NORTH ATLANTIC TREATY ORGANIZATION

SCIENCE AND TECHNOLOGY ORGANIZATION

AC/323(HFM-297)TP/1141

www.sto.nato.int

STO TECHNICAL REPORT **TR-HFM-297**

Assessment of Augmentation Technologies for Improving Human Performance

(Évaluation des technologies d'augmentation visant à améliorer les performances humaines)

This report documents the findings of the Human Factors and Medicine Research Task Group 297.

Published December 2023

Distribution and Availability on Back Cover

NORTH ATLANTIC TREATY ORGANIZATION

SCIENCE AND TECHNOLOGY ORGANIZATION

AC/323(HFM-297)TP/1141

www.sto.nato.int

STO TECHNICAL REPORT **TR-HFM-297**

Assessment of Augmentation Technologies for Improving Human Performance

(Évaluation des technologies d'augmentation visant à améliorer les performances humaines)

This report documents the findings of the Human Factors and Medicine Research Task Group 297.

The NATO Science and Technology Organization

Science & Technology (S&T) in the NATO context is defined as the selective and rigorous generation and application of state-of-the-art, validated knowledge for defence and security purposes. S&T activities embrace scientific research, technology development, transition, application and field-testing, experimentation and a range of related scientific activities that include systems engineering, operational research and analysis, synthesis, integration and validation of knowledge derived through the scientific method.

In NATO, S&T is addressed using different business models, namely a collaborative business model where NATO provides a forum where NATO Nations and partner Nations elect to use their national resources to define, conduct and promote cooperative research and information exchange, and secondly an in-house delivery business model where S&T activities are conducted in a NATO dedicated executive body, having its own personnel, capabilities and infrastructure.

The mission of the NATO Science & Technology Organization (STO) is to help position the Nations' and NATO's S&T investments as a strategic enabler of the knowledge and technology advantage for the defence and security posture of NATO Nations and partner Nations, by conducting and promoting S&T activities that augment and leverage the capabilities and programmes of the Alliance, of the NATO Nations and the partner Nations, in support of NATO's objectives, and contributing to NATO's ability to enable and influence security and defence related capability development and threat mitigation in NATO Nations and partner Nations, in accordance with NATO policies.

The total spectrum of this collaborative effort is addressed by six Technical Panels who manage a wide range of scientific research activities, a Group specialising in modelling and simulation, plus a Committee dedicated to supporting the information management needs of the organization.

- 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 Panels and Group are the power-house of the collaborative model and are made up of national representatives as well as recognised world-class scientists, engineers and information specialists. In addition to providing critical technical oversight, they also provide a communication link to military users and other NATO bodies.

The scientific and technological work is carried out by Technical Teams, created under one or more of these eight bodies, for specific research activities which have a defined duration. These research activities can take a variety of forms, including Task Groups, Workshops, Symposia, Specialists' Meetings, Lecture Series and Technical Courses.

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

Published December 2023

ISBN 978-92-837-2455-1

Table of Contents

	Page
List of Figures	**vii**
List of Tables	**ix**
List of Acronyms	**x**
HFM-297 Membership List	**xii**
Executive Summary and Synthèse	**ES-1**
Chapter 1 – Research Task Group (HFM-297) Work Plan	**1-1**
1.1 Justification for RTG HFM-297	1-1
1.2 RTG HFM-297 Objectives and Scoping	1-2
1.3 Evolution of the Assessment Framework	1-3
1.4 Augmentation Attributes	1-4
1.4.1 Enhancement	1-4
1.4.1.1 Augmentation Tools and Interfaces	1-6
1.4.1.2 Building Efficiencies	1-10
1.4.1.3 Automation	1-13
1.5 Summary	1-14
1.6 References	1-14
Chapter 2 – Evaluating Technology Against a Human Performance Temporal Construct	**2-1**
2.1 Performance Domains Left and Right of a “Bang” Event	2-1
2.1.1 What Is the BANG Event?	2-2
2.1.2 Left of the BANG Event	2-2
2.1.2.1 Institutional Training and Education	2-3
2.1.2.2 Operational Readiness and Sustainment	2-3
2.1.2.3 Mission Preparation and Rehearsal	2-3
2.1.2.4 After Action Review	2-3
2.1.3 Right of the BANG Event	2-4
2.1.3.1 Real-Time Support	2-4
2.1.3.2 Remote Control	2-4
2.1.3.3 Monitoring and Optimization	2-4
2.2 Assessment Framework Definitions	2-5
2.2.1 The SWOT (Strengths, Weaknesses, Opportunities, Threats) Analysis	2-5
2.3 Topics to be Covered	2-6
2.3.1 Augmentation Technology for Adaptive Instruction for Accelerated Readiness (Chapter 3)	2-6

2.3.2 Augmentation Technology for Mission Prep and Rehearsal (Chapter 4) 2-6
2.3.3 Augmentation Technology for Real-time Support and Remote Control (Chapter 5) 2-7
2.3.4 Augmentation Technology for Performance Monitoring and Optimization (Chapter 6) 2-7
2.3.5 Augmentation Technology for After Action Review (Chapter 7) 2-8
2.4 References 2-8

Chapter 3 – Adaptive Instruction and Accelerated Readiness 3-1
3.1 Introduction 3-1
3.2 Requirements for Implementing AIS 3-2
3.2.1 Human Performer 3-2
3.2.2 Technology and Hardware Requirements 3-2
3.2.3 Data Requirements 3-3
3.3 Strengths, Weaknesses, Opportunities, Threats (SWOT) Analysis 3-3
3.3.1 Strengths 3-4
3.3.1.1 Optimized Training Time and Learning Outcomes 3-4
3.3.1.2 Established Implementation Techniques 3-4
3.3.1.3 Domain Agnostic 3-5
3.3.2 Opportunities 3-5
3.3.2.1 Nascent Applications to Instructor-in-the-Loop Training and Collaborative Learning 3-5
3.3.2.2 Learning Ecosystems Investments 3-5
3.3.2.3 Distributed Implementation Potential 3-6
3.3.2.4 Adaptive Performance Aiding 3-6
3.3.2.5 Enhanced Connectivity 3-6
3.3.2.6 Continued Advancement in Synthetic Environments 3-6
3.3.3 Weaknesses 3-6
3.3.3.1 Low-to-Moderate Maturity 3-6
3.3.3.2 Ecosystem Investment Requirements 3-7
3.3.3.3 Labor Intensive Development 3-7
3.3.3.4 Maturity Level of Natural Language Processing 3-7
3.3.4 Threats 3-7
3.3.4.1 Data/Cyber Security 3-7
3.3.4.2 Relatively Fast Pace of Technology Evolution 3-7
3.3.4.3 High Initial Investment 3-7
3.3.4.4 Skill Erosion 3-8
3.3.4.5 User Adoption 3-8
3.4 Conclusion 3-8
3.5 References 3-8

