the task performance time in a task network model-type HBR, with some variability also assigned.

Table 7-6: Motor or Hand Control Duration (ARL, 2005)
 D is the transit distance to the target, S is the target size,
 N_{\min} = minimum number of keystrokes

Hand Movement (Fitts' Law-Welford Variant)	$I_M \log_2 [D/S + 0.5]$ I_M = Slope Constant nominally 0.1 sec/bit	(Welford, 1968)
Pushbutton or Toggle	400 ms	(Harris, Iavecchia & Bittner, 1988)
Rotary Dial	730 ms	(Harris et al., 1988)
Cursor Movement with Trackball	$I_M * \log_2 (D/S + 0.5)$ seconds I_M is a constant nominally 100 ms/bit	(Harris et al., 1988, from Fitts' Law)
Cursor Movement with Mouse	$1.03 + 0.06 \log_2 (D/S + .5)$ seconds	(Card et al., 1983)

Cursor Movement with Joystick	$K_D + 0.100 \log_2 (D/S + .5)$ seconds K_D is the intercept distance	(Card et al., 1983)
Cursor Movement with Step Keys	$0.98 + 0.074 (D_x/S_x + D_y/S_y)$ seconds D_x is the horizontal distance to target D_y is the vertical distance to target S_x is the size of a vertical step (default = 0.456 cm) S_y is the size of a horizontal step (default = 0.246 cm).	(Card et al., 1983)
Cursor Movement using Text Keys	$0.66 + 0.209 N_{\min}$ seconds 0.209 seconds Keystroke rate (seconds/keystroke) that approximates the typing rate for random words	(Card et al., 1983)
Single Finger Keying Rate	0.140 [0.060 to 0.200] seconds	(Card et al., 1983)
Typing Rate	0.209 seconds/keystroke	(Card et al., 1983)

EPIC and ACT-R will again serve as the exemplar cognitive architectures to elaborate on manual aspects of HBR. EPIC (Kieras, 2004), keeps track of the position of the hands and for some actions, the fingers, in the x, y, z space. Motor actions can occur in parallel with other types of actions and consist three phases: preparation, initiation, and execution. Further, depending on certain characteristics, the execution phase of one motor action, for example, can overlap with the preparation phase of another motor action. This allows for a rapid succession of motor actions while still accommodating the requirement for all three phases. As a modeling convenience in the sense of pre-assembled motor sequences, EPIC includes different action “styles,” “punch” (e.g., hitting a button), “ply” (e.g., moving a joystick or mouse), “point,” which is similar to “ply” but with a target location, and “keystroke.” The following are some of the defaults used in EPIC: preparation time for an action with respect to a given “feature” in the environment, 50 ms; initiation time, 50 ms; and burst time (or the minimum possible manual execution time), 100 ms. Additionally, EPIC specifies such parameters as key closure at 25 ms; ply minimum, 100 ms; keystroke execution, 280 ms; and keystroke point, 1100 ms. ACT-R/PM is structured similarly and specifies highly similar, but not identical values (Byrne, 2001): feature preparation time, per feature, 50 ms; burst time, 50 ms; Fitts’ coefficient for peck, 75 ms/bit; Fitts’ coefficient, mouse movement, 100 ms/bit; and minimum aimed movement time, 100 ms. To actually use these times, of course, the full architecture (either EPIC or ACT-R/PM) includes additional parameters, constraints, and ties to the central cognitive processing representation that puts these times in context.

7.5 LEGS, TRUNK, AND WHOLE BODY MOVEMENT

Movement, as with the other types of motor activity discussed in this chapter, can be viewed as the means to an end, the output end of the perception-cognition process, and the physical means to accomplish a goal. Walking is built of individual steps, each step requiring coordinated muscle activity, energy expenditure, etc., including feedback control loops. As the individual steps are executed, at some point, muscle fatigue or more general fatigue may set in, having an impact on the execution of each subsequent step. At some point, this fatigue may become the limiting factor in performance of the mission. In a benign environment, walking does not require any significant amount of conscious cognitive attention; however, the effect of various stressors (e.g., a heavy load, an uneven or slippery surface, low light conditions) may require additional conscious attention to coordinate movement during navigation, path planning, orientation and obstacle avoidance. When developing an HBR for which movement by foot is central to the problem under study, then the full gamut of movement-related factors will be important.

For the models of legs, trunk, whole body or overall movement, the type and variety of source material is even more varied than for the other types of motor actions discussed in this chapter. Gross motor and movement models for HBR draw heavily from two disparate sources: research in physiology, biomechanics, and kinesiology as well as instantiations in combat models and training simulations. Note that this chapter limits the level of detail discussed to the surface features of movement. The list of things it does not cover is quite long: It does not cover the strength requirements or muscle-level exertions for gross motor tasks (e.g., walking, running, lifting, carrying). It does not cover the tight neurological coupling between perception, attention, action planning, and action execution; muscular and path adjustments for environmental factors such as obstacles, surface traction, and evenness, motion; etc. (These sorts of factors were also omitted from the discussion of hand-arm fine motor tasks.) Then again, it is acknowledged, consistent with the HBR framework being used here, that physical fitness measures are linked to trait variables, and more momentary measures of muscle fatigue, nutrition, and hydration are linked to state variables as discussed in Chapter 6 on moderators. An HBR could include representation of deep physiological measures such as blood glucose, muscle lactate, etc., if sufficient data are available to link those measures to ultimate performance. The definition of a moderator in this context is a bit fuzzy. For example, when considering a soldier in standard issue gear, is the weight of the boot, for example, a moderator (Jones, Toner, Daniels & Knapik, 1983)? What temperature is normal? What road surface is considered baseline?

As a starting point, consider a baseline walking rate, 0.19 s/ft (as cited for IMPRINT: Clark, 1972; Harris et al., 1988). This would be useful for unencumbered movement, perhaps indoors or on smooth terrain for relatively short distances. Another approach is maximum sustainable speed. Hayes (1996) provides a detailed break out of maximum sustainable speed over a 4-hour period based on maximum sustainable metabolic rates (Pandolf et al., 1977). It was assumed that the soldier was carrying a 20 kg load that included clothing, equipment, and water. The analysis varied cloud cover (solar load), levels of humidity, terrain slopes, and road/terrain surface types and “solved” for the maximum sustainable speed for low, medium, or high risk to the individual for heat stress. Data are provided in m/s, km/h, and mph. The full set of findings is quite extensive, so only a few key observations from the report are included here as examples. Moderating effects of temperature and humidity on maximum sustainable speed are greater the higher the levels of temperature, humidity, or both. Solar load had little moderating effect. An increase of terrain slope from 0% to 15% had a significant lowering effect on maximum sustainable speed, and above that (15-50%) there was little effect but the speeds were quite slow (<1 m/s). Negative grades (going downhill) had an effect, but were less severe than positive grades. Roads of blacktop or dirt were about the same, but loose sand had a significant slowing effect. These data have been represented in the Infantry Warrior Simulation (IWARS²), which also includes

² <http://iwarsonline.com/home.shtml>

movement parameters for different postures or positions. The maximum speed for a standing position is ~7 m/s; for crouching, ~2 m/s; and for prone or crawling, ~0.5 m/s.

Another useful approach to capturing leg, trunk, whole body, or movement tasks comes from virtual reality-based training research and development for which taxonomies have been developed to capture and/or represent natural human motion. Templeman, Sibert, Page, and Denbrook (2006) describe a set of properties shown in Table 7-7. While no values are offered and the intended application is simulations with high fidelity visualization, variables are described that could be included for some level of motor action completeness in HBR. Additionally, Templeman et al. (2006) offer a framework derived from navigation for describing motor actions pertinent to infantry soldiers. They use the terms “heading,” “course,” and “displacement” where heading refers to the direction the head is facing, course refers to the direction the trunk is moving, and displacement is distance traveled. This effectively gives an articulated body so that the infantry soldier can look in one direction while moving another, as would be the case with certain scan while moving tactics. Further, they offer classes of motions: “steering” where the heading and course are the same; “canted” has heading and course maintain at a fixed offset; “oblique” where heading is fixed while course may vary; and “scanning” when heading varies independently of course.

Table 7-7: Properties Important for Realistic Body-Driven Simulations (Templeman et al., 2006)

Actions	Effects	Actions + Effects
Which body segments need to be modeled?	What degrees of freedom and range of motion must be output?	Is the simulation open- or closed-loop control?
How much effort is required?	What rate and accuracy need to be produced?	
Can multiple actions be coordinated?		

7.6 MOTOR TASKS, WORKLOAD CHANNEL LIMITATIONS AND INTER-CHANNEL CONFLICTS

Motor tasks run the gamut from reflexive action to highly controlled and practiced expert performance, from involuntary eye movements to running cross country. All of these actions require some particular body part for performance and when that part(s) is/are engaged, it/they cannot be invoked for some other, simultaneous task. This aspect of workload channel limitations is fairly clear. However, the degree and nature of the cognitive performance associated with a motor action, and therefore, whether any other motor tasks or cognitive tasks can be performed simultaneously is another matter. This entire topic can be considered in the category of performance moderators, but it should be clear that there are potentially complex interaction effects in high-fidelity modeling of motor tasks.

Another aspect of movement that might be better considered in a discussion of moderators is motion sickness. For instance, why is the world perceived to be essentially static while the projection of the world on the retina floats incessantly as a soldier moves through terrain? This is a fine demonstration of motoric-cognitive co-operation and involves complex brain work, considering not only the optic flow, but also the output from the vestibular system, the muscle activity of the eyes, and the position of the joints involved in posture. The brain concludes from the mixed information that the optical flow is the net effect of the soldier’s

movements and consequently not of a moving world. The calculation needs to be accurate since any error will be noticed as shaking surroundings. Indeed, when the feedback loop fails, the world starts moving. This can be observed by anyone shaking his or her head with increasing frequency. The human system has evolved to stand or to move under its own power, not to move in fast machines. Thus, under some circumstances, the mechanism fails and creates perceptual dissonance, resulting in motion sickness as vehicle passengers, amusement park visitors, aviators, and students using immersive training simulators know all too well (Johnson, 2007).

As a final word for this chapter, as with other types of HBR research on motor tasks, research continues. A recent article in a special section on modeling proposed that modeling motor actions needs to consider that the various factors that govern reach and movement such as ease of joint rotation, required speed, obstacles, etc., are better represented by flexibly ordered constraint hierarchies and not strictly sequential execution (Jax, Rosenbaum, Vaughan & Meulenbroek, 2003). Moreover, motor execution is modified by individual differences or momentary differences with respect to strategies.

Chapter 8 – TASK EXECUTION AND ERRORS

Lieutenant Anderson's men hurried to the hostage rescue because he had tasked them to move and arrive at a certain time. The mission will be jeopardized if they take too long to get there. The men struggle against their thirst and exhaustion to meet this goal. En route to the abandoned warehouse, they encounter a person who could be an enemy. Do they challenge the individual? Do they check their SA display for updates? They know the rescue mission is time critical. They do not want to let down their comrades. What if the individual attacks? What if one of the platoons needs medical attention? Ultimately, Lieutenant Anderson's men manage these dynamic goals and tasks successfully in real time.

In the SAS-017 report RTO-TR-047 (Dompke, 2001) it was concluded that many shortfalls in operational modeling are not attributable to a lack of HF knowledge per se but rather to a lack of proper implementation of HF knowledge. At the time, level and scalability of representation, composability, intent handling, behavior moderator implementation, and inclusion of credible cognitive processes were among the challenges. Seven years later, many of these shortfalls still exist. This chapter offers a commentary on some approaches to enabling the representation of human behavior, especially the representation of error in M&S. Note, however, that specific software approaches (e.g., HLA), object- or agent-oriented programming, client-server relationships) are not discussed. The thrust is on the organization of HBR that will enable accounting for the perception, cognition, moderators, and motor tasks described in this report.

Of course, practicality influences – or even dictates – the specifics of any given HBR implementation, especially when it is inserted into or linked to an existing simulation. Even within practical limitations, cases such as the one described in “Driving without Looking: Storing Behavioral Data Information in the Terrain Model.” (Kornman & Longtin, 2001) should be avoided. In that case, in the effort to have a more efficient simulation, aspects of line-of-sight perception and the rules of the road were placed onto road objects rather than with the vehicle driver. In other words, human perception, attention, cognition, and decision making capabilities were placed on objects in the environment, not incorporated into the HBR. It may have been efficient in the short run, but such a structure will not accommodate HBR in the long run.

Models can be organized such that tasks are executed by a series of hard-coded rules that vary as to their generality or specificity. However, this approach requires extensive and detailed scenarios coupled with rules that foresee *every* possible situation. Of course, even a very large number of decision rules cannot realistically anticipate every possibility. A more robust approach is to give the HBR the capability both to “respond” to as full a set of events as possible in the simulated environment and to carry out goal-directed actions.

One notional example of such an HBR is as follows: A military movement-to-contact task is described parametrically by the mission purpose and planned arrival time; distance to be traveled, waypoints, formation, load, terrain, and threat potential and location. The HBR entity “plans” its movement based on these factors, noting that “purpose” includes a goal of what the arrival condition should be. Thus, the HBR model keeps track of fatigue and heat tolerance through a calculation of work rate (speed, load, terrain) and the resulting body temperature and blood glucose level states. If these are not met, resting and eating are necessary to continue. Under threat, the movement may be slowed due to extra search tasks being carried out. The size of the formation can also slow movement, particularly for formations that involve 360 degrees of vision, for the rear of the formation cannot see backward and forward at the same time. This approach unifies a number of key factors into a single movement-to-contact task network. The same basic network could be augmented by more cognitive goals and tasks or by moderators such as moving in chemical-biological protective gear.

8.1 TASK NETWORKS: REPRESENTATION, SELECTION AND RESOURCE LIMITS

As a starting point for this discussion, assume that tasks have a beginning, a middle, and an end and thus lend themselves to discrete event-type or agent simulation. This holds quite well for tasks at the level of seconds and minutes and has been embodied in the human performance modeling tools in common use today such as IMPRINT¹ and IPME², both of which are based on the MicroSaint³ simulation language. A classic reference on this level of task representation is “The Psychology of Human-Computer Interaction” (Card et al., 1983). It lays out the GOMS formulation for examining task performance using building blocks of task elements. GOMS was later put into a computational form with a number of variations (John & Kieras, 1996). Figure 8-1 represents a simple task network with branching, which could be governed by different sorts of logic. The paths could run in parallel, or one path could be chosen on the basis of a probability or some other logic or parameter value. For example, if this represented a target ID sequence, 3a could be the decision that the target was a friend, 3b a foe, and 3c a neutral.

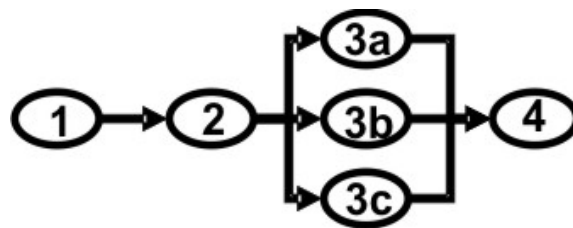


Figure 8-1: Notional Discrete Event Task Network with Branching.

Thus, in a task network, the elements (mission, function, tasks, subtasks, or whatever nomenclature is used) generally are arranged in the typically expected sequence but with some branching logic, variability, or decision rules. However, since humans rearrange task execution “on the fly,” dynamically, realistic HBR requires the same. Even though a task network approach seems to imply that tasks are carried out in a structured way, a “network” can be built with no hard-coded sequences at all. Instead, triggers can be used to access “free-standing” functions or tasks such that task execution can be tripped by scenario events (e.g., a change in weather, an enemy appearance, the arrival of ammunition supplies) or accumulated parameter values (e.g., time on task, mental workload, hydration level). Still, it can be argued that task networks are a better fit to routine procedures than to cognitively controlled actions. As has been discussed in earlier chapters, cognitive functioning can be represented at the level of task output as is the case in a typical task network model or at the level of detailed, millisecond-level processes as in the various cognitive architectures. Current research on modeling recognition primed decision making both pushes the limits of task network modeling and blurs the distinction between the two levels of modeling (Warwick & Hayes, 2003; Warwick & Hutchins, 2004).

Representation assumptions. An actual military unit “naturally” makes inferences when an order is received. An order for unit Alpha to “Go with unit Bravo to point Zero” implies that units Alpha and Bravo go to a meeting point and leave together, follow the same route, and arrive in time at point Zero. They understand the

¹ <http://www.arl.army.mil/ARL-Directorates/HRED/imb/imprint/Imprint7.htm>

² <http://www.maad.com/index.pl/ipme>

³ <http://www.maad.com/index.pl/products>

subtasks required, the timing and coordination, the waypoints, how to adjust when unexpected circumstances arise, etc. This logic also extends to the fact that soldiers know how to put one foot in front of the other. Of course, soldiers have been walking since toddler-hood and have been extensively trained, etc., all of which must either be assumed or implemented in a model. When constructing a model, decisions must be made to what degree the perception, cognition, and action and the associated implied knowledge must become explicit and what can be abstracted away. Factors such as the level of granularity of the particular M&S being used, the questions to be asked, etc., will help shape the appropriate level of representation. It is also the case that understanding what is indeed being assumed helps to ensure that the resulting model is as correct and rich as possible. Moreover, there are ways to structure HBRs that support the end goal.

Selecting a particular behavior. Key to the representation of realistic behavior and strategy is the selection of a certain behavior from a variety of behaviors. Alternative behaviors could stem from attempts to counteract the effects of stress (working less due to exhaustion, moving swiftly or ducking under enemy threat, falling asleep when sleep deprived, etc.) or from reactions to events in the environment (e.g., focusing attention to the location of an adversary, fascination for a partial problem), from the consequences of carrying out ordered tasks that contain incompatible subtasks (e.g., stopping moving to aim accurately, interrupting reasoning to talk to someone), or from individual preferences. One particular challenge in task execution is the implementation of strategy and the associated implementation of judgment and decision making. Even within tactically correct task execution, there is variability in the precise strategy employed. How would the behavior be chosen from this possible set?

A first cut at determining which task to select is often to ask military subject matter experts (SMEs). This is often a good starting point, but it must be acknowledged that even SME judgment is limited when it comes to making estimates that can adequately populate a distribution, when new technologies or equipment are being modeled, or when millisecond-level cognitive tasks are involved. Thus, various scientific theories can be brought to bear. For example, in reinforcement learning theory, behavior is optimized by finding strategies that balance drivers with goals where rewards for desired behavior are balanced with the costs. The cost will be minimized to just match the rewards and since the costs include various factors (discomfort, pain, exhaustion, fear, etc.), there is also room for avoiding certain cost factors. This view on behavior implements in a relatively straightforward way factors such as training, hardening and experience (reduction of perceived costs) as well as positive and negative motivation (manipulation of rewards by praise and punishment). Another theory proposed by Selfridge (1959) and generalized by Jackson (1987) is the concept of a Pandemonium in which “demons” are fighting for attention, each trying to “solve” the problem at hand. Dynamic behavior is then the subsequent handling of priorities, fed by stress and task demons. A simple and task-independent algorithm would satisfy the demon that shrieks loudest, requesting a certain behavior. Note, the COGNET architecture (Zachary, Ryder & Hicinbothom, 1998) grew from this Pandemonium notion.

At some level, the “screaming demons” could be conceptualized as the sort of level of activation notion at the heart of ACT-R (Anderson & Lebiere, 1998) such that the most active item in memory is selected for execution. This implementation is based on decades of psychological literature on learning, memory, and attention and happens within the framework of goal orientation (i.e., the idea that human behavior is predominantly goal-directed and that a task is selected that helps to meet the current goal).