Chapter 4 – Mission Preparation and Rehearsal 4-1
4.1 Introduction 4-1
4.2 Human Performer Within the Mission Prep and Rehearsal Domain 4-2

4.3 Augmentation Technologies for Mission Planning 4-2
4.3.1 Virtual Presence 4-2
4.3.2 Immersive Technologies 4-3
4.3.3 Artificial Intelligence 4-3
4.4 Strengths, Weaknesses, Opportunities, Threats (SWOT) Analysis 4-4
4.5 References 4-5

Chapter 5 – Real-Time Support and Remote Control 5-1
5.1 Introduction 5-1
5.1.1 Scope 5-1
5.2 Digital Backbone of Data Network Infrastructure 5-1
5.3 Real-Time Support 5-2
5.4 Remote Control 5-4
5.4.1 Human Performance Constructs 5-4
5.5 Strengths, Weaknesses, Opportunities and Threats (SWOT) Analysis 5-5
5.6 References 5-6

Chapter 6 – Cognitive Monitoring and Optimization 6-1
6.1 Introduction 6-1
6.2 Monitoring 6-2
6.2.1 Subjective Sleepiness and Fatigue 6-2
6.2.2 Objective Performance 6-2
6.2.3 Psychophysiological Indicators of Sleepiness and Fatigue 6-3
6.2.4 Monitoring Summary and Conclusions 6-4
6.3 Mathematical Performance Prediction Modeling 6-4
6.3.1 Countermeasures 6-5
6.4 Conclusion and SWOT Analysis 6-6
6.4.1 Strengths, Weaknesses, Opportunities, Threats (SWOT) Analysis 6-7
6.5 References 6-8

Chapter 7 – After Action Review (AAR) 7-1
7.1 Introduction 7-1
7.2 User Requirements and Limitations 7-2
7.2.1 User Roles and Skills 7-2
7.2.2 User Limitations and Challenges 7-3
7.2.2.1 Limitations at the Level of AAR Preparation: Performance Observation, Metrics and Analysis 7-4
7.2.2.2 Challenges in AAR Conduct: Reviewing Events and Presenting Feedback to Trainees 7-4
7.2.2.3 Challenges in AAR Conduct: Group Discussion Facilitation 7-5
7.3 AAR Augmentation Technologies Context 7-6
7.3.1 Augmentation Supports for AAR Preparation 7-6
7.3.2 Augmentation Supports to AAR Conduct: Event Review and Visualization 7-7

7.3.3 Augmentation Supports to AAR Conduct: Facilitation of Group Discussion 7-8
7.4 Strengths, Weaknesses, Opportunities, Threats (SWOT) Analysis 7-9
7.4.1 Strengths 7-11
7.4.2 Weaknesses 7-11
7.4.3 Opportunities 7-11
7.4.4 Threats 7-11
7.5 Recommended Near-Term Applications 7-12
7.6 Science and Technology Investments 7-12
7.6.1 Current Research 7-12
7.6.2 Recommended Research Investments 7-13
7.7 References 7-14

Chapter 8 – Insights, Lessons Learned, and Conclusion 8-1
8.1 Summary of Augmentation Domain Chapters 8-1
8.1.1 General Overview 8-1
8.1.2 Synthesis of SWOT Analyses Across Domains 8-4
8.2 Recommendations and Considerations on the Application of Augmentation Technologies 8-7
8.2.1 Best Practices in Applying Augmentation Technologies 8-7
8.2.2 Caveats and Areas that Require Further Research 8-7
8.3 Longer Term R&D Investment Recommendations 8-8
8.4 The Evolving Augmentation Tech Space and the Challenges of Reporting Strategies for Future RTGs (Mostly Copied from Early Report Outline) 8-10
8.5 Concluding Thoughts 8-11
8.6 References 8-12

Annex A – Instantiation Card (iCard) Template and Examples A-1
A.1 References A-4

Annex B – HFM-297 Engagements and Extramural Activities B-1
B.1 End-User Research and Development Workshop; West Point, NY USA B-2
B.2 International Ergonomics Association (IEA) World Congress HFM-297 Paper Session; Florence, ITALY B-4
B.3 HFM-297 Panel at Canadian Armed Force's "Ready, Set, Innovate" Symposium; Canadian forces BASE Borden, Canada B-5
B.4 Interactive Matrix Workshop; Stockholm, Sweden B-5
B.5 Interservice / Industry Training, Simulation and Education Conference (I/ITSEC) HFM-297 Special Demonstration Event; Orlando, FL USA B-6
B.6 COVID-19 Webinars B-6
B.7 References B-8

List of Figures

Figure		Page
Figure 1-1	Augmentation Technology and Human Performance MindMap	1-5
Figure 1-2	Augmentation Model in Support of Sensory and Cognitive Enhancement	1-6
Figure 1-3	Promotional Picture of Commercial HMDs	1-8
Figure 1-4	Promotional Picture of Commercial Binocular and Monocular AR HMD Systems	1-8
Figure 1-5	Categories of AR Devices	1-8
Figure 1-6	(a) Example of a Spherical Dome Projection System and (b) Example of a CAVE Setup	1-9
Figure 1-7	Examples of Haptic Feedback Gloves	1-10
Figure 1-8	Diagram of the OODA Loop by Patrick Edwin Moran Licensed Under CC BY 3.0	1-11
Figure 2-1	The Domains of Human Performance Across an Operational ‘Bang Event’ Timeline	2-2
Figure 2-2	Internal and External Organizational Factors Aligned to a SWOT Framework	2-5
Figure 5-1	Edge Compute Real-Time AI and Video Analytics Supporting Search and Rescue, Object Localization, Image Enhancement, Behavioral Analytics, Predictions, etc. to Low Latency Interactive Human Machine Applications	5-2
Figure 5-2	Example of Real-Time Applications Used in Strategic, Operational and Tactical Functions	5-3
Figure 7-1	Typical Canadian Army AAR Phases	7-2
Figure 7-2	Sampling of AAR Technologies in Live, Virtual and Constructive Environments	7-6
Figure 8-1	Gartner Hype Cycle for Emerging Technologies, 2022	8-3
Figure A-1	(Top) Technology iCard Template – Used to Establish an Aggregated Representation of a Technology and the Underlying Evidence Showing its Effectiveness Across all Applications; (Bottom) Application iCard Template – Used to Represent a Focused Perspective on a Technology Applied Within the Boundaries of Defined Tasks, Conditions and Standards	A-2
Figure A-2	iCard Example on Low Current Brain Stimulation Developed During the HFM-297 Meeting at DRDC Toronto	A-3
Figure A-3	iCard Example on Cognitive Training Developed During the HFM-297 Meeting at DRDC Toronto	A-3

Figure A-4	iCard Example of Adaptive Instructional Systems Developed During the HFM 297 Meeting at DRDC Toronto	A-4
Figure B-1	Virtual Pilot Seen through Augmented Reality Headset	B-3
Figure B-2	Pilot Training Virtual Reality Configuration with Low-Cost Interfacing	B-3
Figure B-3	Emergency Operations Center Virtual Command Map	B-4
Figure B-4	Draft Interactive Matrix for Organizing Insights and Recommendations Across the Augmentation Domain of Interest	B-6
Figure B-5	I/ITSEC Focus Event Description and Demonstration Participants as Highlighted in the Conference Program	B-7

List of Tables

Table		Page
Table 3-1	Adaptive Instruction and Accelerated Readiness SWOT Analysis	3-2
Table 4-1	Mission Preparation and Rehearsal SWOT Analysis	4-4
Table 5-1	Real-Time Support and Remote Control SWOT Analysis	5-5
Table 6-1	Cognitive Monitoring and Optimization SWOT Analysis	6-7
Table 7-1	AAR Phases as per Villado and Arthur (2013)	7-1
Table 7-2	Summary of User Challenges and Augmentation Technology Supports for Different AAR Phase	7-8
Table 7-3	SWOT Analysis for AAR Augmentation Domain	7-9
Table 8-1	Commonalities Across Augmentation Domains SWOT Analysis	8-5
Table B-1	HFM-297 Workshop and Demonstration Events	B-1

List of Acronyms

2D	2-Dimensional
3D	3-Dimensional
3GPP	3rd Generation Partnership Project
5G	5th Generation (networks)
AAR	After Action Review
ACT-R	Adaptive Control of Thought – Rational
ADL	Advanced Distributed Learning
AETC	Air Education and Training Command
AI	Artificial Intelligence
AIS	Adaptive Instructional Systems
API	Application Programming Interfaces
AR	Augmented Reality
AREA	Augmented Reality Enterprise Alliance
ARLEM	AR Learning Experience Model
ATF	Anatomical Transfer Function
BwCSC	Bundeswehr Command and Staff College
CAVE	Cave Automatic Virtual Environment
CFB	Canadian Forces Base
CIPO	Context-Input-Process-Output-Outcome
COVID-19	Coronavirus Disease 2019
CTAT	Cognitive Tutor Authoring Tools
CWES	Canadian Weapons Effect Simulation
DM	Data Mining
DRDC	Defence Research and Development Canada
DSTL	Defence Science and Technology Laboratory
E2DT	Emerged and Emerging Disruptive Technologies
ET	Exploratory Group
FLXD	Flow Driven Experiential Learning
FRMS	Fatigue Risk Management System
GIFT	Generalized Intelligent Framework for Tutoring
GPS	Global Positioning System
HFM	Human Factors and Medicine
HMD	Head-Mounted Display
HRTF	Head-Related Transfer Function
IEA	International Ergonomics Association
IEEE	Institute of Electrical and Electronics Engineers
IOS	Instructor Operating Stations
ISD	Instructional Systems Designer
ITS	Intelligent Tutoring System
I/ITSEC	Interservice/Industry Training Simulation and Education Conference

JCATS	Joint Conflict and Tactical Simulation
LMS	Learning Management System
LTE	Long-Term Evolution
LVC	Live-Virtual-Constructive
MILPERSGEN	Military Personnel Generation
ML	Machine Learning
MPPM	Mathematical Performance Prediction Models
MR	Mixed Reality
NATO	North Atlantic Treaty Organization
NIAG	NATO Industry Advisory Group
NLP	Natural Language Processing
NMSG	NATO Modeling and Simulation Group
OODA	Observe, Orient, Decide and Act
PC	Personal Computer
R&D	Research and Development
RTG	Research Task Group
SA	Situation Awareness
SAGAT	SA Global Assessment Technique
SE	Synthetic Environment
SEW-AT	Submarine Electronic Warfare Adaptive Trainer
SME	Subject Matter Expert
SPAM	Situation Present Assessment Technique
STE	Synthetic Training Environment
STO	Science & Technology Office
STTC	Simulation and Training Technology Center
SWOT	Strengths Weaknesses Opportunity Threats
SWAT	Special Weapons and Tactics
TES	Transcranial Electrical Stimulation
TLA	Total Learning Architecture
UCATT	Urban Combat Advanced Training Technology
UMP	Unified Model of Performance
USMA	United States Military Academy
VBS3	Virtual BattleSpace 2
VR	Virtual Reality
VUCA	Volatile, Uncertain, Complex, and Adversarial
xAPI	eXperience Application Programming Interface
XR	eXtended Reality

HFM-297 Membership List

CO-CHAIRS

Dr. Benjamin GOLDBERG*
Human Research & Engineering Directorate
UNITED STATES
Email: benjamin.s.goldberg.civ@mail.mil

Dr. Jerzy JARMASZ*
Defence Research and Development Canada
CANADA
Email: jerzy.jarmasz@drdc-rddc.gc.ca

MEMBERS

Dr. Thomas ALEXANDER[1]
Federal Institute for Occupational Safety and Health
GERMANY
Email: alexander.thomas@baua.bund.de

Dr. Elizabeth BIDDLE*
The Boeing Company
UNITED STATES
Email: elizabeth.M.Biddle@boeing.com

LTC Vincent CAPALDI*
Uniformed Services University of the Health Sciences
UNITED STATES
Email: vincent.f.capaldi.mil@mail.mil