Whatever the theoretical foundation, the model must keep track of the factors that will guide the selection from among alternative behaviors and must select one for execution at the required points. In Figure 8-2 one concept of a dynamic task network is illustrated. The tasking comes from the mission *context*, with a presumed, default starting order of task execution. The priority ultimately associated with a task depends on

TASK EXECUTION AND ERRORS

external environmental factors and internal HF (such as traits, perceived cost and benefits, history of goal activation, etc.). Alternative goals as moderated by the *state* of the human may be considered according to the theory or approach being used for decision making.

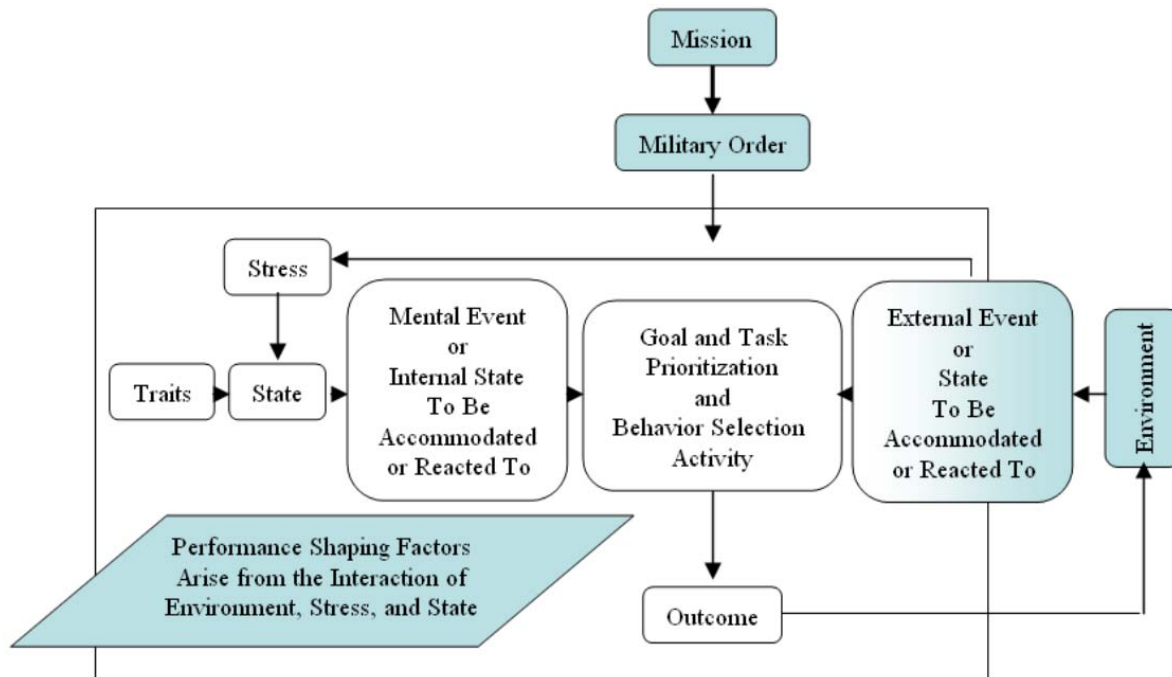


Figure 8-2: Behavior Selection and Resulting Performance Outcome Dependency on Task Priority.

Workload or processing resource limitations. A particularly interesting question is how one hypothesized aspect of cognitive state – workload or processing limitations – affects task execution. When a task is executed, processing resources are being used. As discussed in earlier chapters, there is the notion of the “full” exploitation of a resource by one task that leaves no room for other tasks that require the same resource. For some motor tasks, this conflict is straightforward. Tasks such as typing and playing the piano are incompatible. Some tasks may be carried out simultaneously (e.g., parallel branches in a task network representation). However, for other types of tasks, the “calculation” of whether sufficient resources remain is more difficult. Various workload approaches exist to determine how task execution unfolds if resources are “short.”

One approach to resolving the resource limitation question and its effect on error in an HBR is to make reasoned judgments about the kinds of resources available, the “amount” required by a task (i.e., task demand), and the consequence when insufficient resources are available. This is demonstrated in Figure 8-3. (See the discussion in Chapter 6 on representation of task demand as a performance moderator.) For example, if insufficient resources are available when a task is needed to be performed, the task may be:

- Performed less accurately due to perception or decision making errors;
- Performed more slowly, which adds time to the overall performance;
- Omitted, which, depending on its criticality, could be an error; or
- Delayed and performed at a later time.

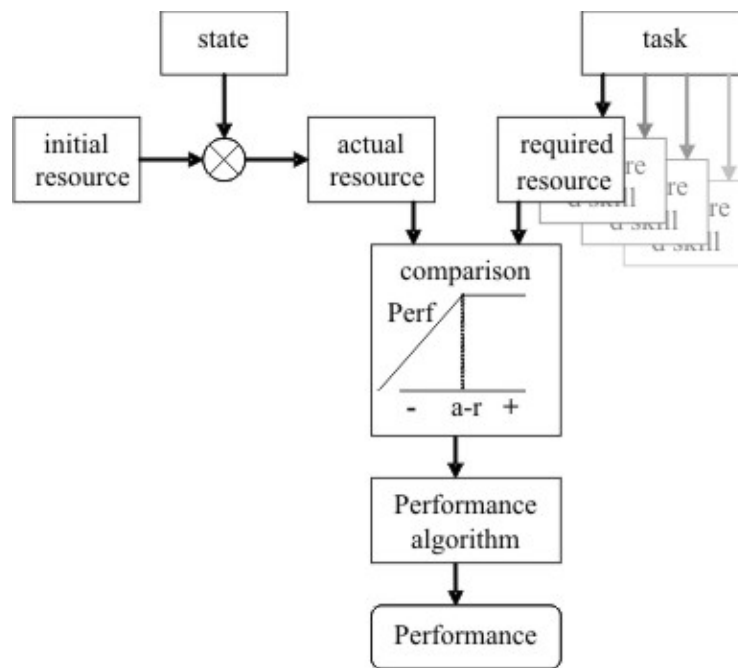


Figure 8-3: Task Performance is Moderated by Resource Demand and Availability.

A task requires resources, which may be met by the actually available resources. If so, the performance is not affected, despite the reduction of the initial resources by the current state. However, if the actual resource is less than required, performance will likely deteriorate.

Or the task could be given to another human or to some automated system function. For HBR purposes, the simplest expression would be to calculate a percentage of resources used based on a linear, weighted sum of the resources used, and if the total demand exceeds availability, there will be some decrement. Alternatively, complex matrices have been developed to guide a modeler through the assessment of within task demand, between task demands, all depending on the nature of the resource – perceptual, attentional, cognitive, motor, etc.; for example (e.g., see Wickens’ Multiple Resource Theory implemented in IMPRINT ARL, 2005). Of course, both of these approaches rest on the assumption that the amount of resources available and required can be estimated reliably and meaningfully at any point in time and in any context. Also it is not difficult to see that the linear model is deficient. When a man walks but cannot see, his performance will not drop a little, but will drop considerably, despite the fact that vision may be estimated as a minor required resource. A multiplicative model might be more realistic: performance is the multiplication of relative availabilities. Even better would be a model that knows to handle minimum and maximum requirements. The “amount” of cognition required to walk is limited but still essential since cognition is required to know where to go. Thus, a more advanced model would consider that the performance is multiplicative for all resources involved but with respective thresholds above which no further improvement may be expected.

The question of parallel versus serial in human information processing is an on-going debate in the literature. This debate is embodied in different cognitive architectures, for example, with EPIC arguing for no limit on central (cognitive) processing resources and ACT-R arguing for completely serial processes. Thus, a full implementation of an HBR that links traits, states, and the environment at a fine-grained physiological and cognitive level to ultimate system performance is the goal; however, the full set of research findings to model this definitively does not exist. Figure 8-3 shows how this could take effect in a model. As explained

TASK EXECUTION AND ERRORS

previously, the actual resources are assumed to depend on the traits (controlling the basic resources) and the states (adding dynamics to the resource). This needs to be compared to the task demands as explained above (required resource). Although Figure 8-3 does not highlight the fact that resources are not independent, of course these interdependencies should be considered. A task may require several resources and the actual use of resources is often higher because of interactions between resources. For example, when gross motor action is required for a task, the resource for fine motor action is reduced in a sort of parasitic way such that fine motor action is effectively disabled. An analysis of resource interactions shows that many of these interactions are asymmetrical: one hampers the other but not the reverse. Gross motor actions affect fine motor actions, perception, and cognition (for instance, particularly for high work rates and imminent exhaustion). The use of cognitive or perceptual capabilities may seem to have little effect on motor resources such as moving the whole body while walking; however, if the person is looking down at a hand-held device while walking, even the minimal perception required for walking is fully used by the other task of looking at the hand-held device. Likewise, using perceptual and cognitive resources to talk on a cell phone while driving appears to have a significant impact on attention but it may also have a significant effect on the fine motor resources required to control an automobile.

8.2 FUNDAMENTALS OF ERROR

Delivery of the supplies to the medical center is well underway when Colonel Schmidt receives an urgent message about a possible hostage situation in a remote section of the city. He immediately realizes that this implies a number of steps to prepare for his new tasking. While he is in the midst of checking unit status reports to support replanning for the rescue mission, he is interrupted again with a message about a power outage in a remote neighbourhood ... with no other indicators of hostile activity on the situation displays... he returns to checking the unit status reports. The system prompts him regarding exactly what information he has reviewed so far and highlights important aspects of the remaining information that should be reviewed to ensure getting the total picture.

The occurrence of human errors as an output of task execution is fundamental to many, even most implementations of HBR. That is, when HBR-based tasks are executed in a simulation, less than perfect performance, mistakes, deviations from optimal, etc., are all expected as a natural part of HBR. Although the human entities in simulations can be programmed to make mistakes, those sorts of models are typically deterministic and arbitrary. Indeed, it is the ability of robust and valid HBRs to represent and predict errors that is especially useful for various applications such as system concept exploration, design evaluation and, in the right circumstances, synthetic teammates or forces used for training. If likely errors in human-system performance can be predicted, then concepts can be refined and systems can be designed and re-designed repeatedly to mitigate or design out errors without the cost of “bending metal” and full scale field testing. When trainees are ready for a realistic trial of fundamental skills, the representation of error in simulation exercises can enhance overall learning.

This section presents the fundamentals of error, drawing upon a fairly extensive literature in error analysis; describes approaches to representing error in task execution; and discusses some additional factors to be considered related to error. Before proceeding, it is useful to distinguish the terms “variability” and “accident” from the term “error.” “Variability” refers to the precision over time or repeated outcomes of an action. Variability is observed both between people (individual differences) as well as within-person differences. As discussed below, variability may give rise to an error but is not an error itself in this context. On the other hand, “accident” refers to an unintended event, usually one that has a negative consequence and often has some cost associated with it such as damage to or loss of property, personal injury, or death. It is the case,

however, that many deliberate actions can culminate in an unintended, “accidental” consequence. It is in this middle ground between variability and accidents that the errors of interest here occur.

A formal definition of error can be found in basic HF textbooks and handbooks: “an action that violates some tolerance limits of a system” (Kantowitz & Sorkin, 1983). The implication of this definition is that errors occur in context. For example, an action that constitutes an error in one case will not be an error in another. Thus, the full context, which for HBR is supplied by the modeled mission, the simulation environment, and feedback loops within the HBR, is integral to our understanding of task execution and modeled error. Figure 8-4 depicts a rough scheme of the variety of interacting causal factors that lead to errors and subsequently, accidents, although of course, given the full context, this is not a deterministic relationship.

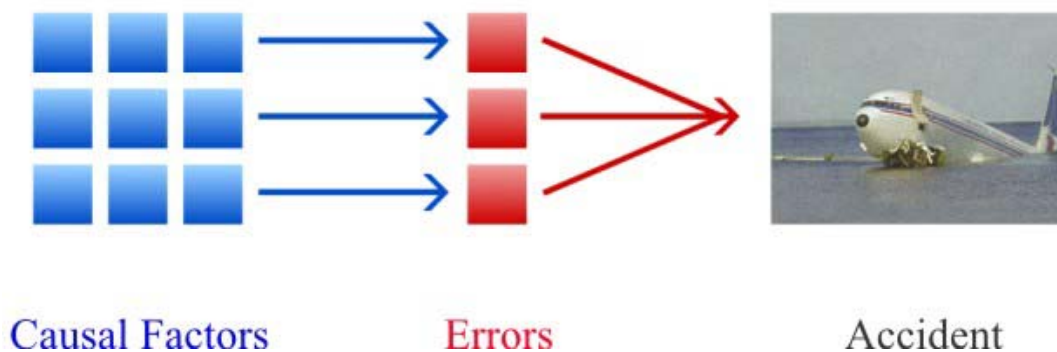


Figure 8-4: Three Levels of Accident Evolvment (Adapted from K  ppler, 2006, 2008).

Errors are most immediately thought of with respect to measures of accuracy, precision, or more generally, quality – percent correct, mm from target, a poor decision (although we can certainly argue about the current state of the art with any of those measures, notably decision quality). Of course, errors can be made with respect to time as well. A response may be too slow or too fast or may be in the wrong sequence. Further, there are complex speed/accuracy trade-offs that can give rise to errors.

Given the association of errors with accidents and cost, errors have been studied intensely by the HF and safety communities. For example, in a review Kirwan (1998) describes 38 approaches and methods for identifying error. He groups them into five categories: reliability-oriented methods, taxonomies, psychologically based methods that point to PSFs, more strictly cognitive analytic methods, and cognitive simulations that are computational. These approaches, although not completely orthogonal, all have some utility for understanding how to model error as a part of HBR. The reliability approach is not discussed in detail in this document but is prevalent in many communities where the overall statistics of accident rate, cost, etc., are of concern. Such base rate data are not necessarily useful as *input* to an HBR but rather, could prove useful as data against which the ability of a model to reproduce the rates over many, many runs might be validated or as input into proactive risk models. (See ARIADNE for one such large scale, web-based data base: K  ppler, 2006; K  ppler, Pre  ler & Specht, 2008)⁴.

Over the years there have been a number of useful error classification taxonomies proposed – the second approach for examining human error listed by Kirwan (1998). One early taxonomy of errors was put forth by

⁴ ARIADNE: <http://www.ariadne-sms.com/>

TASK EXECUTION AND ERRORS

Swain and Guttman (1983) for the nuclear regulatory community. It categorized errors as errors of omission, commission, extra act, sequence, and time. A recent volume of *Human Factors* contained a special section on classifying and understanding human error (Krokos & Baker, 2007). Across the eight articles, a dozen or more taxonomies are referenced. Dekker (2007) calls out some of the key types of taxonomies, which vary according to the orientation or goal of the researcher or practitioner: identification of preconditions or information transfer problems; stratification into cognitive, social, or situational factors; assignment along the information processing pipeline, to name a few. The most frequently cited is Reason's (1990) "Swiss cheese" model. Reason (1990) proposed four categories of error:

- 1) Lapse: memory failure (omitting planned items, place-losing, forgetting intentions).
- 2) Slip: attentional failure (intrusion, omission, reversal, misordering, mistiming).
- 3) Mistake: rule-based mistakes (misapplication of good rule, application of bad rule); knowledge-based mistakes (many variable forms).
- 4) Violation: routine violation; exceptional violations, act of sabotage.

These were further described as either active or latent errors that occur at different levels of an organization, that is, at different levels of proximity to the action. The notion is that each level is associated with factors that can contribute to an accident either directly or indirectly and that when on a given occasion these "holes" line up, an accident occurs (see Figure 8-5).

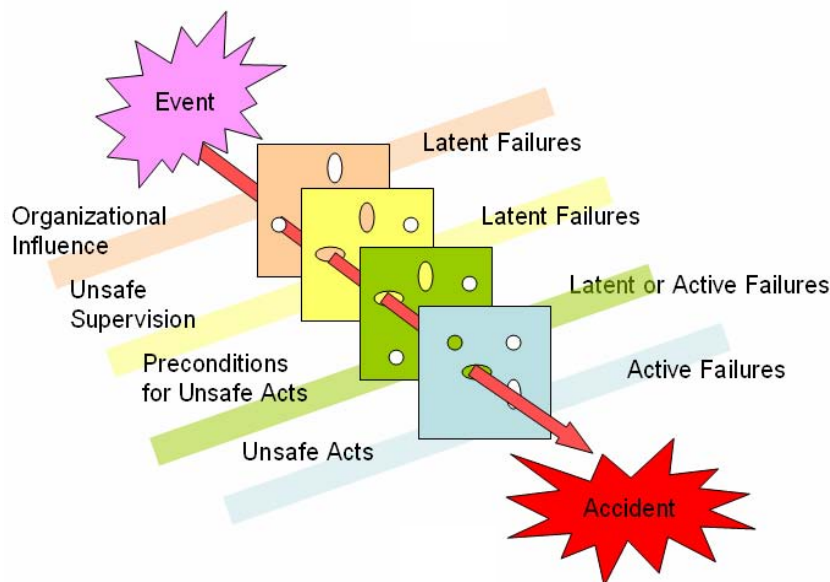


Figure 8-5: The "Swiss Cheese" Model of Human Error (Adapted from Reason, 1990).
Latent failures may be associated with several levels of control.

The Human Factors Analysis and Classification System (HFACS) framework (Wiegmann & Shappell, 2003) shown in Figure 8-6 elaborated upon Reason's approach and has proved to be a widely used method for assessing and understanding accidents (principally in aviation). Using this framework, errors are "back-associated" with various proximal and distal causes ranging from the operator's actions and physical state and local team factors to organizational climate and policies.

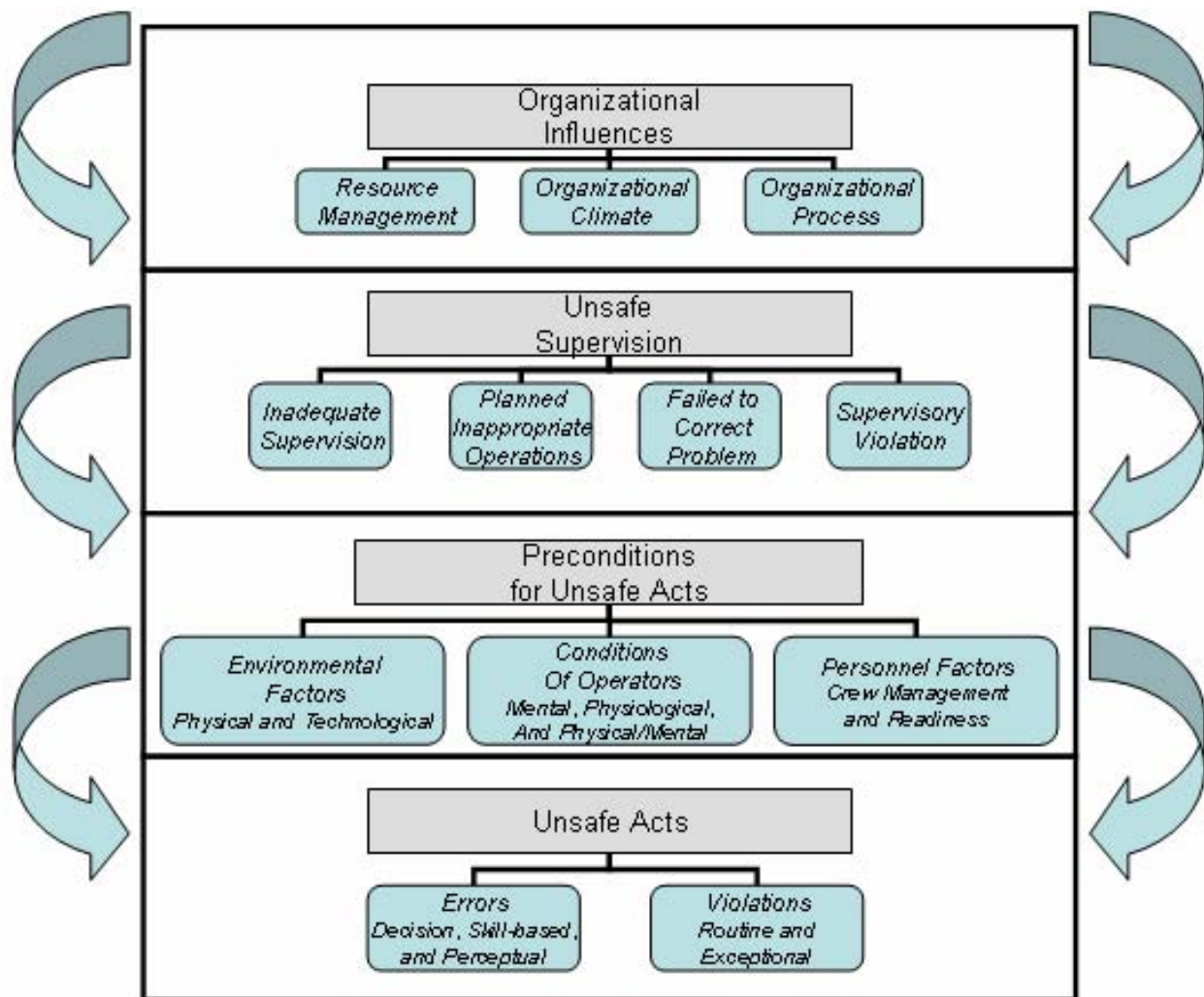


Figure 8-6: The HFACS Framework (Adapted from Wiegmann and Shappell, 2003).