Dr. Emilie DESSY
Royal Military Academy
BELGIUM
Email: emi.dessy@gmail.com

Dr. John Dexter FLETCHER*
OSD
United States
Email: fletcher@ida.org

Dr. Glenn GUNZELMANN*
Air Force Research Laboratory
UNITED STATES
Email: glenn.gunzelmann@us.af.mil

CAPT(N) Thomas HERZIG*
DHA
UNITED STATES
Email: thomas.c.herzig.mil@mail.mil

Mr. Paolo PROIETTI
Leonardo SpA
ITALY
Email: paolo.proietti@leonardo.com

Dr. Patrick SANDOR
IRBA / SSA
FRANCE
Email: patrick.sandor@def.gouv.fr

Prof. Lucio TOMMASO DE PAOLIS
University of Salento
ITALY
Email: lucio.depaolis@unisalento.it

* **Contributing Author**

1 **Former Co-Chair**

ADDITIONAL CONTRIBUTORS

Dr. Thomas BALKIN*
Walter Reed Army Institute of Research
UNITED STATES
Email: thomas.j.balkin.ctr@health.mil

Mr. Ian GREIG[2]
MoD-DSTL
UNITED KINGDOM
Email: igreig@mail.dstl.gov.uk

Dr. Mayowa OLONILUA*
Dstl
UNITED KINGDOWN
Email: moolonilua@mail.dstl.gov.uk

Dr. Peder SJÖLUND*
Swedome
SWEDEN
Email: peder.sjolund@swedome.se

Dr. Robert SOTTILARE[3]
SoarTech, Inc.
UNITED STATES
Email: bob.sottilare@soartech.com

PANEL/GROUP MENTORS

Ms. Christina ALDRIN
FMV (Defence Materiel Administration)
SWEDEN
Email: christina.aldrin@fmv.se

CAPT(N) Paul CHATELIER (Retired)
US Naval Postgraduate School
UNITED STATES
Email: Paul.R.Chatelier.civ@mail.mil

* **Contributing Author**
2 **Former Member**
3 **Former Co-Chair**

Assessment of Augmentation Technologies for Improving Human Performance

(STO-TR-HFM-297)

Executive Summary

NATO Human Factors and Medicine (HFM) Research Task Group (RTG) HFM-297 was established to support the "Assessment of augmentation technologies for improving human performance". The members of the RTG met from 2017 to 2022. During this time, the team met at regular intervals (in-person and virtually) to define the scope of the augmentation technologies it would consider. This involved developing a framework to guide the controlled assessment of augmentation tools and methods against a set of defined metrics aligned to human performance and task outcomes. A goal was to analyze the relative merits of human performance as supported by augmentation and to identify cross-domain themes used to establish suggested best practices. This involves identifying recommendations for the continued application of, and research on augmentation technologies to support human performance in military contexts. These activities were supplemented by frequent engagements with military domain experts and requirements holders, and research and industry subject matter experts.

To manage the scope of augmentation technologies considered in the analyses, the RTG decided to focus primarily on technologies that "augment" the task environment and task process (e.g., synthetic environments, interface technologies) rather than the operator directly (e.g., exoskeletons, neural implants). Despite this distinction, maintaining a strict dissection between "environment-" and "operator-" aligned augmentation proved challenging. Nevertheless, with a scope defined, RTG 297 proceeded to develop a framework to analyze these technologies and integrate evidence from the research literature with operational requirements. The framework adopted for this investigation involved the application of the Strength-Weaknesses-Opportunities-Threats (SWOT) analysis methodology. The following five performance domains are broken down through a formalized SWOT analysis within the report, organized along an operational timeline from force generation to operations to post-operations lessons learned:

- Adaptive Instruction and Accelerated Readiness.
- Mission Preparation and Rehearsal.
- Real-Time Support and Remote Control.
- Cognitive Monitoring and Optimization.
- After Action Review.

The SWOT analyses applied to the performance domains yielded a number of themes common to all task domains and human performance requirements. These include:

- The centrality of human performance front-end considerations and human factors principles to the successful application of augmentation technologies;
- The inherent dependence on data and information technology infrastructure in the effective implementation of these technologies, and the need to sustain investment and efforts in developing data standards and overall data strategies to ensure interoperability and extensibility;
- Security, reliability, privacy and ethical considerations will play a determining role in augmentation technologies as they become increasingly adopted by military organizations;

- The very dynamic (rapidly evolving) nature of the technologies themselves and the mission sets to which they could be applied create significant challenges for systematic assessments of their effectiveness and value, in particular for the traditional, report-based format typical of RTG studies;
- Nonetheless, the evidence reviewed by the RTG makes it clear that a number of augmentation technologies already have established track records in training (e.g., adaptive instructional systems, visual synthetic environments) and operational (e.g., augmented reality) settings; and
- A number of evolving technologies (e.g., machine learning, performance monitoring, haptic interfaces for virtual reality) hold significant promise for near- and far-term applications in support of military human performance and training, but further research is required before operationally valid.

Reflecting on the RTG's challenges with defining a scope for the study, an analysis framework, and engaging with subject matter experts to ensure the study's relevance, the members of this group recommend that NATO STO consider more dynamic and responsive processes and formats (e.g., web-based reporting outputs using community-sourced information) for conducting studies on rapidly evolving technical domains such as augmentation technologies for human performance and training.

Évaluation des technologies d’augmentation visant à améliorer les performances humaines

(STO-TR-HFM-297)

Synthèse

Le groupe de recherche (RTG) HFM-297 en facteurs humains et médecine (HFM) de l’OTAN a été créé afin de soutenir « l’évaluation des technologies d’augmentation visant à améliorer les performances humaines ». Les membres du RTG se sont réunis de 2017 à 2022. Pendant cette période, l’équipe s’est retrouvée à intervalles réguliers (en personne et virtuellement) pour définir le champ des technologies d’augmentation qu’elle étudierait. Cela impliquait d’établir un cadre d’évaluation contrôlée des outils et méthodes d’augmentation, au moyen d’un jeu d’indicateurs définis mesurant les performances humaines et les résultats des tâches. Le but était d’analyser les mérites relatifs des performances humaines soutenues par l’augmentation et d’identifier les thèmes interdomaines utilisés pour établir des suggestions de bonnes pratiques. Il s’agissait d’émettre des recommandations permettant de poursuivre l’utilisation des technologies d’augmentation et la recherche à ce propos, afin de soutenir les performances humaines dans les contextes militaires. Ces activités ont été complétées par de fréquents échanges avec des experts et des porteurs d’exigences dans le domaine militaire, ainsi qu’avec des spécialistes de la recherche et de l’industrie.

Pour gérer le champ des technologies d’augmentation prises en compte dans les analyses, le RTG a décidé de se concentrer principalement sur les technologies qui « augmentent » l’environnement de la tâche et le processus de la tâche (par exemple, les environnements synthétiques, les technologies d’interface) plutôt que l’opérateur (par exemple, les exosquelettes, les implants neuronaux). Malgré cette distinction, le maintien d’une stricte séparation entre l’augmentation portant sur l’ « environnement » et celle portant sur « l’opérateur » s’est avéré délicat. Néanmoins, une fois le champ défini, le RTG-297 a élaboré un cadre pour analyser ces technologies et intégrer des preuves issues de la littérature de recherche dans les exigences opérationnelles. Le cadre retenu pour l’étude impliquait l’utilisation de la méthodologie SWOT (analyse des forces, faiblesses, opportunités et menaces). Les cinq domaines de performance suivants sont décomposés à l’aide d’une analyse SWOT formalisée dans le rapport, organisée selon une chronologie opérationnelle qui va de la production de la force aux opérations, en passant par les leçons retenues à l’issue des opérations :

- Enseignement adaptatif et accélération de la préparation ;
- Préparation à la mission et répétition de la mission ;
- Assistance en temps réel et commande à distance ;
- Surveillance et optimisation cognitives ;
- Compte rendu après action.

Les analyses SWOT appliquées aux domaines de performance ont produit un certain nombre de thèmes communs à tous les domaines de tâches et aux exigences de performances humaines. Ces thèmes sont les suivants :

- Les considérations préalables aux performances humaines et les principes des facteurs humains sont au centre de l’application réussie des technologies d’augmentation ;

- La mise en œuvre efficace de ces technologies dépend intrinsèquement de l'infrastructure des données et des technologies de l'information et les investissements et les efforts en matière d'élaboration de normes de données et de stratégies globales de données doivent être maintenus pour garantir l'interopérabilité et l'extensibilité ;
- La sûreté, la fiabilité, la confidentialité et les considérations éthiques joueront un rôle déterminant dans les technologies d'augmentation à mesure de leur adoption par les organisations militaires ;
- La nature très dynamique (évoluant rapidement) des technologies elles-mêmes et les ensembles de missions auxquels elles pourraient être appliquées sont des obstacles importants à l'évaluation systématique de leur efficacité et de leur valeur, en particulier dans le format traditionnel, à base de rapports, typique des études du RTG ;
- Toutefois, les preuves examinées par le RTG montrent clairement qu'un certain nombre de technologies d'augmentation ont déjà des antécédents établis en matière de formation (par exemple, systèmes d'enseignement adaptatif, environnements visuels synthétiques) et de contextes opérationnels (par exemple, réalité augmentée) ; et
- Un certain nombre de technologies évolutives (par exemple, l'apprentissage automatique, le suivi des performances, les interfaces haptiques de réalité virtuelle) sont très prometteuses à court et long terme pour des applications soutenant la formation et les performances humaines militaires, mais leur validation opérationnelle nécessite d'autres recherches.

Après avoir réfléchi aux défis du RTG pendant la définition du champ de l'étude et du cadre d'analyse et après avoir échangé avec des experts pour garantir la pertinence de l'étude, les membres de ce groupe recommandent à la STO de l'OTAN d'envisager des processus et formats plus dynamiques et plus réactifs (par exemple, des résultats de rapport accessibles par un navigateur Internet et utilisant des informations issues de la communauté) pour mener des études dans des domaines techniques évoluant rapidement, tels que les technologies d'augmentation pour la formation et les performances humaines.

Chapter 1 – RESEARCH TASK GROUP (HFM-297) WORK PLAN

Benjamin Goldberg, Jerzy Jarmasz, and Peder Sjölund

This chapter summarizes the background and justification for Research Task Group (RTG) HFM-297, Assessment of Augmentation Technologies for Improving Human Performance, for the period September 2017 – June 2022. It describes the objectives and provides an ontological representation of the human performance spectrum and the technology sector aiming to influence the performance factors aligned to those tools, methods and best practices. Furthermore, the scope of the reporting activity is defined and definitions for the attributes aligned to augmentation, performance, and the human performer are provided. This includes a summary of the background and measurement constructs of human performance as it relates to HFM-297 and a general definition and review of augmentation technologies and their alignment to the task domains they support. It is important to note that this is not an all-encompassing review of the available technologies for use today, but rather an investigation on the impact of augmentation tools and methods across performance constructs that matter from an operational context.

Following this introductory chapter, the subsequent chapter describes the RTGs methodology for organizing human performance across a spectrum of operations and presents the Strengths Weaknesses Opportunities Threats (SWOT) framework used to analyze technology constraints. This establishes the guidelines for the remaining chapters that present SWOT analyses across the augmentation domains of adaptive instruction/accelerated readiness, mission prep and rehearsal, performance monitoring and optimization, real-time support / remote control, and After Action Reviews. In the final chapter, a review of overall recommendations and discussion of the challenge of reporting against a volatile technology sector with continual advancement and investment in the rapid maturation of these capabilities is presented.

1.1 JUSTIFICATION FOR RTG HFM-297

When initiating this research activity, the outcomes of two Exploratory Teams were combined. This included a Technical Activity focused on the evaluation of Augmented and Mixed Reality in support of Training and Education requirements (HFM-ET-153), and a second Technical Activity aimed at evaluating the use of Augmentation Technology to support real-time performance needs and optimizing operational outcomes (HFM-ET-154). At the guidance of the HFM Panel, the team's assignment was to examine augmentation capabilities across a contextualized spectrum of human performance, and to provide a report that identifies best practices for applying these technology types around task characteristics and associated performance constraints driving augmentation. A secondary goal is to document trends, investments, and evolving opportunities across the research and development landscape. This approach aims to provide recommendations on best practices for managing procurement of augmentation technologies to support performance needs, while identifying research priorities required to mature a capability for use at scale within an operational setting.