Although the HFACS framework is quite inclusive, it is worth mentioning that taxonomies may still be tailored or developed to achieve a certain analysis goal. This has been recommended for the construct of SA generally (Baxter & Bass, 1998) and specifically with respect to aircraft pilot errors (Leiden et al., 2001). Nevertheless, no matter what the taxonomy, it begs the question of how the HBR actually gives rise to errors. Does the HBR invoke variability, wrong actions, or performance moderators? It turns out that the remaining three approaches for examining human error listed by Kirwan (1998) – psychologically based methods that point to PSFs, more strictly cognitive analytic methods, and cognitive simulations – fit naturally into the next sections and are part of the discussion below.

8.3 A NETWORK-CENTRIC VIEW OF ERROR: THE GAP BETWEEN THE REAL AND THE PERCEIVED WORLD

The concept of network-enabled military operations holds the promise that a multitude of sensors will bring the complete picture – all available information to the soldier. The idea is to extend the soldier's ability to perceive

TASK EXECUTION AND ERRORS

and therefore, to understand the situation. At the same time, the network environment has many points where variability and perhaps error can be introduced: the sensors, the network, the system software, and display technology will all contribute some variability (e.g., deviation in target x, y, z location; transmission lag; incorrect correlations of disparate sensor inputs; overlapping symbology). Also, the individual soldier's innate capabilities and limitations and the effects of moderators will contribute variability. The result is that the SA "achieved" by the soldier will not correspond perfectly to the "real" world. In Figure 5-1, the concepts of the real and the perceived world were introduced. In the dynamic environment of military operations, keeping track of friendly force movements is quite an achievement, supported by information exchange and GPS-based automated systems (e.g., Blue force tracking). A little more difficult is keeping track of neutrals and stay informed of static information; even more difficult still is following adversaries who purposely conceal their moves. The nature and extent of the delta between the real and perceived world is a major research area for HF and an opportunity for HBR to help provide descriptive and predictive insight.

Humans are continuously trying to understand the perceived world and form hypotheses about it. New information is considered in light of the person's existing understanding. If the new information is consistent, then the current belief persists and is strengthened; however, if the new information is inconsistent, it may be ignored, discounted, or misinterpreted. The difficulty of reconciling these sorts of differences can lead to errors and accidents. If someone acts according to his/her perceived world, can this be called an error? Only in hindsight, with the real world at hand, may it be called an error. The challenge is to understand and predict circumstances where these discrepancies are prone to occur.

8.4 REPRESENTING ERROR IN TASK EXECUTION

For the simplest GOMS examples of typing and text editing, error was not considered separately. It was treated as part of performance variability, either in performance time or in the ability of the person to self-monitor and do quick corrections so that the end performance still appeared error free (Card et al., 1983). For the more complex examples, different rules for method selection were invoked to account for what was termed "method-abortion," "method-failure," and "big" errors. These latter "big" errors, which technically were typing or method errors, were put into this separate category since they accounted for substantially longer times. This is useful to note as it points to a "threshold" beyond which variability and error can become system accidents with an associated cost.

In stochastic discrete event simulations, HBR error can arise when the extremes of the distribution of the task times or accuracies are sampled for a given model run. Or, error can be represented as probabilities in the branching logic. That is, as illustrated in Figure 8-1, different probabilities can be assigned to the branches following Task 2 so that for a given model run, either Task 3a, 3b, or 3c will follow Task 2. With repeated model runs, the three branches will be followed roughly according to the assigned probabilities. The error arises when the task on one of the branches constitutes an error, such as an incorrect action. More complex logic can give rise to other types of errors. For example, the consequences of error, which may range from inconsequential to fatal, may be included. The consequence can be propagated through the HBR to include subsequent degradations in time or accuracy or mission failure (see IMPRINT, ARL, 2005, for an example). Note, although this method is highly useful for many HBR applications and there may be SME or base rate data available to inform the setting of the probabilities, this method is a rougher approximation than the other methods described next.

Another approach to producing error in line with Kirwan's (1998) list of methods is to invoke performance shaping factors or moderators (see Chapter 6 for a more complete discussion of moderators). The idea is that the mission and the environment combine or interact with the traits to produce states and thus influence the

activity or action that is produced. Take the example of sleep deprivation; time of day, hours of sleep loss, caffeine intake, whether the individual is a morning or a night person, etc., will affect the time and accuracy of task performance. Belyavin and Spencer ((Belyavin & Spencer, 2004) describe the modeling of performance as a function of alertness, an aggregate state measure in a discrete event simulation. The model includes the link from the external factors (or context) to the measures of alertness (or traits and resulting state) and from there to the measures of performance. A similar approach is reported by Allender et al. (1997) for a variety of PSFs, which includes a differential mapping of various moderators to performance for different types of tasks (i.e., motor vs. perceptual vs. cognitive) but omits any accounting of trait or explicit physiological variables. Of course, the key step is making the link to performance, which becomes the output activity of the HBR. Still, degradations in performance do not necessarily culminate in an error, but rooting such degradations in moderator effects is a principled way to generate performance that can lead to error.

For HBRs modeled using the discrete event simulation approaches just described, certain parameters must be set in order to induce “errors” or something labeled as an error. One method is to set thresholds at which task performance is deemed unacceptable in some way. For example, a task that takes too long or is not performed up to some accuracy standard might be deemed an error. Another method is to construct the model with all known or reasonably imaginable task branching alternatives that comprise task commissions, omissions, and sequence-type errors. Further, thresholds can be combined with branching logic such that going down a particular branch is tripped by the threshold variables, which could be time or accuracy or some other model variable such as mental workload, alertness, or SA. With these methods, the model output would include performance times, error frequency counts, etc., for later analysis. Finally, when the HBR output interacts with some level of a system model, then the combined output is where the result of errors can be assessed more completely. This is where errors can be shown to culminate in accidents or mission failure, to be frequent and therefore risky, or to be intolerable in the full system context.

Error representation in HBR can be driven to a still deeper or more fine-grained level using cognitive architectures such as ACT-R⁵ or EPIC⁶. Whereas the variability in larger-grained models can be ascribed only roughly to perception, memory, or motor errors (the cognitive analytic category in Kirwan, 1998 list), when the HBR is at a finer level, these processes and the associated errors are represented directly. This finer level of abstraction is consistent with what Kirwan lists as cognitive simulations. Error can arise from the natural stochastic variability of the processes, including, in ACT-R, an explicit noise variable. Or attention may be diverted or misdirected so that an element in the environment may be misperceived. Also, a process might not achieve a sufficient activation level such that a fact may be mis-remembered or the wrong sequence of actions may be executed. Models of this type abound. Kelley et al. (2001) used ACT-R to account for soldiers making map recall errors while engaged in a cross-country navigation task. Gunzelmann et al. (2005) combined explicit sleep deprivation effects with ACT-R to achieve degraded performance and the concomitant errors. Errors at this level, just as at the task level, can be descriptive, diagnostic, and predictive but still rest on the completeness of the basic HBR. The models must be sufficiently rich and complete to generate the possibility of errors. Also, going back to the original definition offered for errors with respect to context, the more of the full system context available to the HBR, the better the error modeling will be.

8.5 ADDITIONAL ERROR FACTORS

When representing error in HBR, there are a few additional factors to be considered. One factor is the level of expertise being represented. A novice will make very different errors than an expert. If the HBR is being used

⁵ <http://act-r.psy.cmu.edu/>, 18Dec2007.

⁶ <http://www.eecs.umich.edu/~kieras/epic.html>, 18Dec2007.

TASK EXECUTION AND ERRORS

to explore system operational concepts, it may be best to model trained, experienced users. However, if the HBR is being used to help understand system training requirements, then it will be important to represent naïve users as well. Indeed, the very nature of the change in errors with the growth of expertise can be useful to help guide system design.

Although moderators have been presented as a way to produce performance errors, it is especially worth noting that modeling moderator extremes is particularly useful in exploring the full parameter space for human-system interaction. The errors that result from extremes in sleep deprivation or high-Gs are expensive and potentially dangerous to obtain. Modeling offers a way to examine performance cost-effectively and safely. It is also the case that HBR offers a practical venue for examining the errors resulting from multiple, simultaneous moderators; however, the exact ways in which moderators combine to affect performance is not yet fully understood and so this step must be taken with caution.

Another practical consideration for error modeling that applies as good modeling practice is to include model complexities in a step-wise fashion. First ensure that the baseline model executes correctly (verification), then add time or accuracy variability (noise) through a plausible mechanism. More complex branching logic or attentional variables could then be added to increase the representativeness of the HBR. In this way, the modeler will be assured that an error observed in a simulation results plausibly from the HBR and not from a programming error in the model logic.

It may be redundant to say at this point in the document, but it is important to be clear that the errors discussed here are not merely deviations from some mathematically derived optimal solution or something other than what a particular SME would have done. Rather, errors are produced by the very nature of human activity. This may be especially true in models that include higher-order decision making. If a model of a commander making a tactical decision on the battlefield results in the battle being lost, we cannot simply say that the decision was an “error.” The full context must be understood. Did the commander have access to the needed information? Was the information in an accessible, easily understood and remembered format? Was he or she sleep-deprived? Was he or she time-stressed? This, then, harkens back to Reason’s (1990) Swiss cheese model where a set of actions aligns to yield a bad result. This example also points up the power of including errors in HBR to more fully explore the parameter space in applications ranging from concept exploration and system design to training.

Part Three – The Future of Human Behavior Representation



Chapter 9 – OPERATIONALIZATION OF HUMAN BEHAVIOR IN THE MODEL ENVIRONMENT

9.1 A PROPOSED FRAMEWORK FOR HUMAN FACTORS IN SIMULATED OPERATIONS

The previous chapters were dedicated to identifying the essentials of human behavior and its representation as entities in effects based modeling. In this chapter we attempt to bring these elements together to provide practical guidance for the practical inclusion of HF in models and simulations that support the high-level problems outlined in Section 1.1.6. Many of the suggestions included here are relevant to any modeling endeavor, not just human behavior representation in military modeling and simulation.

The first proposal is to take systematic steps in the set up of your study, involving both the aims of the study and the approach to what is supposed to happen in your scenario. This is condensed in Section 9.2 “good modeling practice” of this chapter. The second proposal is the implementation plan of Section 9.3, showing how you could organize your simulation to avoid unnecessary complexity, while enabling human behavior of your entities. The third proposal is a combined scheme of analysis environment, HBR and EBO, serving as guidance for architectural choices. Finally, in Section 9.4, some application areas are addressed and put along three dimensions that may help to define your analysis environment.

9.2 GOOD MODELING PRACTICE

Good practice is meant here as a way to reliably include HF in a simulated operation. We did not call it best practice because it is simply not known what the best practice would be for all applications. Models and tools tend to be optimized for certain applications and may lack power or precision in other fields. This guidance may not resolve all requirements for a successful simulation, but it may help you in taking the right steps towards that goal.

Good practice in operational modeling involving HF takes the following steps:

A) Analysis of your study:

- 1) What is your client’s question?
- 2) What is it that you want to determine to help you answer your client’s question?
- 3) What is the quality of result you need: do you want absolute results or do you want to compare conditions; how many conditions do you wish to vary; do you want free or controlled behavior?
- 4) Identify constraints on the analysis: Cost, manpower availability, deadlines, information availability

At this point, it would not matter if you wish to perform a modeling study or do an experiment.

B) Finding a suitable study environment:

- 1) What is the setting of the study? Is it part of a known scenario and how does the study environment fit in?
- 2) Level of abstraction. Identify the level of abstraction and the components of the analysis required in the abstraction: Environment, platform, human behavior or cognition.

- 3) Credibility of the study environment. Consider if your selected environment and events fulfill both the demands for credibility with the military and relevancy for the subject of the study. Recall that the outcome of the study is highly influenced by the selection of the environment.
- C) Detailing what happens in your analysis environment:
- 1) Define the operational aim in your scenario. Select an analysis environment, including own units and opponents, define a mission and an aim for your units. Take care that the outcome is likely to discriminate between the variables under study.
 - 2) Select Measures of Effectiveness. Achieving your aim is the success factor. Define the cost factors and their weights, taking the mission into account.
 - 3) Planning of operation for achieving aim. Set up a tactical plan, deploying your units and exploiting their capabilities (including equipment). Do the same for the opponent.
 - 4) Break out of plan in tasks. Define tasks for the various units, the order to carry them out and synchronizations between units.
 - 5) Select Measures of Performance (MoP) that will form the basis of the Measures of Effectiveness (MoE). Define the outcomes on which you measure the task performance for the various tasks.
- D) Identify the essential human behavior representation aspects involved:
- 1) Which perceptual, motor, physiological, or mental factors are at stake, and to what extent does equipment affect the human competencies? Is the behavior typical of drills and rules and what would happen under various types of stress? What types of decisions are taken and who does so? Is catastrophic disorientation or errors possible? Consult a Soldier Day interactive checklist or equivalent. References: U.S. Mil handbook 46855 (AAMC, 1999); UK Defstan 0025; IEEE 1220 (IEEE, 1998); ISO System Engineering; U.S. DOD Acquisition Guide (DOD, 2003, 2004).
 - 2) Determination of HF involved in tasks. Define which human functions are relevant for the tasks in the particular analysis environment and how accurately you want to implement those.
 - 3) Set up task and behavior alternatives. Organize task components in simultaneous and sequential components. Define alternative behaviors that are associated with the stressors that may affect the units.
 - 4) Determination of key HF parameters. Which measurable or observable parameters do you regard as representative for the stress of the units, depending on the required accuracy, required detail, and type of entity?
- E) Defining the appropriate sub models:
- 1) Choose the level of detail or accuracy you think is necessary to simulate the operation sufficiently well and select appropriate software.
 - 2) Selection of PSF associated with HF parameters. Select the modules for the relevant HF calculating the key HF parameters from the data available in the task and analysis environment.
 - 3) Selection of task performance algorithms. Define how the performance on each task depends on the task parameters, environmental parameters, and the key HF parameters.

F) Run the simulation and draw conclusions:

- 1) Run the simulation and evaluate the result using the MoE and MoP chosen. Do extended analyses such as sensitivity analysis (checking to what extent your results depend on various conditions). Study interactions of conditions to avoid overly simplified conclusions. Find causes for unexpected results and mistrust outcomes if not every step makes sense. The model is to stimulate your understanding – it is unlikely a complete representation of the truth.

9.3 IMPLEMENTATION OF THE GOOD PRACTICE IN AN OPERATIONAL MODEL ARCHITECTURE

The backbone for an HF architecture in the EBO context is summarized in the previous list. The operation has an aim that is accomplished through the completion of tasks assigned to entities that typically include a plan and subtasks to be executed. The quality of the plan can only be judged from the effect that is created (a score on MoEs), but subtasks can be evaluated separately, by MoPs. Performance will depend on the entity state during the simulation that will, in general, be task, time and environment dependent. The state variables are inputs to the entity activities (behaviors) through PSFs. A conceptual representation of how the human model fits into the overall simulation architecture is shown in Figure 9-1.

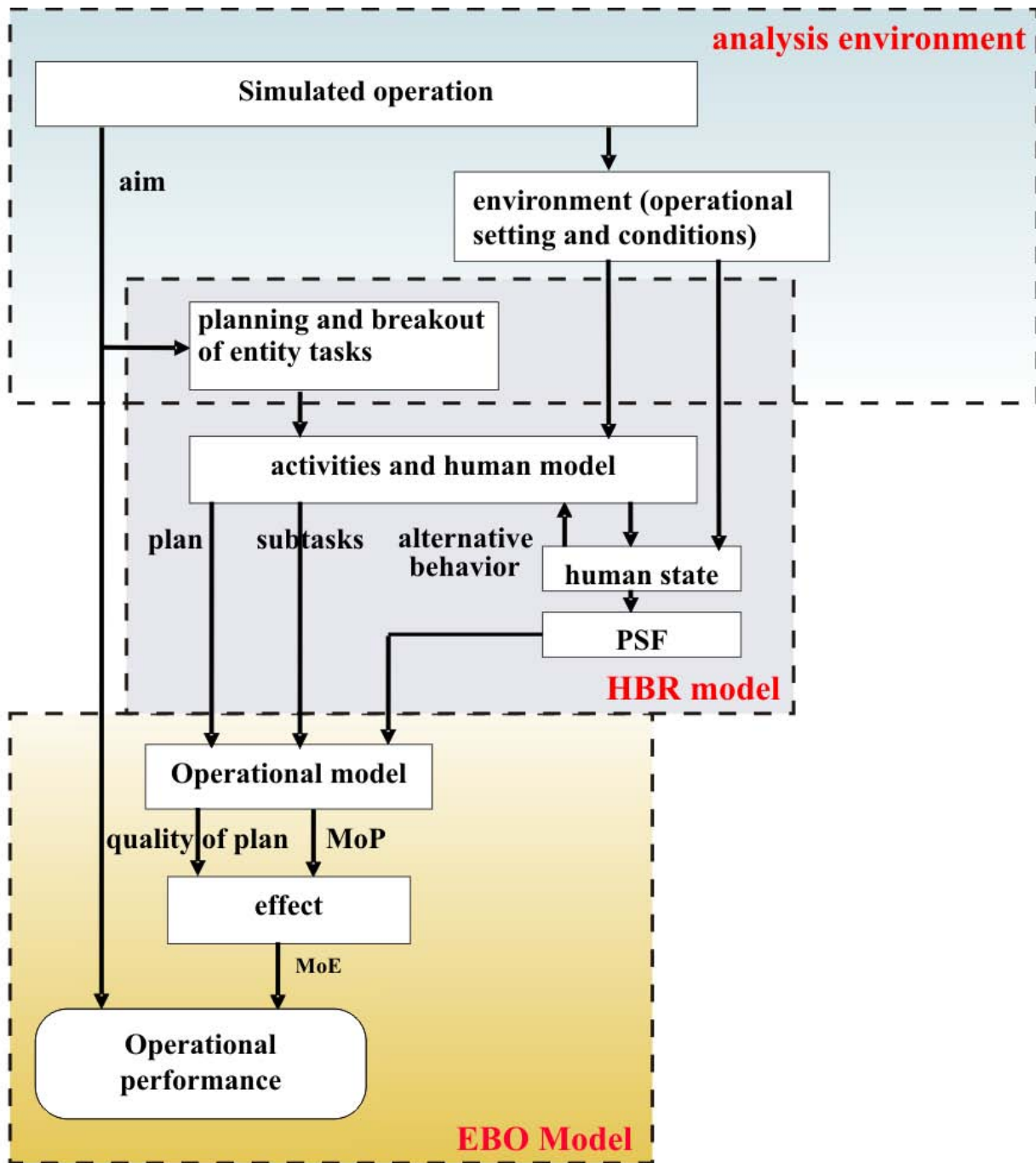


Figure 9-1: Synthesized Model Including HF in Operational Performance.