Utilizing technology to enhance the performance capability of a human operator is not a new area of critical interest. The continual maturation of various technologies (e.g., wearable devices, augmented reality and multi-modal immersion, assisted cognition, etc.) has created a necessity to systematically evaluate their applicability to training and operational environments across NATO alliance nations, and to do it at appropriate intervals to maintain awareness of the current state of the possible. The requirement to track this technology sector has been met by the initiation of many activities: multiple RTGs of NATO STO HFM and NMSG, and activities sponsored by the NATO Industry Advisory Group (NIAG) have provided background in augmentation technology areas. Many of the original insights on this topic are summarized in NATO Emerged and Emerging Disruptive Technologies (E2DT) #8 "Virtual and Augmented Reality and Cognitive Interfaces" [1]. However, since the conclusion of that work (2010), there have been significant advancements and technology innovations, which have the potential to revolutionize training and operation practices across many job and team functions. These technology trends are summarized in NATO's "Science and Technology Trends 2020 – 2040: Exploring the S&T Edge" [2], which provides further justification and alignment to the performance dimensions reported upon by this RTG.

Of particular interest is the ability of augmentation technology to enhance, mediate and improve learning, performance, retention, and transfer of skills from training to operational contexts. Research within various NATO nations and several commercial enterprises have produced capabilities and technologies to facilitate instructional and intelligently-guided experiences during training and real-time operations and job functions. In addition, performance monitoring, optimization and real-time support through multi-modal immersion are areas of increasing investment in applied research. However, the use of these technologies must be carefully considered across each task and each performance domain they're injected to support, as building generalized claims of the technology effect is not feasible. In other words, the application of supportive technologies must always be fit-for-purpose. An analysis of these evolving technologies will offer NATO a research-informed framework for evaluating this industry sector, and for assisting the community at-large in baselining available tools that meet performance needs within a controlled context. This includes best practices for developing and delivering more effective training (e.g., reduced time/cost to competency, reduced decay and skill fade, better cognitive function) and performance aids that significantly impact outcomes and increase survivability.

It's important to also consider these assessments beyond tactical combat operations. In addition to more traditional peace-keeping and emergency response tasks, the increasing importance of so-called "grey-zone" or hybrid operations [3], whereby hostile operations may occur below the threshold of declared combat operations, will require broader and more flexible skill sets spanning the full range of security and defence missions [4], [5]. To account for this, it is necessary to examine technologies across a wide range of task characteristics that can generalize based on time and performance constraints that the technology operates within. As a result, the nature, extent, availability, and feasibility of these opportunities will be researched and reported utilizing a SWOT analysis framework across multiple augmentation enhancement domains.

1.2 RTG HFM-297 OBJECTIVES AND SCOPING

This report will not provide an all-encompassing representation of the Augmentation Technology sector. Rather, it aims to develop guiding concepts and frameworks with which to evaluate specific application domains for augmentation technologies, and to provide an initial assessment. It will thus be a high-level overview of the task type domains these technologies are applied within, with specific attention focused on the characteristics and effectiveness of their application as it relates to time and performance constraints. This scoping exercise was established due to multiple factors. At a time when there was restructuring of leadership within the RTG itself, the COVID-19 Pandemic began to impact day-to-day life across the globe. The RTG activities were delayed as we learned to manage professional life in a distributed virtual capacity. Additionally, the scope of what would be reported against was adjusted due to the uncertainty of the times, and the extent to which the pandemic would persist and impact required activities to drive an effective RTG.

In view of these obstacles, the objective we collectively defined to steer this RTG is establishing military and operational relevance around performance critical skills and tasks, and leveraging the expertise within the team to provide valid recommendations on how the technology is maturing and where best its application applies. Meeting this objective involves defining relevant use cases and application domains to evaluate the feasibility and strengths of a capability with respect to a performance context. Managing that within an RTG is a challenging endeavor. Accordingly, a theoretical temporal representation of human performance around a hypothetical "Bang" event (i.e., a performance event aligned to a task or sub-task within a mission context) to organize the use cases and provide structure to our assessments was established. This temporal framework will be described in detail in Chapter 2 and will serve as a high-level taxonomy to organize insights on human performance impacts across augmentation technologies of interest.

Furthermore, the aim is to review and analyze opportunities for moving new and emerging augmentation technologies from state-of-art to state-of-practice for training and operations. The objective is to assess the effect of new and emerging system interaction capabilities for individual and units on learning, retention, and performance, and their ability to manage cognitive load and the time/cost to reach a required level of competency through on-the-job support, or through focused training and education practices. This will include identifying weaknesses and opportunities that can inform future research and development investment strategies to support technology-driven training and on-the-job performance requirements. To meet these objectives, the following were completed when investigating augmentation technology within a specified task domain, as reported in the latter chapters of this report:

- Identify current and emerging technologies.
- Execute objective SWOT (Strengths, Weaknesses, Opportunities, Threats) Analysis across a series of defined human performance domains [6].
- Define factors impeding the maturation and adoption of augmentation technologies.
- Discuss research strategies and priorities that will significantly impact desired human performance effects.

1.3 EVOLUTION OF THE ASSESSMENT FRAMEWORK

As mentioned above, the nature of the augmentation technology domain presented challenges for defining a precise scope for the RTG's activities, and for developing an assessment framework for augmentation technology applications. An initial attempt to limit the RTG's activities to a tractable subset of the augmentation technology space was made by defining augmentation in terms of augmenting the user's task environment (e.g., input/output interfaces, artificial intelligence methods for augmenting simulated environments and tasks, monitoring performance to adapt the task environment), rather than augmenting the user (e.g., enhancing perceptual, cognitive or physical capabilities directly across a broad range of tasks). Thus, at a rough cut, mixed and extended reality interface technologies, Artificial Intelligence (AI) for synthetic agents, intelligent tutoring systems, and non-invasive performance monitoring technologies were "in," and invasive neural implants, exoskeleton and pharmacological task performance enhancement were "out." Nevertheless, this first cut still left the problem space for the RTG considerably large in scope.

To further characterize the research space for this activity, the RTG iteratively developed a visual Mind Map of the research topics, leveraging online software. Mind Map exercises are applied to brainstorm a topic, and provide mechanisms to organize, visualize, and clarify the relational components of a specified focus area and theme [7]. This resulted in a high-level ontological framework representing the factors that need to be considered from a technology-influenced performance paradigm (see Figure 1-1). It also provided a more refined definition of the RTG's focus: to explore technologies that directly target training and job aid augmentation approaches that account for limitations associated with cognitive and physical performance. A desired end-state is a system of systems approach that leverages multiple augmentation modalities for the purpose of optimizing mission and training outcomes on a task by task and individual by team basis across the full human performance operational timeline. This includes mechanisms to personalize training on individual strengths and weaknesses, monitoring cognitive and affective states for augmenting interaction components, leveraging data and Artificial Intelligence where feasible to establish overmatch, and building tools that enhance the limitations of the human perceptual systems and cognitive endurance faced in operational situations.