This model is an aggregation of several previous figures to show the interactions of HBR modeling with the analysis environment for Effects Based Operations. Step C of the modeling approach outlined in Chapter 9.2 may be recognized in the analysis environment, while Step D plays inside-the-box Human Behavior Representation model. Depending on the architecture, the planning of the modeled operation is done by an analyst (thus belonging to the analysis environment) or by an entity in play (thus to the HBR model).

9.4 SUITABILITY FOR APPLICATION FIELDS

The properties of a modeling environment limit the validity to certain applications only. It seems practical to rate modeling environments on three dimensions: level of fidelity, complexity, and granularity. **Fidelity** expresses the level of uncertainty in the output of the model (or conversely, the degree of match or faithfulness), with particular reference to differences (or similarities) between simulation outcomes and reality (absolute error) and differences between two simulated conditions (relative error). (Note, the outcomes of two different simulations may both have a bias, but the difference between the two should be reliable.) This dimension is also related to validity, which can be classified as content, construct, and predictive validity. (See the next chapter for a more detailed discussion of validity.) Correct predictions at the submodel level do not necessarily result in correct overall predictions since the construct may lack some essential parameters or neglect interactions. Fidelity is thus about more than just producing believable behavior. It determines for which applications the simulation model is suitable. A lower level of fidelity is probably good enough to support the development or learning of procedures compared to that needed for effective mission rehearsal. **Complexity** is a measure of the number of variables and their interactions that are used to determine the output or outcome. Handling a weapon, applying a simple rule, or training for physical endurance are described by a smaller or better documented set of parameters compared to COA evaluation, maintaining SA, or replanning, which involve many interacting parameters both at the individual level and across multiple levels of entities. Accordingly, more complex applications have an enhanced risk of error, caused both by software instability and lack of insight into the interactions among variables. **Granularity** can refer to size and organizational level of entities involved or to the time scale and detail involved. In the first case, constructive modeling of larger operations includes many entities, requiring large computational and manning resources. If a modeling architecture allows for such scaling up, that is of great value. However, larger entities are not merely aggregations of smaller entities, but have different sets of properties. Simulation models with designed to support multiple levels of granularity offer the functionality to investigate a question coarsely at higher levels and more detailed when individuals are addressed. This is useful when a high level problem, such as policy, organizational or doctrinal, needs directions for solution. The solution on this high level can then be interpreted in terms of more detailed solutions for lower levels. In the second case, granularity is interpreted as the time scale/level of detail modeled and this is where the modeler would choose between representations at the millisecond level vs. seconds, minutes, hours, days, etc. The same scaling issues apply: a simulation can be designed to handle multiple time scales or different simulations might be required, depending on the question.

In a manning application, the roles of a crew of three for a new rotary wing helicopter is studied to determine if the required tasks can be performed without mental overload and the results compared to the classical crew of four to determine if one option is better than the other, demanding a relative accuracy rather than absolute fidelity. The study indicates, however, that the task-critical issue is not having too much to do but resolving ambiguous problems. Stress thus plays an important role and consequently all the stress-coping factors need to be included on top of perceptual-cognitive variables. The number of variables increases, making it a moderately complex problem. The typical approach is to task network the total crew, involving detail of competence and resourcefulness down to a fine granularity. If the crew would have been an operations team at a warship, such as DD21, the many entities involved would make it a very large study. An alternative approach would be to reduce detail in the model in favor of quicker understanding of the key issues and then refine selected issues.

Figure 9-2 gives some estimates of modeling application fields, plotted along these dimensions. Mission rehearsal is one of the most demanding, since it involves all aspects of the operation and is intolerant for timing and resource errors. After action review will involve just as many aspects, but absolute values are not

so relevant; the emphasis is on the COA, classifying this lower on the fidelity dimension. Obviously, a study that ranks high on all dimensions is very demanding. In all likelihood, no one would do that kind of study but rather would limit the study by setting boundaries and doing consecutive, smaller studies, starting for instance with doctrine development, continuing with a manning study, and finalizing with materiel selection. These studies move down along the granularity and complexity dimensions, never demanding high fidelity.

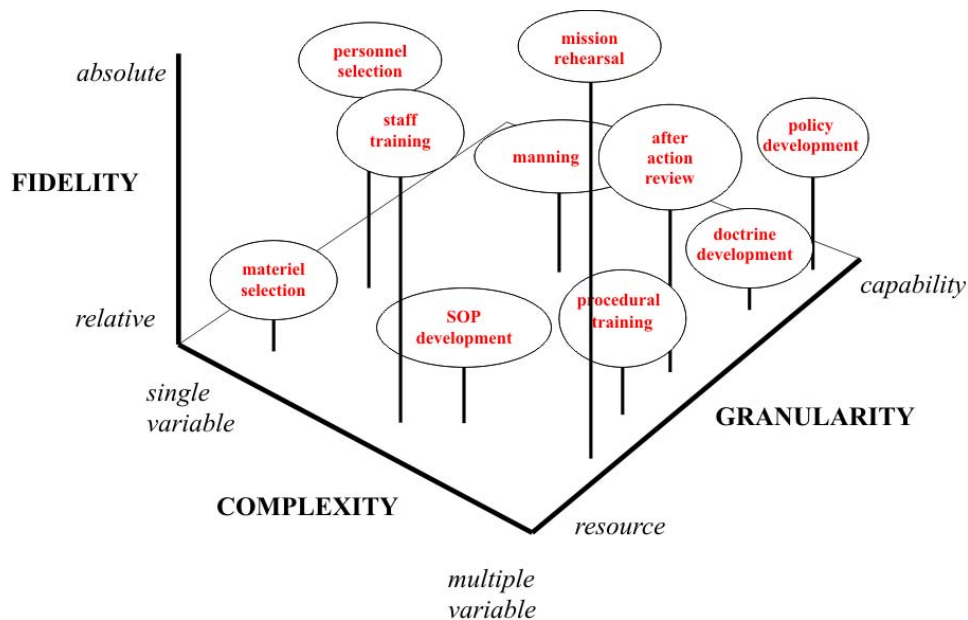


Figure 9-2: Modeling Requirements Analysis for Various Military Applications.

Positioning of application fields in a 3D space of Fidelity, Complexity and Granularity. Rarely a single model will be feasible for a wide range of applications.

9.5 OPERATIONALIZATION CONCLUSIONS

9.5.1 HBR is on its Way

The progress on incorporating HBR into military M&S has been slow but is steadily increasing. There is a small but dedicated community working on behavior representation, meeting annually at the Behavior Representation in Modeling and Simulation (BRIMS) conferences. Compared to the inventory of Pew and Mavor (1998) and the following studies (see Section 1.6), the same issues are under research, but within these issues steady improvements have been made. Representation of cognition is receiving a great deal of attention, and combined approaches have proved successful in simulating operators in rather different environments. However, the cognition models have largely been developed to solve laboratory problems rather than operational problems. The merger of cognitive, physiological and operational models is still underdeveloped.

9.5.2 Observed Weaknesses

The character of military operations has changed quickly over the past few years, posing new challenges such as accommodating the attitudes of the local population, considering means other than weapons to manage attitudes, adjusting to rapid changes in threat levels, countering the IED threat, etc. Models have not engaged

these challenges in a significant manner. Also, team processes, the building of SA and the causes or occurrence of errors are still relative weaknesses.

9.5.3 Human States and Task Performance

Modeling the effect of human state on performance is thought to improve performance prediction. State parameters represent the taxing of the human by the tasks, are measurable in the human (at least conceptually), and are well established in many human sciences domains. The transition from task and task environment to performance is divided into two parts: the first, a scientifically established relation between the task and the states; and the second, a performance shaping function relating state to task performance. Operator states are a great help when considering the cumulative effects of a multitude of tasks over time within the simulation. In the cognitive domain state variables are more difficult to define and measure.

9.5.4 New Developments in Modeling

In this report, new developments have been identified that aim at a balanced approach of operational problems, involving entities that work in concert to achieve objectives, thereby using their cognitive and physical capabilities. Effects based operations are more common now, leaving the way to achieve the aim to the unit. The unit therefore needs more information and decision skills. In simulations, this demands more representation of cognition for each entity involved. EBO also leads to new performance metrics and brings new significance to the commander's intent. Also new is the approach to explicit generation of alternative behaviors. Incorporating human capabilities, competencies, and constraints in behavior models will help to capture some of the variability observed in practice. Alternative behavior is also reflected in the variations in course of events. Better analysis of these variations may demand another statistical approach to modeling than repeated runs with randomized parameters.

9.5.5 Architectures are Still Stove Piped

The implementation of an architecture in software to capture entity behavior is a major effort. Many architectures are on the market (commercial, academic and government, see Section 2.7), but they seem to be clustered by representation approach for specific applications. An important cluster forms the predominantly cognitive architectures while another has a focus on operations. Architectures are still stove piped and many either do not permit interactions among approaches or provide awkward interface procedures to support inter-application communication. Interface technologies that support such interactions, for instance HLA, would be advantageous for providing leaky stove pipes that allow applications to excel in an area of expertise but contribute to the overall effort to produce a more effective, distributed simulation environment suitable for exploring evolving military challenges. Currently, only preliminary steps have been taken to produce a human HLA Federation Object Model that is widely accepted as being suitable for most military M&S.

9.5.6 A Recommended Practice to Involve HF

Proper involvement of HF requires careful consideration of the analysis case. A procedure has been described in this document to support the user, starting with the problem analysis, through the set up of a study and the requirements on the simulation tool to the analysis of results. Although this approach is in part common practice in HF research, it is not in many operational settings. The value of this recommendation will increase when the same sets of operational conditions and human states are used throughout simulation and experimentation. The lists of useful parameters need to grow in a natural way, rather than being forced on the community.



Chapter 10 – VERIFICATION AND VALIDATION

When a complex simulation model is constructed, it is essential that it performs in the way it was intended to perform and that the representation of the system it provides is fit for purpose. Verification is the process of checking that a simulation has been constructed as designed while validation is the process of ensuring that the design meets the simulation requirement. The application area has implications for the validation approach and the level of fidelity required. For simulations involving representations of Human Behavior both tasks are potentially complex. Verification is outlined first and validation second.

10.1 VERIFICATION

The process of verification for a simulation involving HBR is similar to that for any other complex model. The implementation of each component must be matched against the user requirements and the associated system specifications. The system specification for the HBR and its component models should be driven by the current state of human sciences knowledge (the algorithms that reflect content and construct validity) but it will typically be constrained by the user requirements, such as cost or development time. The process can be simplified by using previously verified packaged components for the HBR and matching the parameter values used in the simulation with those specified at the outset. Any interactions between packaged components may require separate verification although, since the characteristics of human performance and behavior at the lowest level are generic, it is likely that even many of the interactions may merely require checks of the parameter values employed. If the implementation matches the system specification, the model is verified.

10.2 VALIDATION

The definition of validation used by the U.S. Department of Defense Modeling and Simulation Office (DMSO Glossary (DMSO, 2006) is

The process of determining the degree to which a model or simulation is an accurate representation of the real-world from the perspective of the intended uses of the model or simulation.

but validation has been decomposed into several aspects (Anastasi, 1988; Cronbach & Meehl, 1955), each of which plays a role in determining the validity, or fitness for purpose, of a model.

In principle, a simulation can pass the predictive validation criterion if it is able to predict the pattern of real-world data that were not used to build the model. In other words, the same data cannot be used to parameterize the model **and** to be the criterion that the model must match. The downfall of this logic is that any simulation involving components of HBR could satisfy this criterion and yet be built with individual components that would not meet the target if considered in isolation. This is because there are components of human physiology and psychology that may be represented in a full HBR model that are homeostatic – provide negative feedback in control systems terms – in that they tend to restore a defined state. Since the defined state will be known, it is possible to have incorrect details in these models – in terms of open-loop properties – but the defined state is appropriately restored and in this way, the overall model appears valid, although it is incorrect in detail. A number of cognitive behavioral elements are intrinsically unobservable and a more thorough approach should be adopted based on the methods originally proposed by Cronbach and Meehl (1955) to validate psychometric models. The definitions of the three validity criteria are as follows:

VERIFICATION AND VALIDATION

- *Construct validity* is attained if the model is built using accepted theoretical constructs about how the object in question functions or accepted abstractions of the object to be modeled are deemed suitable for the intended use.
- *Content validity* is attained if the range of applicability of the model, that is, the range of independent variables and component models, meets the requirements criteria of its intended use – e.g., predictive validity covers the required range of application.
- *Predictive validity* is attained if a model is capable of reproducing real-world observations to the required degree of fidelity for the proposed application of the model.

Construct validity should be applied to each of the HBR components individually and as a whole. Typically, the construct validity of the behavioral model component involves a military SME assessment while the other aspects require a human sciences judgment on the suitability of the basis of the models. Since explicit output of many of the individual constructs is frequently unobservable in the human, the model has to be assessed as to whether it is based on best current theory. A danger here is that SMEs whose expertise is on the military mission scenario or task rather than on the psychologically based behavior run the risk of offering opinions based on no more than “That’s the way I would do it,” thus resulting in HBR components that are peculiar to an individual style or particular set of experiences. If at all possible, multiple SMEs should be queried and a consensus reached or adjustments made for inter-SME variability.

Content validity should also be applied to each of the component models separately and to the way the components interact. The key question is whether the phenomena represented by the models span the range demanded by the requirement and whether the parameters used to define the models span a plausible space of values in that context. The majority of the judgments again have to be based on SME opinion, backed by measures where they are available.

Predictive validity is tested by comparing the output of the model with real-world observations. If possible, experimental observations should be used to make the comparison. Subjective judgment is likely to be less stringent than statistical comparison with data, although in extremis SME opinion may have to be accepted. Since there is negative feedback likely to be present in the simulation, it is important that more than one output measure is used for comparison since one measure can match observed behavior when it is composed of incorrect component models.

Predictive validity remains the Achilles heel of modeling. It is difficult to validate a model that includes human behavior comprehensively because of the complexity and interactions described in this report. Experimental data are required to match the model data, but it may be the case that the system being modeled does not yet exist, or even if it does, it may not be feasible to obtain sufficient data. Still, the pursuit and documentation of predictive validity is of prime importance. Model validation is the foundation for being able to make strong statements and generalizations about performance in practice based on model outcomes.

There is considerable debate at the moment in the literature about the precise approach that should be used to assess the degree of conformance of any model with real-world observations (Robinson, Duursma & Marshall, 2005). The notion of performing a standard statistical test as to whether the model and reality conform is regarded as insufficient. For example, using a standard set of observations with a common procedure to assess model validity will yield the same answer for all applications, but the single “YES/NO” answer may be appropriate for one model application but inappropriate for another. The core of the problem is defining the required level of conformance between model and reality for different applications. For many applications, it is sufficient that the *nature* of the phenomena predicted by the model conforms to observation

but inaccuracy can be tolerated in the *magnitude* of the predictions. For example for some applications, it might be important that a model including HBR forecasts that a soldier will engage an enemy target, but it does not need to forecast how rapidly the engagement occurs with great precision. For other applications, the speed of human response may need to be forecast with great precision. Any statistical procedure used to assess predictive validity should be based on the needs of the current application. It is unlikely that the application of a single generic measure such as Pearson correlation coefficient between observed and predicted values will provide an appropriate assessment, and the different requirements for model conformance should be reflected in different statistical assessments. A detailed discussion of the issues that arise in the validation of HBR can be found on the U.S. DMSO web site¹.

The key message is that before predictive validity is assessed for a particular application, it is important that criteria for model conformance with reality are considered carefully so that an appropriate assessment is made.

10.3 MODEL COMPONENTS

Based on the requirements of both verification and validation, it is highly desirable that the complete simulation is built from individually verified and validated HBR components. This reduces the complexity of verifying model components and establishing construct and content validity at the component level. It does not remove all the questions since different simulations are likely to employ different combinations of basic model components, including moderator models. While the representation of specific tasks is quite likely to be simulation specific, a large part of the relationship between external and internal moderators and internal states is generic and susceptible to generic verification and validation.

10.4 VALIDATION AND STANDARDIZATION

As outlined in preceding paragraphs, full validation for a simulation involving HBR can be a costly and time-consuming undertaking. It has been observed (Gluck, 2006) that relatively few studies invest in validation since it may comprise 30% of the cost of the study and is perceived to be research and development rather than application. As a result, potential consumers of HBR, in rather circular logic, argue that validation activities fail to add value since it cannot be demonstrated that the results do not directly resolve the application question as posed. Routine validation of HBR models is at best extremely limited and at worst, non-existent.

A possible way forward is to focus modeling effort on the development of standardized components that are susceptible to validation at an acceptable level individually. This is clearly applicable to at least part of the problem of modeling moderators and can be applied to a number of aspects of more general task performance and behavior. The development of this approach clearly favors architectures in which it is easy to verify that a standardized model component has been employed and to identify the parameter set that is varied and the values adopted. Examples of this approach include theory-based architectures (e.g., ACT-R: Anderson et al., 2004), and the micro-model based tools (e.g., GOMS and its various forms: John & Kieras, 1996) as well as those that trace back to Human Operator Simulator (HOS: Wherry, 1976) micro-models including MIDAS, IMPRINT, and IPME. The theory-based architectures have been built up essentially by recursive experimentation, model building, data fitting, and finally prediction, thus providing a validated foundation for further models using that architecture. The basic notion of micro-models is that robust performance data

¹ http://vva.dmsomil/Special_Topics/HBR-Validation/hbr-validation.htm

VERIFICATION AND VALIDATION

(i.e., time and accuracy) are gleaned from the empirical psychological literature and made available to models in a “plug-and-play” or “copy-and-paste” fashion, perhaps as averages with associated variability. In the case of moderators, those established baseline figures are varied as a function of some other factor such as environmental temperature, individual aptitude, or situational stress also as supported by empirical research. In some cases it might be sufficient to start with validated data, say for a fairly simple model or a model of human-system performance where the system has not yet been built. Usually, though, a better test is validation of an assembled model against data collected from actual human-system performance.

One difficulty with such standardization is that it can freeze development at a time when there remains relative uncertainty in the scientific community at large about which model form is most appropriate for a particular component. At the moment, the community appears to be most comfortable with a path that leads to higher costs of application, in that new model frameworks are developed to meet new problems, so that scientific evolution can continue. The customers for HBR bear this additional cost and the lack of standards may be holding back the widespread application of HBR to military problems. The criticality of the timing of standards was emphasized by Pew (Allender et al., 2006) who urged the community to not wait for formal standards developed only to satisfy software interoperability requirements. Further, Pew recommended maintaining and exploring a variety of HBR architectures rather than adopting a *defacto* standard since this is still an area for scientific research rather than merely applied technology.