The resulting MindMap categorized the primary factors when considering a framework to drive an evaluation strategy. Ultimately, the primary performance effects and variables considered within our research are aligned to cognitive, procedural and decision-making task requirements that have time and accuracy constraints. With those elements identified, the primary categories driving our assessment were aligned against the Augmentation Stimulus that a technology is managing and the Performer Response a technology is designed to moderate and enhance. This Performer Response is then aligned against a set of Effects and Outcomes that drive the performance measurement constructs. However, each of these categories are dependent on the specific context they are being evaluated within (i.e., when asking how Technology X supports Task Y, the answer will start with "well that depends…"). This includes the type of Task an augmentation technology is influencing, the Environment in which that task and technology are interoperating, ultimately establishing a requirement to deconstruct the review into military relevant applications and use cases that can assist in generalizing the key findings across RTG activities. These criteria and constraints are also represented in the MindMap.

The various elements and constraints identified in the MindMap provided a conceptual background guiding the RTGs activities, including reviewing relevant literature aligned to the attributes of interest and obtaining input from Subject Matter Experts (SMEs) to flesh out performance dependencies at the task level and inform assessment of augmentation applications in those contexts. These associated attributes are defined in the following sections of this chapter.

1.4 AUGMENTATION ATTRIBUTES

In this section, we provide general definitions across the factors of Augmentation Technology, Performance Effects and Outcomes, and Performer Characteristics (see Figure 1-1). These definitions are used to differentiate the human dimension variables associated with a given performance context. The study leverages Pedersen and Duin's human-centric framework of augmentation technology [8] to provide a general definition of the tools and methods in question and the performance effects they are designed to mediate. They identify three technology value systems: 1) Enhancement; 2) Automation; and 3) Building Efficiencies.

1.4.1 Enhancement

As reported by the NIAG Study on augmentation technology [9], and established by the SIENNA Project [10] human enhancement is "a modification aimed at improving human performance and brought about by science-based and/or technology-based interventions in or on the human body." This system is broken down into three sub-categories that address perceptual, cognitive and physical performance dependencies.

Sensory Enhancement is the most common form of human augmentation reported in the literature. The goals of this performance category involve a need to experience more than the natural senses and perceptual systems provide. Sensing augmentation is achieved by creating an enhanced world through digitally reproduced physical and system representations. Through an interfacing modality (e.g., visual, auditory, haptic), established models can be visualized and perceptually superimposed to create an enhanced world within the physical boundaries it was designed (see Figure 1-2). These enhancement techniques are based on perceptual faculties that direct attention, drive anomaly detection and stimuli recognition, triggering decision-making processes and behavioral responses linked to an interfacing environment. When considering augmentation devices and interfaces, establishing sensory enhancement is critical to facilitating cognitive enhancement.

Furthermore, sensory enhancement can align to interfacing virtual technologies that create immersive environments to elicit and support task execution, with varying degrees of fidelity based on the requirements of the user and the task. This spectrum of immersion and interaction is referred to as Milgram's reality-virtuality continuum [11], and is used to operationally define the spectrum of interfacing eXtended Reality (XR) interfaces and the underlying interactions they support. Based on the user objectives and the task constraints, varying types of augmentation tools may be more appropriate than others, especially when considering training use characteristics versus operational use characteristics, and the underlying goals and constraints of the interfacing human performer.