A second difficulty with component standardization is that the validity may or may not hold when components are arbitrarily put together. Assumptions and tests for a single component may require additional validation to test for additivity, interactions, etc. It is also the case that when components have been validated within different theoretical frameworks, combining them should not be done haphazardly. Additional validation may be required.

It is worth mentioning that there is a growing number of cases where model validation has been accomplished, or attempted with some reasonable success. Allender et al. (1995) reported on a verification, validation, and accreditation (VV&A) effort, where accreditation is a further step of obtaining approval from an authoritative office, a “sign-off-on,” as it were. The subject of the VV&A was the set of tools called HARDMAN III (Hardware vs. Manpower III), the precursor to the IMPRINT tool. The validation effort relied heavily on construct and content validity and somewhat less on predictive validity but used additional factors that were deemed to add to the assessment of being “fit for purpose” in the larger, applied sense, that is, worthy of being accredited. Those additional criteria included establishment of software configuration management procedures, availability of documentation, detailed data definitions, appropriate level of detail or model granularity, suitability of output reports, and ability to support timely analysis and decision making requirements. This same tool set was subsequently used as a part of a formal military operational test where the model was used to structure the test conditions based on predicted operator performance and workload and the operational test results were then used to validate the model predictions (McMahon, Spencer & Thornton, 1995).

More recently in the Gluck and Pew (2005) report of the AMBR project, considerable attention is paid to validation, especially the chapter by Campbell and Bolton (2005). Two points are of interest here. For one, for predictive validity, the classic goodness-of-fit measures may be too restrictive for many complex HBR-based models. Other measures such as pattern matching, *a priori* predictions and comparisons with other models may be suitable for many applied validation cases. Second, they make the distinction between validating an architecture versus a model built with an architecture and that validating one does not constitute validation of the other. Indeed, it may not be possible to formally validate an architecture and we may have to accept or reject architectures based on the confidence gained through application to a wide range of problems.

The same distinction applies to tools and models built with those tools, all of which means that the validation requirement, indeed the challenge, does not go away. On the plus side, however, is the fact that validation can accumulate, so-to-speak, and that previous validation efforts can serve as a precedent, a history for the next.

10.5 STATISTICAL METHODS IN EXPERIMENTATION AND SIMULATION

As noted in Chapter 6 on modeling the effect of moderators, there is inherent variability in both human behavior and performance. The sources of variability are both inter- and intra-individual, and these distinct sources have to be recognized in any statistical analysis of observations. These properties must be recognized and represented in any modeling that involves HBR. It also follows that if the stochastic properties of human behavior are embodied in a simulation, Monte Carlo methods must be used to generate the simulation results, and appropriate statistical methods must be used to investigate the simulation findings. There is a substantial literature, including many well-known text books, that describe the analysis of human experiments and this literature may be used as reference for the approach to be adopted to the analysis of Monte Carlo trials.

It is not appropriate to provide detailed advice on the methods to be used for undertaking statistical analysis in this report. Some simple observations can be made:

- 1) The methods that are used to investigate observations drawn from human trials should be applied to the analysis of simulations involving HBR;
- 2) The methods used to design experiments involving human subjects can be applied to the design of simulation experiments involving HBR; and
- 3) The statistical principles that underlie the analysis of human experiments should be applied to the problem of comparing the outcomes from a model with observations.



Chapter 11 – FUTURE CHALLENGES

11.1 SCALABILITY, AGGREGATION, AND LEVELS OF ANALYSIS

The larger part of the discussion in Part Two of this document has focused on how to model the behavior of individuals. In the majority of practical applications, it is required that the behavior of an aggregate of individuals is required at the platform level, the level of a formation such as a battalion or brigade or possibly at a higher level than that. It is rarely practicable to model the aggregate by representing each individual within it and describing the interactions between individuals to arrive at aggregate behavior, and it is necessary in practice to model at some aggregate level. Based on the discussions in the chapters in Part Two, it is clear that aggregating individuals into organized teams in platforms and at a higher levels alters the relative importance of the different elements that are required to represent an individual. The effect of individual differences – the internal moderators – tends to be smoothed out in an aggregate, but the effects of external moderators remain unchanged; interactions between individuals become important since they drive emergent behavior, but decisions made by commanders tend to be representative of individual decisions. It is therefore necessary to consider carefully how the factors described in Part Two should be incorporated in aggregate simulations; extra attention must be paid to the impact of the interactions between individuals that are represented in team constructs such as cohesion, communication, and backing up, and the effect of external stressors must be addressed carefully.

Although neither Colonel Schmidt's precise scenario nor Lieutenant Anderson's exact mission had been modeled, a series of M&S exercises predicting operational outcomes had been conducted. HBR modelers worked with large-scale M&S developers to make sure that the appropriate granularity was incorporated. They were able to draw from a database of various levels of HBRs, from individual Soldiers and small teams that could act with a large degree of autonomy in simulations, to brigade- and battalion-sized forces that required command intervention by personnel. Likewise they were able to select from a database of standard tasks including navigation and shooting, to reconnaissance and higher level decision making, all drawing on component models of human perception and cognition to generate plausible actions given uncertainties in the simulation rather than idealistic responses that arise from being aware of everything at every moment. In this way, HBR helped to inform the trade-offs that must occur in the application of effects-based tactics and force employment against the higher level metric of operational effectiveness.

There is a difference between the approach that can be used to represent structured military entities with well-defined internal organization and that employed for the description of civilian society. The special problems of societal modeling are discussed in Section 11.2.

This report has discussed a wide range of aspects of representing human behavior in constructive simulations, including the effects of moderators of all kinds (external and internal), the principles of cognition, representation of error and the description of basic motor activity. There are architectures in which a number of these elements have been captured and in which they can be represented relatively conveniently for the user, although there is no single architecture in which all have been captured and implemented. It is unusual to require the complete gamut of human behavior representation in a single simulation, and this characteristic implies that a modular architecture is the key to the development.

The challenge for the HBR community is to develop one or more modular architectures that enable developments ranging from free-standing agents that incorporate all the elements described in this report to

representations of complex crews with a coarser granularity and that incorporate a range of human behavior representations suitable for aggregated entities.

11.2 SOCIETY LEVEL MODELING

11.2.1 Introduction

The underlying idea of using detailed scientific knowledge in theater-level simulations is that aggregate models are more valid and reliable than separate models on different unit levels. The main issue behind aggregate modeling is to balance the need for precision and accuracy of HF with the operational needs for a specifically sized unit in mind. Another issue of concern is the extent to which HF knowledge that is scientifically sound but limited in scope (at the individual level) can be combined in large scale simulations and thereby become operationally relevant.

With growing interest for EBO and DIME-PMESII (Diplomatic, Informational, Military and Economic actions to accomplish Political, Military, Economic, Social, Infrastructure, and Informational effects) there is however a need for even larger scale level simulations. This need deals with human characteristics and behavior of populations on society level (roughly speaking: hundreds to hundreds of thousands). It also implies another time frame of interest: weeks to years rather than seconds to hours. The balance of detail and operational needs, as well as the issue of aggregating effects is just as important in these simulations. In addition, emergent rather than command-chain-decision effects and behaviors become more relevant, especially in the contexts of civilian interpreted laws and norms (Holland, 2008).

11.2.2 From Individuals to Societies

Before we consider specific operational problems and needs for society level models and simulations, let us approach the problem from an individual HF modeling standpoint. From a physical perspective, one could argue that because all behavior is generated in a human body and the natural unit is the individual, HBRs should intrinsically include this representational level to build upon. The naïve suggestion would then be that the step from representing “one man using a weapon” to “one thousand men using their weapons” is simply a matter of multiplication. Even if generating large numbers of individuals characterized by random values on individual attributes according to normal distributions, it would not automatically include the structural aspects and interactions of a population of agents. Although such interactions conceptually increase geometrically in number with population size, the interactions are likely restricted to “neighboring” entities. A lesser concern with this approach is that there may be computational limitations in multiplying software agents to represent city or country populations.

Conversely, one could suggest usage of rule-based and task-based architectures (e.g., Soar, ACT-R, IPME) to directly create “top-down, group-agents” simply by changing the level of abstraction (one agent representing an arbitrary number of individuals). While some concepts may find a useful reference (e.g., a society’s long-term memory) other do not (e.g., anthropometry) and built-in HF measurements, such as neuron activation levels or reaction times, would be rendered meaningless.

New frameworks with appropriate concepts and processes will be required for society-level modeling in constructive simulations. The U.S. National Research Council has identified gaps in the IOS (Individual, Organizational, and Societal) research domain and to meet these challenges, new research programs have been initiated with topics such as model federation, theory development, dynamic adaptability, and new data collection methods (Zacharias, MacMillan J. & Van Hemel, 2008).

Approaching something as abstract as culture and society there should also be some initial limitations to what phenomena we are interested in and what the time frames are for natural variation in these phenomena. Figure 11-1 illustrates different time frames of cognitive or task modeling and the modeling of societies.

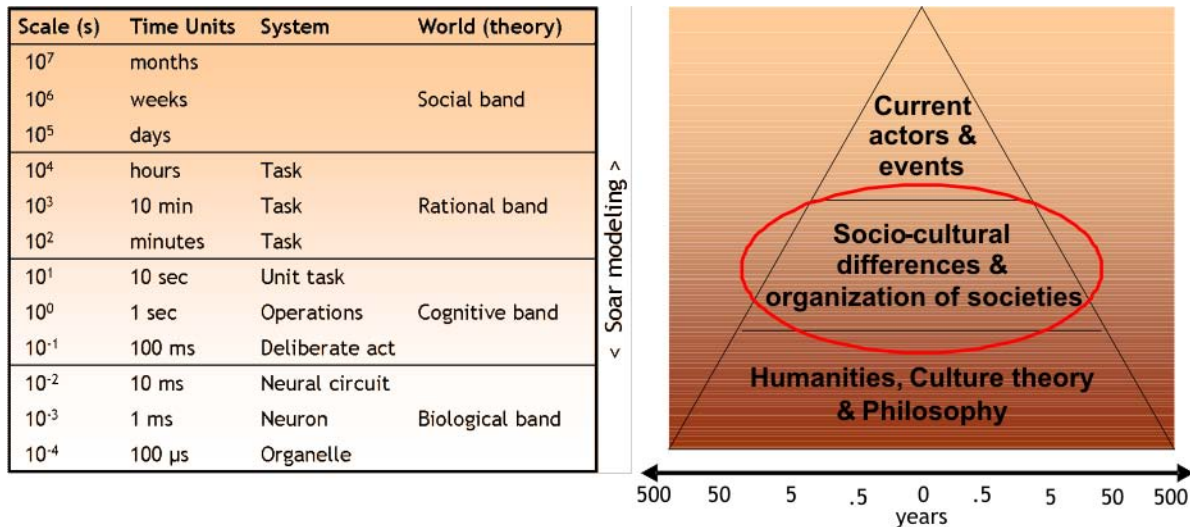


Figure 11-1: Time Scales of Human Action (left) and Culture-Related Research Areas (right).

Appropriate time frame for cognitive models in Soar (Newell, 1990) is milliseconds to hours, which could be contrasted with the shaping of institutions and foundations characterizing different societies, developed over decades to hundreds of years. Modeling the socio-cultural differences in operational areas worldwide (circled in red), we would likely be able to make better political-strategic and military-strategic decisions. However, common population behavior could occur in the top of both figures (days to years). Both of the figures imply that an upper level segment builds upon characteristics of a lower level segment.

It is evident that modeling individual HF and modeling human societies are different research domains but with intersecting and shared interests. A model of opinion dissemination on a society level may benefit from a more detailed individual model of cognitive dissonance (Bertie, Himmelweit & Trigg, 2006). Likewise, cognitive models could operate on a social environment just as the natural phenomena they attempt to represent. The decision making model of our Colonel Schmidt in the scenario storyline of this document would benefit from input from large-scale opinion dissemination models describing attitudes both in the area of operation and in those cultures present in the coalition.

11.2.3 Operational Needs

In traditional military operations, symmetric as well as asymmetric, it is of highest importance to be aware of the changing character of the opponent in order to win the war (“Know your enemy” – Sun Tzu). This includes military assets and tactics but also the population mind set and culture, which can be targeted with both kinetic and psychological means. However, the main need for society-level models would be in MOOTW (Military Operations Other Than War) or in the post-combat phases of war, where the people’s hearts and the minds must be addressed in order to win the peace.

MOOTW include operations such as peacekeeping, support as well as targeting of rebels, police work, and humanitarian efforts. Typically MOOTWs require restraint through restrictive and complex rules of

FUTURE CHALLENGES

engagement, a clear political and legitimate context, and years of commitment to reap positive results. Problems and needs in these missions could include:

- Change critical opinions through non-kinetic means such as campaigns and policies;
- Maintain and create new relationships with culturally very different counterparts through actions from soldier face-to-face interaction to overall foreign policy;
- Use currently available channels and develop new channels for communication to different segments of populations;
- “Soften” and “work up” opponents or other actors to achieve effects such as trust and openness;
- Avoid situations where conflicts, segregations or “perceived unfair circumstances” exist between different groups in the area of operations; and
- Police violent behaviors, organized crime and radical ideologies that conflict with the mission’s purpose.

A thorough understanding of the local social and political environment in the area of operation is required at a strategic level to meet these objectives. The time frames, “cultural presets” and opportunities for change in behaviors and attitudes on populations will vary greatly in different parts of the world. On the operational level, planning and execution of operations must support the strategic mission goals, occurring in collaboration with potential kinetic operations, and take present local circumstances into account.

11.2.4 Subject Matter Domains

There are two main issues which must be resolved in order to provide useful long-, mid- and short-term decision support and training for military and civilian personnel involved in foreign operations:

- What type of conceptual models, computational models and simulations support the operational needs?
- What data sources should be used for building such models and simulations?

Clearly, the challenges are enormous in an historic perspective, and the gap of knowledge must be met on many fronts.

Society-level modeling should take a multi-disciplinary research approach, just as HF modeling has its natural empirical counterpart in multiple research domains (e.g., psychology, ergonomics, and physiology). A list of subject matter domains that, at least on the surface, would seem to provide useful input for society level models is as follows:

- **Anthropology** – Especially the area of anthropology dealing with national and regional differences could feed stereotype models with datasets from factors such as honor, face, taboos and traditions. One example of typical cultural behaviors is the agent-based model describing the Afghan social structure qualm (Geller & Moss, 2007).
- **Economics** – In particular, Keynesian macro-economic models describing growth, inflation, unemployment and the influence of national economic policies. Economic models are of concern for several reasons: Measures such as Gross Domestic Product (GDP) may serve as an indicator of economic health and percentage of GDP on military expenses as an indicator of hostility or perceived threat. Furthermore, well-known factors for militant recruiting are poverty and unemployment; hyper-inflation is often associated with war or its aftermath.

- **Multi-Agent Systems** – These agent-based frameworks allow for individual- (micro), social- (meso), and society- (macro) level modeling. In particular, one can demonstrate macro-level effects as a simulated consequence of changes in micro-level rules, which, for instance, could provide an answer to the question, “Will more interpersonal meetings lead to a less segregated society?” (Axelrod, 1997) Examples of multi-agent system architectures are: Repast Symphony, Ascape, and MASON.
- **Marketing, Advertising, and Media** – In developing psychological operations (PSYOPS), such as leaflets or internet advertising for different cultural domains, techniques and knowledge for catching attention and using appropriate cultural values may be copied directly from the corporate world. Also of interest are techniques for characterizing target audience segments such as the media dissemination models developed by Gonzalez-Avella et al. (2007) and Bertie et al. (2006).
- **Organizational Science** – Characterization of groups and leadership in political as well as corporate organizations may be appropriate for conceptual models. Karabaich (2002) provides one approach for a taxonomy in this area. Likewise, the idea of system thinking from the corporate world may be helpful in visualizing society-level influences, providing cues for common recurring themes through archetypes (Senge, Roberts, Ross, Smith & Kleiner, 1994).
- **Political Science** – In this very broad field, areas of interest range from basic knowledge of political systems, institutions, and power bases to diplomacy, conflict resolution, and political psychology. Taylor, Waltz, and White (Taylor, Waltz & White, 2008) provide a recent simulation framework for modeling power structures and transactions of power between actors.
- **Religion and Ideology** – A thorough understanding of religious and other ideologies would help us understand the logic, motivations, and norms of different societies. Limiting sectarian violence (as seen in the Middle East) and supporting beneficial religious tenets (such as the haram of growing opium) through information operations may be more efficient than kinetic operations. Makowsky (2007) provides an example of a computational model of the utility of extremist religious membership.
- **Social Psychology** – When modeling small groups (in this context 10-100 entities) social factors such as group dynamics, team performance and crowd behaviors may be relevant to operational areas. FIRO (Fundamental Interpersonal Relations Orientations: Schutz, 1958) is a famous conceptual model of group dynamics, and the work of Nguyen et al. (2005) is an example of a computational crowd model.
- **Sociology** – Of particular interest are ideological movements and influence of ideas “thought contagion processes” within societies. Oliver and Myers (2000) provide examples of empirically based conceptual models in this field. Demographics and other population statistics data sources may be of value for both model construction and validation. Sznajd-Weron and Sznajd’s (2000) work is an example of non-empirical but computational model of opinion evolution within a closed community.

With models and datasets from each of these domains, gaps can be filled and contextual pictures will be more complete. Application areas for models of societies range from cross-cultural training to operations planning and intelligence gathering. Computational models within all these research domains fall in the middle block of Figure 11-1. These models may build upon historical and theoretical concepts of culture (lower block) but may not investigate them per se. Further, it may be possible to statistically infer characteristics for current actors and events (upper block) from data sources and models within the listed domains.

11.3 DATA ACQUISITION

Underlying even the simplest model is a set of data that defines the performance characteristics of the entity being modeled. In contrast to models of physical systems where the associated performance characteristics are often well understood or even documented in detail and easily incorporated into computational models, HBR modelers are faced with the challenging task of obtaining and incorporating data related to aspects of the human system that may be ill defined or where incomplete or minimal to no data exist. The depth and breadth of available data in the realm of human performance is further exacerbated by the range of pertinent behavioral characteristics of interest to modelers from microscopic low-level perceptual processes to the macroscopic elements of social-cultural behavior. Indeed, as one expands the scope of interest from the perceptual to the socio-cultural, not only do the quality and availability of data become scarce, but computational theories from which data could potentially be derived are less prevalent.

The HBR practitioner is then faced with a limited set of options as to how best address the issue of data availability including:

- 1) The conduct of human experimentation to obtain data where no previous data exist;
- 2) The identification and application of existing HF datasets contained within the relevant literature; and
- 3) Derivation of data based on sound theory and practice.

Obtaining data via human experimentation or the observation of human activity is generally a time-consuming and costly process and is often difficult to incorporate within the context of a typical modeling effort. However, if capturing human performance data is required, efforts to reduce the associated costs involved with this approach can be made, including piggy-backing data collection activities during the operational or simulated use of a system in conjunction with the use of automated data collection and performance monitoring devices tied to instrumented systems.

To reduce the time associated with identifying existing human performance datasets, the concept of establishing large repositories of human performance data that would be available to the modeling community as a whole appears as an initially attractive option. However, apart from housing the most elemental of behaviors at the keystroke level, the sheer scope associated with the myriad combinations of human behavior of interest to modelers makes this approach quickly intractable, at least within a single database. The best approach to supporting easily accessible data is the establishment of domain-specific databases that are tied to specific levels of analysis. In conjunction with purely information databases, the HBR community at large should be encouraged to continually produce and publish open-source micro-models that are based on existing HF knowledge such as those available from (Boff et al., 1986a; Boff, Kaufman & Thomas, 1986b; Boff & Lincoln, 1988), making them modular and suitable for use in military M&S at various levels of detail.

11.4 CURRENT SIGNIFICANT GAPS

This report has described the requirements for the development of HBR for military simulations. As part of the discussion, the current state of the art has been described. There is considerable experience in the HBR community in the representation of structured military teams as crews of military platforms. Structured military teams frequently involve the representation of a relatively restricted range of behaviors in a constrained context, and this can be accomplished in a range of architectures. There are shortcomings in almost all of the areas described: Perception, Cognition, Motor Tasks and Errors, although there is considerable research activity in the NATO nations. The area that is least developed at the individual level is

the representation of emotion as a component of the model. This is frequently unimportant for the representation of the behavior of members of military teams but is highly relevant for non-military players in Peace Support operations.

The representation of loosely structured groups in simulations to capture the behavior of civilian elements demands a more substantial repertoire of behaviors than is exhibited by the structured military team. This presents a significant challenge to the development of general purpose architectures.

11.5 RAPID PROTOTYPING

This report has provided a general overview of the needs of HBR, taking into account the full gamut of influences ranging from perception and cognition through internal and external moderators to critical aspects of human error. It is clear from the description of the phenomena that implementation of an HBR that takes account of just critical elements of a model is a significant undertaking. As a result there are significant costs associated with the development of even relatively simple HBRs from scratch. The previous section has discussed the problems of data acquisition to support the development of HBR and proposed that repositories of micro-models could provide a key element of cost reduction if implemented in the appropriate architecture.

In Chapter 6, the problems of modeling moderators are described, and the concept of internal state is introduced. It was noted that the relationship between internal and external moderators and internal state is generic in that it applies in all modeling contexts. As a result these elements of HBR are suitable for incorporation in standard libraries and can be re-used with confidence if the relationship between moderator and state has passed verification and validation. The second part of the problem, typified by the relationship between internal state and performance or behavior, cannot be generalized to the same degree. To render the relationship between state and performance generic, both IMPRINT and IPME use a taxonomy of tasks to enable the construction of generic PSFs that can be used to represent this relationship in all applications (Belyavin, 1999). There are clear cost reductions implicit in the use of generic models to represent the effects of moderators on performance and behavior in a wide range of internal and external conditions. The use of such frameworks does not eliminate the need for data acquisition when building a new HBR, since any activity must still be allocated to the taxonomy and this requires at the least an SME assessment.

The key challenge is to develop architectures that at the least embody those elements of HBR that are generic in the form of readily accessible libraries of models. In those cases where data are needed to support an engineering model, the most cost-effective approach is to develop an overarching modeling approach that reduces the need for data collection to support a specific analysis by providing a generalization. A case in point is the model library embedded in the U.S. Army's IMPRINT tool, which began as HARDMAN III (Hardware vs. Manpower III). The original concept for HARDMAN III was to maintain a library of task-level models for all major fielded systems. To that end, task network models were built for 22 systems using test data, results of experiments, SMEs, etc. However, the models were out-of-date or out-of-synch almost immediately due to small changes in system configuration, upgraded components, or software updates that were substantial enough to require changes to the task network model. Maintaining this sort of library would be quite costly. Instead, the common practice has been to use the library models to "jump start" new modeling endeavors. The library models have been used as baseline models and as examples of how to construct certain kinds of models. They have also been used as sources from which to cut portions and to paste into a new model. This fits with the more general goal of model reuse, which can refer to both the HBR model structure and the embedded data.

FUTURE CHALLENGES

Derivation of data is required for virtually all HBR efforts when the needed data do not exist for the precise mission, environment, and set of conditions to be modeled. Basing the data on sound theory and practice should go without saying, but, of course, all too often in practice, rough guesses are used. The cognitive architectures – such as ACT-R – show considerable promise in terms of the capacity to generalize and avoid rough guesses but still demand considerable data input and software development. The strength of the approach outlined in this document is that when the fundamentals of human perception and cognition are used as the basis for HBR, it naturally gives rise to performance moderated by various stressors and performance and the associated errors. It is also the case that for HBR on the order of milliseconds to minutes, many modelers have found that the data for larger tasks can be “constructed” out of smaller or “micro” tasks. The challenge is to blend the capacity to generalize provided by cognitive architectures with the convenience of the engineering approach embodied in task network models or simple software agents, thus producing a cost-effective architecture for the development of HBR. This was the basis for efforts such as HOS (HOS: Wherry, 1976), GOMS (GOMS: Card et al., 1983), or more recently, embedding a cognitive architecture within a task model (Warwick et al., 2008).

Chapter 12 – CONCLUSIONS AND RECOMMENDATIONS

This report crosses the domains of science, operations and research policy. The HFM-128 group drew a large number of conclusions, addressing issues in each of the three groups. The recommendations are thus not only to NATO, as is usual, but also to the scientific community and the operational modeling professionals. From these conclusions the most significant recommendations were distilled and grouped around six issues.

12.1 MODELING PRACTICE

12.1.1 Conclusions

- 1) There is an extensive body of literature on human performance and behavior that can be applied to the development of models and agents, although the data are often not represented as formal computational models. The data should be exploited for the development of military M&S of human behavior.
- 2) The current HF knowledge is insufficiently exploited and can be improved using the good practice outlined in this report.
- 3) There is a need to control the complexity of a study when a larger operational scale is simulated both by aggregating multiple units into a larger unit and by making the assertion that some phenomena appear the same at all scales.
- 4) The aggregation of the behavior of multiple units into collective behavior needs further development.
- 5) The extension of the EBO concept to lower organizational levels and the individual level representations increases the demand for more realistic cognitive models in associated entities.
- 6) Representation of non-military personnel in simulations is increasingly needed.

12.1.2 Recommendations to the Scientific Community

- 1) Handle scientific Human Factors concepts (perception, cognition, etc.) as well as more macro, operational concepts (decision making, thinking, etc.) concurrently to ensure that the scientific concepts support the operational community and that the need for specific scientific concepts is understood by the operational community.
- 2) Develop compact, task-specific decision making process models and algorithms for individual entities.
- 3) Develop aggregation methods from individual HF to team and collective HF.

12.1.3 Recommendations to NATO

- 1) Specify the variety of entities involved in up to date operations and their specific behaviors.

CONCLUSIONS AND RECOMMENDATIONS

12.2 MODULARITY

12.2.1 Conclusions

- 1) Modular, composable approaches to model development should be adopted (as suggested by Pew & Mavor, 1998) so that various elements can be added and removed as required without compromising the performance of the remaining elements of the model.
- 2) Recognition of internal states at individual and collective levels enables the development of a modular approach by separating the effect of moderators on state from the impact of state on performance and behavior, enabling the development of sophisticated hybrid modular architectures.
- 3) The framework of state parameters is the blueprint for input-output compatibility between modules and must include flexibility to connect more detailed with more coarse modules.

12.2.2 Recommendations to the Scientific Community

- 1) Develop methods that capture cause and effect relationships or a more complete set of the factors that affect operational behavioral choices, relating these factors to identified state parameters. Identify state parameters in the cognitive and emotional domains.
- 2) Analyze self-pacing strategies and behavioral choices for their effects on task performance.
- 3) Develop generic state to performance predictions (i.e., PSFs) that can be used in a variety of settings and evaluate the extent to which task independent PSFs may be applied to HBR.
- 4) Study complex combinations of performance moderators in the laboratory and the field in order to improve our understanding of how they affect human decision making.

12.2.3 Recommendations to NATO

- 1) Promote and encourage bottom-up initiatives to converge to selected human state parameters for use in modeling and for measurement during operations alike. Expand the list of parameters to encompass greater resolution for a more complete and unified representation of the human.

12.3 VALIDITY

12.3.1 Conclusions

- 1) There is currently no common way to implement, measure and validate human behavior in models. In some models, behavior emerges from internal states, whereas in others, production rules are defined in the scenario. We are in need of robust, human centric approaches for representing human behavior in military simulations.
- 2) Current constructive entities are often brittle, showing unrealistic behavior for even slight departures from the design space due to an over-reliance on simple, rigid rule sets and strict behavioral templates that capture SOPs and idealized, purely doctrinal behavior.
- 3) Validity on the construct and predictive levels is often weak because of the prohibitive cost of extensive validation and because of limitations in experimentation with threatening conditions.

- 4) It is expensive to gather data to drive even simple agents so it is important to reduce the cost of data acquisition. Data acquisition during real operations is of prime value for providing context to laboratory and training data.
- 5) The development of metrics for individual, team and operational performance is a condition for improving validation.

12.3.2 Recommendations to the Scientific Community

- 1) In any modeling effort, show transparently the content and construct, taking the fitness for purpose into account. Make a clear distinction between scientifically based and engineering constructs.
- 2) Provide convincing rationale to sponsoring agencies to provide funding for predictive validity studies.

12.3.3 Recommendations to NATO

- 1) Promote methods to collect data in the operational theater following a standard protocol with selected MoPs and MoEs, including human behavior. Make these data available to the community.
- 2) Encourage all member nations to establish human behavior representation model validation policies and requirements as default components of their research and acquisition investments.
- 3) Set minimum demands on the validity of modeling and simulation efforts published in NATO communications.

12.4 HUMAN ERROR

12.4.1 Conclusions

- 1) There are many references in the literature on the incidence of various types of error and the contributing factors that affect error rates, but little has made it into HBR and agent applications.
- 2) Human error may induce alternative Courses of Actions (COAs) or actions and reactions that result in dramatically different outcomes. In the evaluation both the probability of the outcome and the operational risk must be assessed.

12.4.2 Recommendations to the Scientific Community

- 1) Investigate applying the conditional probabilities on error in specific tasks typically collected by the safety community in conjunction with data on decision making and choice, to determine the ways in which missing or false information may lead to erroneous situation awareness.

12.4.3 Recommendations to NATO

- 1) Start a study on the incorporation of human error models in constructive simulation.

CONCLUSIONS AND RECOMMENDATIONS

12.5 INTEROPERABILITY

12.5.1 Conclusions

- 1) Integration with other M&S representing the mission and analysis environment is both a software specification challenge and an application challenge, involving the incorporation of the mission brief into an agent's goal and knowledge structure as well as involving the creation of a framework that supports interactions among agents.
- 2) HBRs are important for role playing to reduce personnel demands. We will need a basic structural framework that incorporates physics and physiology models with psychology and social models.
- 3) The interfaces between modules define the way information is expressed, processed, and shared in the modeling architecture. Information may be deduced by sense making from (multi-modal) perceptual details or be an attribute of a detected object. It may be expressed as a verbal code or shared in symbolic form. The challenge is to design interfaces that can work with modules of mixed resolution.

12.5.2 Recommendations to the Scientific Community

- 1) Experiment with approaches to interoperability of model modules, performing generic human functions or relating human states to performance, of varying resolution at the input/output level.

12.5.3 Recommendations to NATO

- 1) Start a multidisciplinary study to standardize minimum human model requirements.
- 2) Reopen the discussion on defining a human FOM (Federation Object Model) that supports integration of individual agents in constructive simulation. Develop a model interface environment concept for reuse of models that is consistent with commonly used military simulation environments.

12.6 FUNCTIONAL EXTENSIONS

12.6.1 Conclusions

- 1) There is a need for models of teams and crowds that include the appropriate states and moderators. These models would describe the loose associations of individuals involved.
- 2) Increasingly, the effect of military actions on emotions and attitudes of neutral and hostile parties is considered a significant military effect. Emotions and social dynamics are currently underrepresented in HBR.
- 3) The role of motivation in behavioral choices is currently underexposed. Motivation does not only pertain to the task but in the form of values also to alternative behaviors. We should have a motivational framework that incorporates affective and emotional elements, moderating behavior.
- 4) HBRs need not represent all aspects as psychological process models. Although building models with a theoretical foundation is likely to enhance their generality, empirically based models may be suitable for some phenomena.

- 5) There exists today only a very limited capacity to model entities that must coordinate and communicate in a meaningful manner, such as is necessary for high performing teams.
- 6) Failures in rule-based models of human behavior are most apparent in virtual simulations for training applications where CGFs and semi-automated forces must interact in a realistic, meaningful manner with human-in-the-loop (HITL) participants.

12.6.2 Recommendations to the Scientific Community

- 1) Develop motivational frameworks that moderate behavior on the individual level and predict parameters at the aggregate level (teams/crowds/societies).
- 2) Try to relate constraints in human capabilities to realistic or inadequate behavioral responses as the strain levels expand. Pay attention to limits in the interaction between entities.
- 3) Take on the challenges associated with modeling communication and the dynamics of team coordination and adaptation.

12.6.3 Recommendations to NATO

- 1) Study the feasibility of inclusion of emotions and attitudes in constructive simulation, covering the full spectrum of interactions between friends and enemies, including aspects such as influencing public opinion, both domestically and in the operational theater.
- 2) Study behavior under threat to identify the degree to which SOPs guide rather than control behavior in trained personnel as well as how behaviors deviate from prescribed or rational behavior within the humans in general.
- 3) Set up methods to collect data on C2 performance in the field such as during NATO training or war game exercises.
- 4) Encourage the use and development of Human Behavior Representation within NATO M&S activities or exercises so as to raise awareness of the field in member and PfP nations.

CONCLUSIONS AND RECOMMENDATIONS



Chapter 13 – REFERENCES

- AAMC. (1999). *Human Engineering Program Process and Procedures* (No. MIL-HDBK-46855A). Redstone Arsenal, AI 35898-5270: US Army Aviation and Missile Command, ATTN: AMSAM-RD-SE-TD-ST.
- Adang, O. (1998). *Hooligans, autonomen, agenten. Geweld en politieoptreden in relsituaties*: Samson.
- Adler, N. J. (1990). *International dimensions of organizational behavior* (2nd ed.). Boston, MA: Kent.
- Allender, L., Kelley, T., Archer, S. and Adkins, R. (1997). IMPRINT. The transition and further development of a soldier-system analysis tool. *IMPRINT Quarterly HQDA(DAPE_MR) 300 Army Pentagon Washington DC 20310-0300*, 2-6.
- Allender, L., Lockett, J., Kelley, T. D., Salvi, L., Mitchell, D., Headley, D. B., et al. (1995). *Verification, Validation, and Accreditation of a Soldier-System Modeling Tool*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting.
- Allender, L., Salvi, L. and Promisel, D. (1997). Evaluation of Human Performance under Diverse Conditions via Modeling Technology et al. In *Proceedings of Workshop on Emerging Technologies in Human Engineering Testing and Evaluation, RTO-TR-021 AC/323(HFM-018)TP/19, Human Engineering Testing and Evaluation*. Brussels, Belgium: NATO Research Study Group 24.
- Allender, L., Wallace, J., Gluck, K., Smillie, R., Jackson, L., Tollefson, E., et al. (2006). *Human Behavior Representation Standards: Are We Ready for Them? Are They Ready for Us? (Panel Discussion)*. Paper presented at the Behavior Representation and Simulation (BRIMS).
- Allport, F. H. (1924). *Social Psychology*: Houghton Mifflin.
- Alnaes, E., MacMillan, A. J. F., Hart, S., Davies, J. M., Emmerson, P., Laughery, R., et al. (1998). *A Designers Guide to Human Performance Modeling* (No. AGARD-AR-356). Neuilly-sur-Seine Cedex, France: North Atlantic Treaty Organization (NATO) AGARD.
- Anastasi, A. (1988). Validity: Basic concepts. In A. Anastasi (Ed.), *Psychological Testing* (6 ed., pp. 139-164). New York, NY: Macmillan.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98, 409-429.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (2002). Spanning seven orders of magnitude: a challenge for cognitive modelling. *Cognitive Science*, 26(1), 85-112.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, D., Lebiere, C. and Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 1036-1060.
- Anderson, J. R. and Lebiere, C. (1998). The atomic components of thought.

REFERENCES

- Archer, S., Headley, D. and Allender, L. (2003). Manpower, personnel, and training integration methods and tools. In H. Booher (Ed.), *Handbook of Human Systems Integration*. Hoboken, NJ: John Wiley & Sons, Inc.
- ARL. (2005). *Improved Performance Research Integration Tool (IMPRINT): User's Guide Version 7*. Aberdeen, MD: US Army Research Laboratory, Human Research and Engineering Directorate.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U. and Waldon, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, 105, 442-481.
- Ashby, F. G. and Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin and Review*, 6, 363-378.
- Astrand, P.-O., Rodahl, K., Dahl, H. A. and Stromme, P. (2003). *Textbook of work physiology: Human Kinetics*.
- Aveni, A. (1977). The not-so-lonely Crowd: Friendship groups in collective behaviour. *Sociometry*, 40, 96-99.
- Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41, 203-226.
- Baranski, J. V. and Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology*, 24(3), 929-945.
- Baranski, J. V. and Petrusic, W. M. (2003). Adaptive decision processes in perceptual comparisons: Effects of changes in the global difficulty context. *Journal of Experimental Psychology*, 29(3), 658-674.
- Baxter, G. D. and Bass, E. J. (1998). *Human Error Revisited: Some Lessons for Situation Awareness*. Paper presented at the Proceedings of the Fourth Annual Symposium on Human Interaction with Complex Systems.
- Belyavin, A. (1999). *Modeling the effect of stress on human behavior*. Paper presented at the Proceedings of the 8th Conference on Computer Generated Forces and Behavior Representation, Orlando, Florida.
- Belyavin, A. J. and Farmer, E. W. (2006). *Modeling the workload and performance of psychomotor tasks*. Paper presented at the Behavioral Representation In Modeling and Simulation (BRIMS), Baltimore, MD.
- Belyavin, A. J. and Spencer, M. B. (2004). Modelling performance and alertness: The QinetiQ approach. *Aviation, Space, and Environmental Medicine*, 75(3), A93-A103.
- Berry, J. W. and Kalin, R. (1995). Multicultural and ethnic attitudes in Canada: An overview of the 1991 National Survey. *Canadian Journal of Behavioural Science*, 27 (3), 301-320.
- Bertie, A. J., Himmelweit, S. F. and Trigg, A. B. (Eds.). (2006). *Social norms, cognitive dissonance and broadcasting: how to influence economic agents* (Vol. 584). Secaucus, NJ: Springer.
- Boff, K. R., Kaufman, L. and Thomas, J. P. (1986a). *Handbook of Perception and Human Performance*: John Wiley and Sons, Inc.
- Boff, K. R., Kaufman, L. and Thomas, J. P. (1986b). *Handbook of Perception and Human Performance. Volume 2. Cognitive Processes and Performance.*: John Wiley and Sons, Inc.

- Boff, K. R. and Lincoln, J. E. (1988). *Engineering Data Compendium, Human Perception and Performance* (Vol. I, II, III). New York: John Wiley and Sons.
- Bolton, C. B. and al., e. (1973). *An Anthropometric Survey of 2000 Royal Air Force Aircrew 1970/71* (No. FPRC/1327).
- Bradley, C. M. and Robertson, K. A. (1998). *Combined Stressors: A review* (No. DERA/CHS/PP5/WP980083/1.0). Farnborough, UK: Defence Evaluation and Research Agency, Centre for Human Sciences.
- Bruin, R. d., Verwijs, C. and Vliet, A. J. v. (2007). *Mental Readiness of Teams – development of a team model as a module for SCOPE* (No. Report TNO-DV 2007 A140).
- Bruyn-Martin, L., Bandali, F. and Lamoureux, T. (2006). *Decision making styles: Classification system, contextual analysis and validation of classification system* (No. DRDC Toronto No. CR2006-063): Defence Research and Development Canada Toronto.
- Bunting, A. J. and Wilson, A. L. (1997). *Initial models of human performance for use in IPME* (No. DERA/CHS/PP5/TR97142): Defence Evaluation and Research Agency, Centre for Human Sciences.
- Byrne, M. D. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies*, 55, 41-84.
- Byrne, M. D. and Anderson, J. R. (1998). Perception and Action. In J. R. Anderson and C. Lebiere (Eds.), *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Campbell, G. E. and Bolton, A. E. (2005). HBR Validation: Integrating Lessons Learned From Multiple Academic Disciplines, Applied Communities, and the AMBR Project. In K. A. Gluck and R. W. Pew (Eds.), *Modeling Human Behavior with Integrated Cognitive Architectures: Comparison, Evaluation, and Validation* (pp. 365-395). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Card, S. K., Moran, T. P. and Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Carterette, E. C. and Friedman, M. P. (Eds.). (1978). *Handbook of perception*. New York: Academic Press.
- Carver, C. S., Scheier, M. F. and Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology*, 56, 267-283.
- Chen, H. and Sun, H. (2006). Body-based haptic interaction model for touch-enabled virtual environments. *Presence: Teleoperators and Virtual Environments*, 15(2), 186-203.
- Clark, D. O. (1972). *The MTM systems of the world*. Paper presented at the Annual Institute of Industrial Engineers Conference.
- Collins, M. and Quillian, M. R. (1969). Retrieval from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.
- Cooke, N. J. (2003). *Team Cognition in Distributed Mission Environments. Final technical rept.* Albuquerque: New Mexico University.

REFERENCES

- Cronbach, L. J. and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cross, K. D., Rugge, S. M. and Thorndyke, P. W. (1982). *Cognitive Process in Interpreting the Contour-Line Portrayal of Terrain Relief*. Arlington, Virginia: Office of Naval Research (ONR).
- Daan, S. and Beersma, D. (1984). Circadian gating of human sleep-wake cycles. In (pp. 129-155). New York: Raven Press.
- Dandekar, K., Raju, B. I. and Srinivasan, M. A. (2003). 3-D Finite Element Models of Human and Primate Fingertips to Investigate the Mechanics of Tactile Sense. *Journal of Biomechanical Engineering*, 125, 682-691.
- Darken, C. J. and Jones, B. E. (2007). *Computer Graphics-Based Target Detection for Synthetic Soldiers*. Paper presented at the Behavioral Representation in Modeling and Simulation (BRIMS), Norfolk, VA.
- Day, A. L. and Livingstone, H. A. (2001). Chronic and acute stressors among military personnel: do coping styles buffer their negative impact on health? *J Occup Health Psychol.*, 6(4), 348-360.
- Dekker, S. W. A. (2007). Doctors are More Dangerous than Gun Owners: A Rejoinder to Error Counting. *Human Factors*, 49(2), 177-184.
- Dennis, S. (2005). A memory-based theory of verbal cognition. *Cognitive Science*, 29, 145-193.
- Dennis, S. and Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, 108(2), 452-478.
- DMSO. (2006). VV&A Recommended practices guide. Retrieved 12Dec2007. from <http://vva.dmsomil/default.htm>.
- DND. (2002). *CF Operational Planning Process*. Retrieved. from.
- DOD. (2003). *Department of Defense Directive (DoDD) 5000.1 The Defense Acquisition System*. Retrieved 20Feb2008. from <https://akss.dau.mil/dapc/index.aspx>.
- DOD. (2004). *Defense Acquisition Guidebook, Chapter 6 Human Systems Integration*. Retrieved 20Feb2008. from <https://akss.dau.mil/dapc/index.aspx>; <https://akss.dau.mil/dag/>.
- Dompke, U. C. e. a. (2001). *Human Behaviour Representation* (No. RTO-TR-047 AC/323 (SAS-017) TP/25). Paris, FR: NATO Research and Technology Organization.
- Drumwright, E., Ng-Thow-Hing, V. and Matarić, M. (2006). *Toward a Vocabulary of Primitive Task Programs for Humanoid Robots*. Paper presented at the International Conference on Development and Learning.
- Duncan, R. K. and Grant, J. W. (1997). A finite-element model of inner ear hair bundle micromechanics. *Hearing Research*, 104, 15-26.
- Dyer, J. (1984). Team research and team training: A state-of-the-art review. In F. A. Muckler (Ed.), *Human factors review: 1984* (pp. 285-323). Santa Monica, CA: Human Factors Society.

- Endsley, M. R. (1993). *Situation awareness in dynamic human decision making: Theory*. Paper presented at the 1st International Conference on Situational Awareness in Complex Systems, Orlando, FL.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32-64.
- Essens, P., Vogelaar, A., Mylle, J., Blendell, C., Paris, C., Halpin, S. M., et al. (2005). *Military Command Team Effectiveness: Model and Instrument for Assessment and Improvement*. (No. NATO RTO Technical Report AC/323(HFM-087) TP/59.). Soesterberg, Netherlands: NATO RTA; TNO Human Factors.
- Fallesen, J. J. (1993). *Overview of Army Tactical Planning Performance Research* (No. Technical Report 984). Alexandria, VA.: U.S. Army Research Institute for Behavioral and Social Sciences.
- Farmer, E. W. (2000). Predicting cognitive workload and performance in future systems. *Journal of Defence Science*, 5(1), F4-6.
- Filippidis, L., Galea, E. R., Gwynne, S. and Lawrence, P. (2004). *Representing the Influence of Signage on Evacuation behaviour within building EXODUS* (No. Report 04/IM/119): Fire Safety Engineering Group.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381-391.
- Freed, M. A., Smith, R. G. and Sterling, P. (1992). Computational model of the on-alpha ganglion cell receptive field based on bipolar cell circuitry. *National Academy of Sciences of the United States of America*, 89(1), 236-240.
- Geen, R. G. (2001). *Human Aggression* (2nd ed.): Open University Press.
- Geller, A. and Moss, S. (2007). *Growing qawms: A case-based declarative model of Afghan power structures*. Paper presented at the Fourth Conference of the European Social Simulation Association (ESSA 2007).
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflon.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflon.
- Gigerenzer, G. and Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103(4), 650-669.
- Gillis, P. D. and Hursh, S. R. (1999). *Using behavior moderators to influence CGF command entity effectiveness and performance*. Paper presented at the Eighth Conference on Computer Generated Forces and Behavioral Representation, Orlando, Florida.
- Gluck, K. (2006). *Validation Standards for HBR. Symposium Human Behavior Representation Standards: Are We Ready for Them? Are They Ready for Us?* Paper presented at the 15th Behavior Representation in Modelling and Simulation (BRIMS) Conference, Baltimore, MD.
- Gluck, K., Gunzelmann, G., Harris, J., Scheutz, M. and Kershner, J. (2007). *Combinatorics Meets Processing Power: Large-Scale Computational Resources for BRIMS*. Paper presented at the Behavioral Representation In Modeling and Simulation (BRIMS), Norfolk, VA.

REFERENCES

- Gluck, K. A. and Pew, R. W. (Eds.). (2005). *Modeling Human Behavior with Integrated Cognitive Architectures: Comparison, Evaluation, and Validation*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Gluck, M. A., Shohamy, D. and Myers, C. (2002). How do people solve the “weather prediction” task? Individual variability in strategies for probabilistic category learning. *Learning and Memory*, 9, 408-418.
- Gonzalez-Avella, J. C., Cosenza, M. G., Klemm, K., Eguiluz, V. M. and San Miguel, M. (2007). Information feedback and mass media effects in cultural dynamics. *Journal of Artificial Societies and Social Simulation*, 10 (3), 9.
- Green, D. M. and Swets, J. A. (1967). *Signal detection theory and psychophysics*. New York: John Wiley and Sons, Inc.
- Greenberg, R. J., Velte, T. J., Humayun, M. S., Scarlatis, G. N. and De Juan, E., Jr. (1999). A computational model of electrical stimulation of the retinalganglion cell. *IEEE Transactions on Biomedical Engineering*, 46(5), 505-514.
- Gunzelmann, G., Gluck, K. A., Van Dongen, H. P. A., O'Connor, R. M. and Dinges, D. (2005). *A Neurobehaviorally Inspired ACT-R Model of Sleep Deprivation: Decreased Performance in Psychomotor Vigilance*. Paper presented at the Proceedings XXVII Annual Conference of Cognitive Science Society, Stresa, Italy.
- Halford, G. S., Ford, M., Busby, J. and Andrews, G. (2006). *Literature Review of Formal Models of Human Thinking* (No. DRDC Toronto Report Number CR 2006-056, PWGSC Contract Number: W7711-047932/001/TOR): Defence Research and Development Canada Toronto.
- Harris, R. M., Iavecchia, H. P. and Bittner, A. C., Jr. (1988). *The HOS Micromodels*. Paper presented at the Proceedings of the Human Factors Society 32nd Annual Meeting, Anaheim Hilton Hotel, Anaheim, California.
- Hawkins, H. C., Goss, S., Cain, B., Kerzner, L., Sheppard, C., Willis, R. P., et al. (2003). *Human performance modeling in military simulation: Current state of the art and the way ahead* (No. TR-TTCP/HUM/02/02): The Technical Cooperation Program (TTCP), Group HUM – Human Resources and Performance. Action Group 19. Representation of Human Behavior.
- Hayes, T. R. (1996). *Dismounted Infantry Movement Rate Study*. Natick, MA.: U.S. Army Soldier Systems Command, Natick Research, Development and Engineering Center.
- Heath, R. A. (1984). Random-walk and accumulator models of psychophysical discrimination: a critical evaluation. *Perception*, 13, 57-65.
- Hendy, K. C. and Farrell, P. S. E. (1997). *Implementing a model of human information processing in a task network simulation environment* (No. DCIEM No. 97-R-71). Toronto: Defence and Civil Institute of Environmental Medicine.
- Hendy, K. C., Liao, J. and Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors*, 39(1), 30-47.

- Hersey, P. and Blanchard, K. (1982). *Management of Organizational Behavior* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Hintzman, D. L. (1984). Minerva-2 – a simulation-model of human-memory. *Behavior Research Methods Instruments & Computers*, 16(2), 96-101.
- Hockey, G. R. J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue and circadian rhythms. In *Handbook of perception and Human Performance* (Vol. 2. Cognitive Processes and Performance). New York: John Wiley.
- Hofstadter, D. (1995). Fluid concepts and creative analogies: Computer models of fundamental mechanisms of thought.
- Hofstede, G. (1980). *Culture's consequences. International differences in work-related values*. Newbury Park: Sage Publications.
- Hofstede, G. H. (1991). *Cultures and organizations, software of the mind*. New York.
- Holland, T. (2008). *Towards a Lexicon for the Modeling and Simulation of Emergent Behavior Systems*. Paper presented at the 2008 Spring Simulation Interoperability Workshop.
- Houtmans, M. J. M. and Sanders, A. F. (1984). Perception of signals presented in the periphery of the visual field. *Acta Psychologica*, 55, 143-155.
- Hoyer, W. D. and MacInnis, D. J. (1997). *Consumer Behavior*. Boston: Houghton Mifflin.
- Hummel, J. E. and Holyoak, K. J. (2005). Relational reasoning in a neurally-plausible cognitive architecture: An overview of the LISA project. *Current Directions in Cognitive Science*, 14, 153-157.
- IEEE. (1998). *IEEE standard for application and management of the systems engineering process* (No. IEEE Std 1220-1998). 345 East 47th Street, New York, NY, 10017, USA: Institute of the Electrical and Electronics Engineers Inc.
- Ingraham, L. H. and Manning, F. J. (1981). Cohesion: Who needs it, what is it, and how do we get in to them? *Military Review*, 61(6), 2-12.
- Jackson, J. V. (1987). Idea for a mind. *ACM SIGART Bulletin*, 101, 23 – 26.
- Janis, I. L. and Mann, L. (1977). *Decision Making: A psychological analysis of conflict, choice and commitment*. New York: The Free Press.
- Jax, S. A., Rosenbaum, D. A., Vaughan, J. and Meulenbroek, R. G. J. (2003). Computational Motor Control and Human Factors: Modeling Movements in Real and Possible Environments. *Human Factors*, 45(1), 5-27.
- Jex, S. M., Bliese, P. D., Buzzell, S. and Primeau, J. (2001). The impact of self-efficacy on stressor-strain relations: coping style as an explanatory mechanism. *J Appl Psychol*.
- John, B. and Kieras, D. (1994). *The GOMS Family of Analysis Techniques: Tools for Design and Evaluation*. Pittsburgh PA.: Carnegie Mellon University, Human-Computer Interaction Institute.

REFERENCES

- John, B. E. and Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction*, 3, 320-351.
- Johnson, D. M. (2007). Helicopter simulator sickness: Age, experience, and amount learned. *International Journal of Applied Aviation Studies*, 7(2), 178-201.
- Jones, B., Toner, M. M., Daniels, W. L. and Knapik, J. J. (1983). *The Energy Cost and Heart Rate Response of Trained and Untrained Subjects Walking and Running in Shoes and Boots*. Natick MA: U.S. Army Research Institute of Environmental Medicine.
- Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-292.
- Kai, J., Bridgewater, R. and Spencer. (2001). "Just think of TB and Asians, that's all I ever hear": Medical learners' views about training to work in an ethnically diverse society. *Medical Education*, 35, 250-256.
- Kantowitz, B. H. and Sorkin, R. D. (1983). *Human Factors Understanding People-system Relationships*. New York: John Wiley & Sons.
- Käppler, W. D. (2006). *ARIADNE – Internet Based Error Risk Management and Prevention System*. Paper presented at the Proceedings for the European Annual Conference on Human Decision-Making and Manual Control, Valenciennes, France.
- Käppler, W. D. (2008). *Smart Driver Training Simulation. Save Money. Prevent*. Heidelberg, Berlin, New York: Springer.
- Käppler, W. D., Preßler, R. and Specht, D. (2008). Sicherheitsmanagement mit ARIADNE. In L. Schmidt, C. M. Schlick and J. G. Grosche (Eds.), *Ergonomie und Mensch-Maschine-Systeme* (pp. 437-462). Heidelberg, Berlin, New York: Springer.
- Karabaich, B. (2002). *Toward A Working Taxonomy of Groups*. Paper presented at the Computer Generated Forces.
- Kelley, T. D., Patton, D. J. and Allender, L. (2001). *Predicting Situation Awareness Errors using Cognitive Modeling*. Paper presented at the Proceedings of Human-Computer Interaction International 2001 Conference, Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality, New Orleans, Louisiana.
- Kendrick, D., Neuberg, S. and Cialdini, R. (2005). *Social Psychology: Unraveling the mystery* (3 ed.). Boston: Allyn & Bacon.
- Kieras, D. E. (1997). A Guide to GOMS model usability evaluation using NGOMSL. In M. Helander, T. Landauer & P. Prabhu (Eds.), *Handbook of human-computer interaction* (2 ed., pp. 733-766): Amsterdam: North-Holland.
- Kieras, D. E. (2004). EPIC Architecture Principles of Operation [Electronic Version]. Retrieved 30 Nov 2007 from [ftp://www.eecs.umich.edu/people/kieras/EPICtutorial/EPICPrinOp.pdf](http://www.eecs.umich.edu/people/kieras/EPICtutorial/EPICPrinOp.pdf).
- Kirwan, B. (1998). Human error identification techniques for risk assessment of high risk systems, part 1 : review and evaluation of techniques. *Applied Ergonomics*, 29(3), 157-177.

- Kornman, B. D. and Longtin, M. (2001). *Driving without Looking: Storing Behavioral Data Information in the Terrain Model*. Paper presented at the 10th Conference on Computer Generated Forces, Norfolk Waterside Marriott, Norfolk, VA.
- Krahe, B. (2001). *The social psychology of aggression*: Psychology Press.
- Krokos, K. J. and Baker, D. P. (2007). Classifying and Understanding Human Error. *Human Factors*, 49(2), 175-276.
- Kruschke, J. K. (1992). Alcove: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.
- Laird, J. E., Newell, A. and Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, 33, 1-64.
- Laird, J. E., Rosenbloom, P. S. and Newell, A. (1986). Chunking in soar: The anatomy of a general learning mechanism. *Machine Learning*, 1, 11-46.
- Lanchester, F. W. (1916). *Aircraft in warfare. The dawn of the fourth arm*. London: Constable and Company.
- Landauer, T. K. and Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Lazarus, R. S. and Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York: Springer.
- LeBon, G. (1895). *La Psychologie des Foules*: Alcan.
- Lee, M. D. and Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review*, 11(2), 343-352.
- Leiden, K., Laughery, K. R., Keller, J. W., French, J. W., Warwick, W. and Wood, S. D. (2001). *A Review of Human Performance Models for the Prediction of Human Error*. Boulder, CO: Micro Analysis and Design, Inc.
- Linster, C. and Cleland, T. A. (2004). Configurational and elemental odor mixture perception can arise from local inhibition. *Journal of Computational Neuroscience*, 16(1), 39-47.
- Love, B. C. (2005). In vivo or in vitro: Cognitive architectures and task-specific models. In K. A. Gluck & R. Pew (Eds.), *Modeling human behavior with integrated cognitive architectures: Comparison, evaluation and validation* (pp. 351-364). Mahwah, NJ: Lawrence Erlbaum Associates.
- Love, B. C., Medin, D. L. and Gureckis, T. M. (2004). Sustain: a network model of category learning. *Psychological Review*, 111, 309-332.
- Makowsky, M. (2007). Reducing Religious Extremism: A Model of Religious Populations and Policy. Retrieved from http://economics.gmu.edu/candidates_2007/ReducingReligiousExtremism-MakowskyJobPaper.pdf.
- Manning, F. J. and Fullerton, T. D. (1988). Health and well-being in highly cohesive units of the US Army. *Journal of Applied Social Psychology*, 18(6), 503-519.

REFERENCES

- Markman, A. B. and Gentner, D. (2001). Thinking. *Annual Review of Psychology*, 52, 223-247.
- Maynard, H. B., Stegemerten, G. J. and Schwab, J. L. (1948). *Methods-time measurement*. New York: McGraw-Hill.
- McArdle, W. D., Katch, F. I. and Katch, V. L. (2001). *Exercise physiology: Energy, Nutrition and Human Performance*: Lappingcott Williams & Wilkins.
- McClanahan, T., Feinberg, J., Goalwin, P. and Blemberg, P. (2001). *Human and Organizational Behavior Modeling (HOBM) Technology Assessment* (No. Report MS-00-0019/0028): MSIAC.
- McCormick, E. J. (1970). *Human Factors Engineering*. New York: McGraw-Hill.
- McCracken, J. H. and Aldrich, T. B. (1984). *Analyses of selected LHX mission functions: Implications for operator workload and system automation goals* (Technical Report No. MDA903-81-C-0504 ASI479-024-84). Fort Rucker, Alabama: US Army Research Institute Aircrew Performance and Training.
- McDaniel, M. A. and Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology*, 14, 127-144.
- McMahon, R., Spencer, M. and Thornton, A. (1995). *A Quick Response Approach to Assessing the Operational Performance of the XM983E1 NBCRS through the Use of Modeling and Validation Testing*. Paper presented at the International Test and Evaluation Association Proceedings, Huntsville, AL.
- McPhail, C. (1991). *The myth of the madding crowd*: Aldine de Gruyter.
- Miller, G. A. and Licklider, J. C. R. (1950). The Intelligibility of Interrupted Speech. *The Journal of the Acoustical Society of America*, 22(2), 167-173.
- Miller, N. and Dollard, J. (1941). *Social Learning and Imitation*: Yale University Press.
- Mitchell, D. K. (2000). *Mental workload and ARL workload modeling tools* (No. ARL-TN-161). Aberdeen Proving Ground, MD: US Army Research Laboratory, Human Research & Engineering Directorate.
- Montgomery, J. M. and Spencer, M. B. (1996). *A model to predict levels of alertness during irregular schedules of work and rest* (No. DRA/CHS/A&N/CR/96/007). Farnborough, UK: Defence Research Agency Centre for Human Sciences.
- Morrison, J. E. (2003). *A Review of Computer-Based Human Behavior Representations and Their Relation to Military Simulations* (No. IDA Paper P-3845). Alexandria, VA: Institute for Defence Analysis.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89(6), 609-626.
- NATO. (1998). *NATO Modelling and Simulation Master Plan* (No. Document AC/323 (SGMS)D/2, Version 1.0). Brussels: North Atlantic Treaty Organization, North Atlantic Council.
- NATO. (2004). *NATO Code of Best Practice for Command and Control Assessment* (No. RTO-TR-081 AC/323(SAS-026)TP/40). Neuilly-sur-Seine Cedex, France: Research and Technology Organisation (RTO), North Atlantic Treaty Organisation (NATO).

- Neal, D. M. (1994). A Further Examination of Anonymity, Contagion, and Deindividuation in Crowd and Collective Behaviour. *Sociological Focus*, 26, 93-107.
- Newell, A. (1990). *The unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Nguyen, Q. H., McKenzie, F. D. and Petty, M. D. (2005). *Crowd Behavior Architecture Model Cognitive Design*. Paper presented at the 2005 Conference on Behavior Representation in Modeling and Simulation (BRIMS), Universal City CA.
- Nilsson, N. (1998). *Artificial Intelligence A New Synthesis*: Elsevier.
- North, R. A. and Riley, V. A. (1989). W/INDEX: A predictive model of operator workload. In G. R. B. McMillan, D.; Salas, E.; Strub, M.H.; Sutton, R.; van Breda, L. (Ed.), *Applications of human performance models to system design* (Vol. 2, pp. 81-90). New York: Plenum Press.
- Nosofsky, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental psychology: General*, 115, 39-57.
- Oliver, L. W., Harman, J., Hoover, E., Hayes, S. M. and Pandhi, N. A. (1999). A quantitative integration of the military cohesion literature. *Military Psychology*, 11(1), 55-83.
- Oliver, P. and Myers, D. (2000). *Networks, Diffusion, and Cycles of Collective Action and its Figure*. Paper presented at the Social Movement Analysis: The Network Perspective, Ross Priory, Loch Lomond, Scotland.
- Olsson, P. A. and Regan, M. A. (1998). *A Comparison between actual and predicted evacuation times*. Paper presented at the 1st International Symposium on Human Behaviour in Fire, University of Ulster.
- Paillard, B., Mabillean, P., Morissette, S. and Soumagne, J. (1992). Perceval: Perceptual evaluation of the quality of audio signals. *Journal of the Audio Engineering Society*, 40, 21-31.
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition*, 3(5), 519-526.
- Pandolf, K. B., Givoni, B. and Goldman, R. F. (1977). Predicting energy expenditure with loads while standing or walking very slowly. *Journal of Applied Physiology*, 43(4), 577-581.
- Pattanaik, S. N., Fairchild, M. D., Ferwerda, J. A. and Greenberg, D. P. (1998). *Multiscale model of Adaptation, Spatial Vision and Color Appearance*. Paper presented at the Proceedings of IS&T/SID's 6th Color Conference, Arizona.
- Pew, R. W. and Mavor, A. S. (Eds.). (1998). *Modeling Human and Organizational Behavior. Applications to military simulations. Panel on Modeling Human Behavior and Command Decision Making: Representations for Military Simulations. Commission on Behavioral and Social Sciences and Education*. Washington, DC.: National Research Council. National Academy Press.
- Prince, K., Campbell, J., Picton, P. and Turner, S. (2004). A computational model of acute pain. *International Journal of Simulation*, 6(9).

REFERENCES

- Read, S., Miller, L., Rosoff, A., Eilbert, J., Lordanov, V., Le Mentec, J., et al. (2006). *Integrating Emotional Dynamics into the PAC Cognitive Architecture*. Paper presented at the Proceedings of the Fifteenth Conference on Behavior Representation In Modeling and Simulation, Baltimore, Maryland.
- Reason, J. (1990). *Human Error*. Cambridge, UK: Cambridge University Press.
- Ritter, F. E., Shadbolt, N. R., Elliman, D., Young, R. M., Gobet, F. and Baxter, G. D. (2002). *Techniques for modeling human performance in synthetic environments: a supplementary review*. Wright Patterson Air Force Base Ohio: Human Systems Information Analysis Center State-of-the-art report.
- Robinson, A. P., Duursma, R. A. and Marshall, J. D. (2005). A regression-based equivalence test for model validation: shifting the burden of proof. *Tree Physiology*, 25, 903-913.
- Roth, I. (Ed.). (1986). *Perception and representation : a cognitive approach*. Milton Keynes: Open University Press.
- Rotten, J. F., Barry, T., Milligan, M. and Fitzpatrick, M. (1979). The air pollution experience and physical aggression. *Journal of Applied Social Psychology*, 9, 387- 412.
- RSSB. (2004). *Managing large events and perturbations at stations: Passenger flow modelling technical review*. (No. RS021/R.03.): Railway Safety Standards Board.
- Russell, S. and Norvig, P. (2002). *Artificial Intelligence: A Modern Approach* (2 ed.): Prentice Hall.
- Rybak, I. A., Gusakova, V. I., Golovan, A. V., Podladchikova, L. N. and Shevtsova, N. A. (1998). A model of attention-guided visual perception and recognition. *Vision Research*, 38, 2387-2400.
- Sadeh, A., Keinan, G. and Daon, K. (2004). Effects of stress on sleep: the moderating role of coping style. *Health Psychology*, 23(5), 542-545.
- Sanders, A. F. and Houtmans, M. J. M. (1985). Perceptual processing modes in the functional visual field. *Acta Psychologica*, 55, 251-261.
- Schlick, C. (2006). *Industrial Engineering, Ergonomics, and Work Organization. Unit 6 Time Management II*. Retrieved 28 Nov 2007. from http://www.iaw.rwth-aachen.de/files/ieewo_06_ws2006.pdf
- Schutz, W. C. (1958). *FIRO: A Three Dimensional Theory of Interpersonal Behavior*. New York: Holt, Rinehart & Winston.
- Selfridge, O. G. (1959). *Pandemonium: A Paradigm for Learning*. Paper presented at the Proceedings of Symposium on the Mechanization of Thought processes.
- Senge, P. M., Roberts, C., Ross, R., Smith, B. and Kleiner, A. (1994). *The Fifth Discipline Fieldbook: Strategies and tools for building a learning organization*. New York: Doubleday Currency.
- Shiffrin, R. M. and Steyvers, M. (1997). Model for recognition memory: Rem – retrieving effectively from memory. *Psychonomic Bulletin & Review*, 4(2), 145-166.
- Siebold, G. L. (1999). The evolution of the measurement of cohesion. *Military Psychology*, 11(1), 5-26.

- Silverman, B. G., Cornwell, J. B. and O'Brien, K. (2003). Progress to date on the Human Performance Moderator Function Server (PMFserv) for rapidly generating reusable agents, forces and crowds [Electronic Version]. Retrieved 18Feb2007 from http://www.seas.upenn.edu/~barryg/PMFserv_2003_revisions.doc.
- Silverman, B. G., Johns, M., Cornwell, J. and O'Brien, K. (2006). Human Behavior Models for Agents in Simulators and Games: Part I – Enabling Science with PMFserv. *PRESENCE*, 15(2), 139-162.
- Silverman, B. G., Might, R., Dubois, R., Shin, H., Johns, M. and Weaver, R. (2001). *Toward A Human Behavior Models Anthology for Synthetic Agent Development*. Paper presented at the Computer Generated Forces.
- Smith, R. E. and Bayen, U. J. (2004). A multinomial model of prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 756-777.
- Stolwijk, J. A. J. and Hardy, J. D. (1977). Chapter 4. Control of body temperature. In D. H. K. Lee (Ed.), *Handbook of Physiology – Reactions to Environmental Agents*. Bethesda: American Physiological Society.
- Svensson, E., Angelborg-Thanderz, M. and Sjöberg, L. (1993). Mission challenge, mental workload and performance in military aviation. *Aviation Space and Environmental Medicine*, 64(11), 985-991.
- Swain, A. D. and Guttman, H. E. (1983). *Handbook of human reliability analysis with emphasis on nuclear power plant applications*. Retrieved from.
- Sznajd-Weron, K. and Sznajd, J. (2000). Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(6), 1157-1165.
- Taylor, G., Waltz, E. and White, J. (2008). *PSTK: A Toolkit for Modeling Dynamic Power Structures*. Paper presented at the 17th Conference on Behavior Representation in Modeling and Simulation (BRIMS 2008).
- Templeman, J. M., Sibert, L. E., Page, R. C. and Denbrook, P. S. (2006). *Immersive Simulation to Train Urban Infantry Combat*. Paper presented at the Virtual Media for Military Applications, Meeting Proceedings RTO-MP-HFM-136.
- Thagard, P. (2005). *Mind: introduction to cognitive science*. 266.
- Treurniet, W. C. (1996). Simulation of individual listeners with an auditory model. *Audio Engineering Society*.
- Triandis, H. C., Kurowski, L. L. and Gelfand, M. J. (1994). Workplace diversity. In H. C. Triandis & M. D. Dunnette (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 4, pp. 769-827). Palo Alto, Ca, USA: Consulting Psychologists Press, Inc.
- TTCP. (2000). *TTCP HUM TP2 Strategic Review*. Paper presented at the Report of the Annual Meeting, Human Performance and Resources Group (HUM) Technical Panel HUM-TP2, Training Technology, Annex D, pp33-39, Niagara Falls, CA.
- Vickers, D. (1970). Evidence for an accumulation model of psychophysical discrimination. *Ergonomics*, 13(1), 37-58.

REFERENCES

- Warwick, W., Archer, R., Hamilton, A., Matessa, M., Santamaria, A., Chong, R. S., et al. (2008). *Integrating Architectures: Dovetailing Task Network and Cognitive Models*. Paper presented at the Behaviour Representation In Modeling and Simulation (BRIMS 2008), Providence, Rhode Island.
- Warwick, W. and Hayes, P. (2003). *Developing Computational Models of Naturalistic Decision Making: Methodologies and Perspectives*. Paper presented at the Proceedings of CTA Conference 2003 Advanced Decision Architectures, College Park, MD.
- Warwick, W. and Hutchins, S. (2004). *Initial Comparisons between a "Naturalistic" Model of Decision Making and Human Performance Data*. Paper presented at the Behavioral Representation in Modeling and Simulation (BRIMS), Hyatt Regency Crystal City, Arlington, VA.
- Welford, A. T. (1968). *Fundamentals of Skill*. London, GB.: Methuen & Co. Ltd.
- Werner, J. and Webb, P. (1993). A six-cylinder model of human thermoregulation for general use on personal computers. *Ann Physiol Anthropol*, 12 (3), 123-134.
- Wherry, R. J. (1976). The human operator simulator-HOS. In T. B. Sheridan and G. Johannsen (Eds.), *Monitoring behavior and supervisory control*. New York: Plenum.
- Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 150-177.
- Wickens, C. D. (1984). *Engineering psychology and human performance*. Columbus, Ohio: Charles E Merrill.
- Wiegmann, D. A. and Shappell, S. A. (2003). *A human error approach to aviation accident analysis: The Human Factors Analysis and Classification System*. Burlington, VT: Ashgate Publishing.
- Winslow, R. L. (1989). Bifurcation analysis of nonlinear retinal horizontal cell models. Properties of isolated cells. *Journal of Neurophysiology*, 62(3), 738-749.
- Wissler, E. H. (1985). Mathematical simulation of human thermal behaviour using whole body model. In A. Shitzer & R. Eberhart (Eds.), *Heat and mass transfer in medicine and biology* (pp. 325-373). New York: Plenum Press.
- Zacharias, G., MacMillan J. and Van Hemel, S. (2008). *Behavioral Modeling and Simulation: From Individuals to Societies*. Washington, D.C.: Committee on Organizational Modeling from Individuals to Societies, National Research Council, The National Academies Press.
- Zachary, W., LeMentec, J.-C., Miller, L., Read, S. and Thomas-Meyers, G. (2005). *Steps toward a Personality-based Architecture for Cognition*. Paper presented at the Fourteenth Conference on Behavior Representation In Modeling and Simulation, Orlando: Institute for Simulation & Training.
- Zachary, W. W., Ryder, J. M. and Hicinbothom, J. H. (1998). Cognitive task analysis and modeling of decision making in complex environments. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making Decisions Under Stress: Implications for Individual and Team Training* (pp. 315-344). Washington, DC: American Psychological Association.
- Zsombok, C. E. and Klein, G. (Eds.). (1997). *Naturalistic Decision Making*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

REPORT DOCUMENTATION PAGE			
1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-TR-HFM-128 AC/323(HFM-128)TP/250	ISBN 978-92-837-0074-6	UNCLASSIFIED/ UNLIMITED
5. Originator Research and Technology Organisation North Atlantic Treaty Organisation BP 25, F-92201 Neuilly-sur-Seine Cedex, France			
6. Title Human Behavior Representation in Constructive Simulation			
7. Presented at/Sponsored by Final Report of Task Group 128, Consolidating the Findings of Task Group 143.			
8. Author(s)/Editor(s) Multiple			9. Date September 2009
10. Author's/Editor's Address Multiple			11. Pages 196
12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.			
13. Keywords/Descriptors			
Cognitive modeling	Human information processing	Reasoning	
Cognitive systems engineering	Methodology	Reviews	
Command and control	Models	Simulation	
Decision making	Operations research	Situational awareness	
Human behavior	Perception	Task analysis	
Human factors engineering			
14. Abstract			
<p>Lack of incorporation of human factors in military operational models is still seriously flawing the validity and predictive power. In this report methods are provided to implement the advanced human factors knowledge that is available, without creation of an unmanageable complexity. The core of the method is the use of human states, providing a basis for expressing the stresses experienced by personnel that are important to military outcomes (exhaustion, thermal load, mental workload, etc.). States integrate the effects of previous activity and recovery in a scientifically founded way, allowing models to run time based scenarios without restriction. Aggregated units like teams and larger units have additional HF properties that do not exist at the individual level.</p> <p>Effects based operations have a profound impact on the way an operational problem is solved and consequently also on the requirements placed on simulation models. Modeling EBOperations requires increased representation of cognition of coordinated units, capturing assessment, judgment, and decision making. The choices allow for variable behavior. This report attempts to provide some guidance on how human behavior models can be extended to capture these effects in military modeling and simulation. These considerations are synthesized in an overall scheme and a 19-step process to guide practitioners and analysts through an HF reinforced study case, called good practice.</p>			





BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int



DIFFUSION DES PUBLICATIONS RTO NON CLASSIFIEES

Les publications de l'AGARD et de la RTO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre et la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la RTO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (www.rto.nato.int) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE

Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7, D-53229 Bonn

BELGIQUE

Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research
for Defence, National RTO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Bruxelles

CANADA

DSIGRD2 – Bibliothécaire des ressources du savoir
R et D pour la défense Canada
Ministère de la Défense nationale
305, rue Rideau, 9^e étage
Ottawa, Ontario K1A 0K2

DANEMARK

Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5, 2750 Ballerup

ESPAGNE

SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

ETATS-UNIS

NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE (Correspondant)

Defence Industry & Research General
Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE

Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
P O Box 26
H-1525 Budapest

ITALIE

General Secretariat of Defence and
National Armaments Directorate
5th Department – Technological
Research
Via XX Settembre 123
00187 Roma

LUXEMBOURG

Voir Belgique

NORVEGE

Norwegian Defence Research
Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

PAYS-BAS

Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE

Centralny Ośrodek Naukowej
Informacji Wojskowej
Al. Jerozolimskie 97
00-909 Warszawa

PORTUGAL

Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE

LOM PRAHA s. p.
o. z. VTÚLaPVO
Mladoboleslavská 944
PO Box 18
197 21 Praha 9

ROUMANIE

Romanian National Distribution
Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353, Bucharest

ROYAUME-UNI

Dstl Knowledge and Information
Services
Building 247
Porton Down
Salisbury SP4 0JQ

SLOVAQUIE

Akadémia ozbrojených síl
M.R. Štefánika, Distribučné a
informačné stredisko RTO
Demanova 393, P.O.Box 45
031 19 Liptovský Mikuláš

SLOVENIE

Ministry of Defence
Central Registry for EU and
NATO
Vojkova 55
1000 Ljubljana

TURQUIE

Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi
Başkanlığı
06650 Bakanlıklar
Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)

7115 Standard Drive
Hanover, MD 21076-1320
ETATS-UNIS

The British Library Document Supply Centre

Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

Canada Institute for Scientific and Technical Information (CISTI)

National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants :

Scientific and Technical Aerospace Reports (STAR)

STAR peut être consulté en ligne au localisateur de ressources
uniformes (URL) suivant: <http://www.sti.nasa.gov/Pubs/star/Star.html>
STAR est édité par CASI dans le cadre du programme
NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
ETATS-UNIS

Government Reports Announcements & Index (GRA&I)

publié par le National Technical Information Service
Springfield
Virginia 2216
ETATS-UNIS
(accessible également en mode interactif dans la base de
données bibliographiques en ligne du NTIS, et sur CD-ROM)



BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int



DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

AGARD & RTO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO reports, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your Organisation) in their distribution.

RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of RTO reports as they are published, please visit our website (www.rto.nato.int) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM

Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research
for Defence, National RTO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30
1000 Brussels

CANADA

DRDKIM2 – Knowledge Resources Librarian
Defence R&D Canada
Department of National Defence
305 Rideau Street, 9th Floor
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC

LOM PRAHA s. p.
o. z. VTÚLaPVO
Mladoboleslavská 944
PO Box 18
197 21 Praha 9

DENMARK

Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5
2750 Ballerup

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GERMANY

Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7
D-53229 Bonn

GREECE (Point of Contact)

Defence Industry & Research General Directorate
Research Directorate, Fakinos Base Camp
S.T.G. 1020
Holargos, Athens

HUNGARY

Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
P O Box 26
H-1525 Budapest

ITALY

General Secretariat of Defence and
National Armaments Directorate
5th Department – Technological
Research
Via XX Settembre 123
00187 Roma

LUXEMBOURG

See Belgium

NETHERLANDS

Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY

Norwegian Defence Research
Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

POLAND

Centralny Ośrodek Naukowej
Informacji Wojskowej
Al. Jerozolimskie 97
00-909 Warszawa

PORTUGAL

Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

ROMANIA

Romanian National Distribution
Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6, 061353, Bucharest

SLOVAKIA

Akadémia ozbrojených síl
M.R. Štefánika, Distribučné a
informačné stredisko RTO
Demanova 393, P.O.Box 45
031 19 Liptovský Mikuláš

SLOVENIA

Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

SPAIN

SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

TURKEY

Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi
Başkanlığı
06650 Bakanlıklar – Ankara

UNITED KINGDOM

Dstl Knowledge and Information
Services
Building 247
Porton Down
Salisbury SP4 0JQ

UNITED STATES

NASA Center for AeroSpace
Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)

7115 Standard Drive
Hanover, MD 21076-1320
UNITED STATES

The British Library Document Supply Centre

Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and Technical Information (CISTI)

National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)

STAR is available on-line at the following uniform resource
locator: <http://www.sti.nasa.gov/Pubs/star/Star.html>
STAR is published by CASI for the NASA Scientific
and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
UNITED STATES

Government Reports Announcements & Index (GRA&I)

published by the National Technical Information Service
Springfield
Virginia 2216
UNITED STATES
(also available online in the NTIS Bibliographic Database
or on CD-ROM)