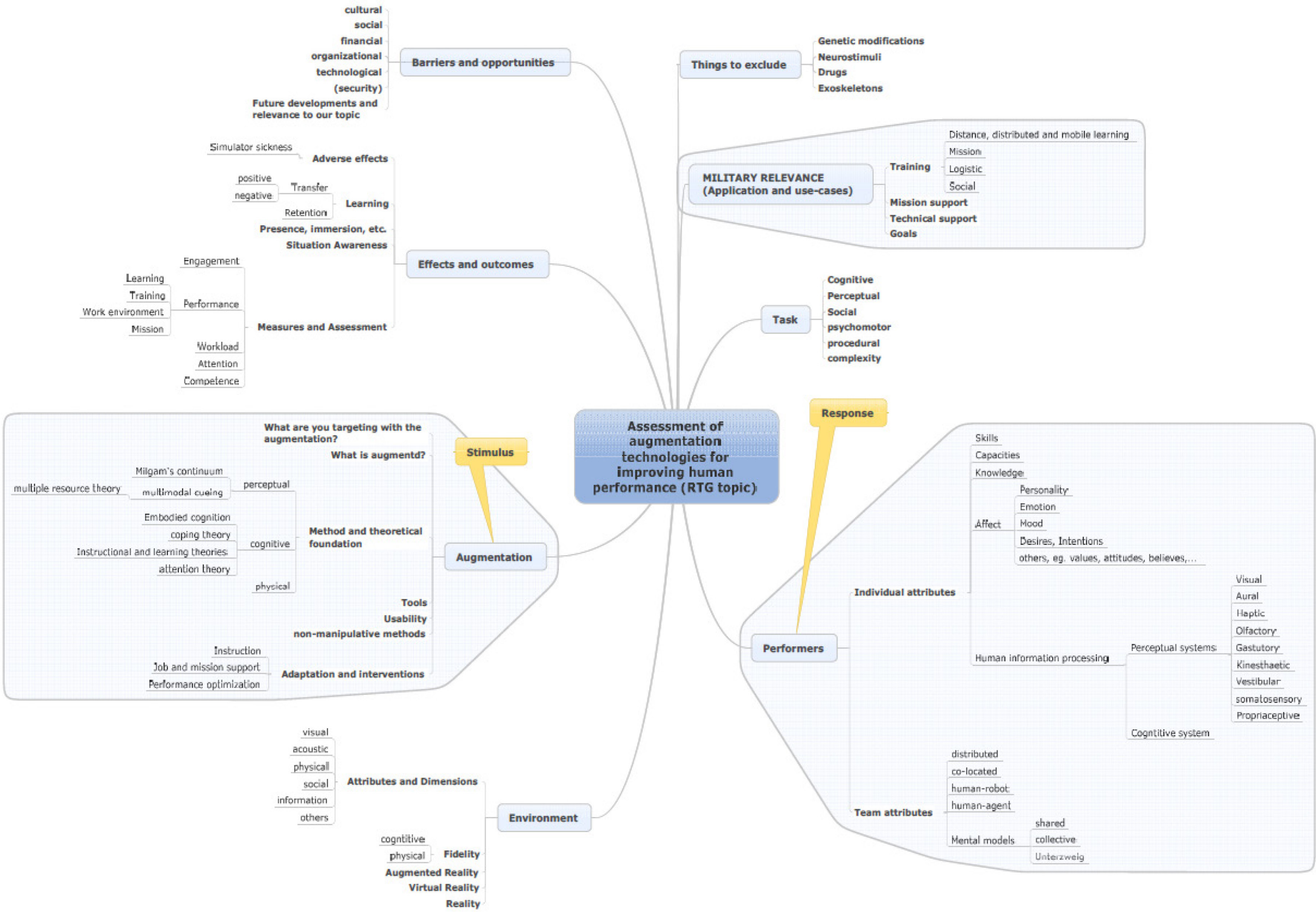


Figure 1-1: Augmentation Technology and Human Performance MindMap.

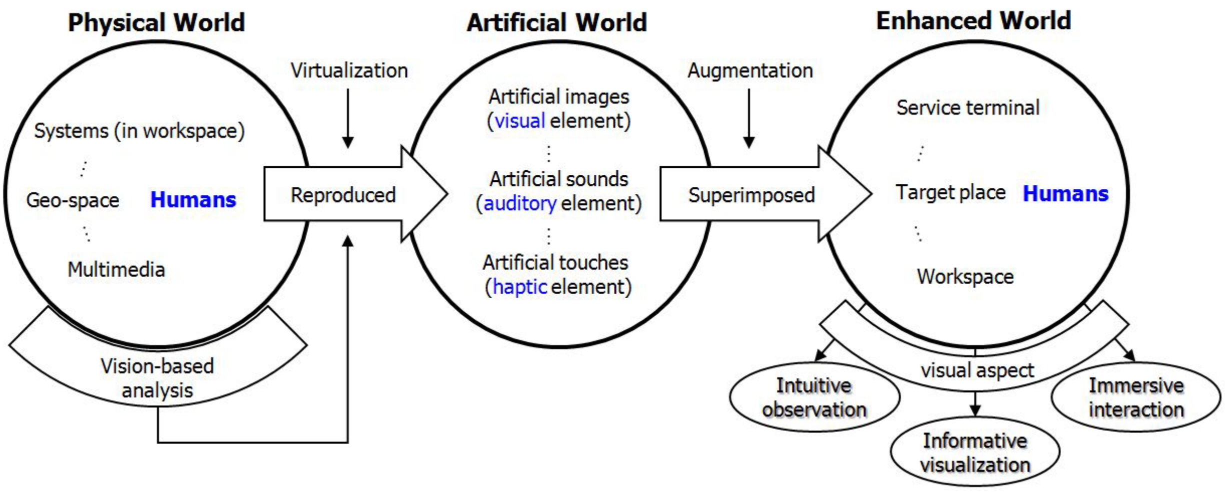


Figure 1-2: Augmentation Model in Support of Sensory and Cognitive Enhancement [2].

Cognitive Enhancement aligns to performance goals that makes an individual smarter, more knowledgeable, remember more, know more, learn faster, learn more efficiently, and override or edit dysfunctional memory [9]. These include tools and methods to influence skill acquisition, leveraging immersive experiential learning paradigms and data-driven techniques to personalize assessment and guide skill acquisition. These technologies are also applied to manage real-time cognitive function during task and mission execution. Technology serves as assisted cognition that addresses specific deficiencies in human perception, cognition, and attentional control. These factors can be aligned to multiple performance variables and constructs, such as cognitive load and workload management, information processing speed and capacity, etc. One construct of importance is situational awareness, which aligns sensory and cognitive domains, and establishes theoretical foundations for predicting environmental behavior and guiding reasoning and decision-making practices. These performance indicators will be reviewed in more detail when discussing building efficiencies with augmentation technology. The following subsections discuss technologies for sensory and cognitive enhancement respectively from a hardware and interaction standpoint.

1.4.1.1 Augmentation Tools and Interfaces

Some of the most common augmentation tools and interfaces used today are related to human perception, with a focus on visual, auditory, and haptic interaction. A large group of technologies in the visual application space are referred to as eXtended Reality (XR) devices. These core tools and methods incorporate virtual, mixed and augmented reality (i.e., VR, MR, AR) paradigms, with an emphasis on enabling humans to interact with an artificial and enhanced world (see Figure 1-2). These technologies are further divided into wearable Head Mounted Displays (HMDs), handheld displays and non-wearable technologies such as spatial displays. Regardless of the interfacing mechanism, these interactive tools enable interaction with an enhanced world through integrated artificial content and visual rendering. While VR and visual caves/domes associate directly with an artificial representation of perceptual stimuli, AR embeds artificial content and interventions within an individual's real-world physical space, producing an enhanced reality with artificial overlays. In other words, AR can be defined as a type of virtual reality in which synthetic stimuli are registered with and superimposed on real-world objects; often used to make information otherwise imperceptible to human senses perceptible [12].

1.4.1.1.1 Wearable Visual Technologies

Most XR visual interfaces make use of an integrated display and an optical system to present visuals to the human eye that are rendered in conjunction with sensor based spatial tracking solutions at high frame rates.

These technologies are the most prominent tools and methods when one thinks of augmentation. It is a volatile tech base, with continual improvements on hardware and software components that are released on a yearly, if not more frequent, basis. At this moment in time, there is no best in class device, and the pace at which this technology is advancing makes it difficult to truly report on their impacts within a human performance frame of reference. As such, monitoring applied research studies over the next five years to empirically evaluate how these technologies influence human performance across a broad spectrum of contexts is important. The goal is to avoid flashy demonstrations that exhibit a maturity level that does not correlate with real-world use cases and constraints.

XR devices can drive mono or stereoscopic presentation of immersive synthetics combined with varying amounts of real-world optical information. Most systems contain a tracking system (head and/or eye position) which maps the wearer's movements and adjusts the images accordingly. Each time the wearer moves their gaze, walks in a particular direction or takes some other form of action, the scene changes accordingly. The tracking system is connected to a computer (which may be tethered to the headset or integrated into it), which adjusts these images so that the wearer is shown a realistic environment with a realistic depth of perception. There are three general types of tracking solutions, two types recording direction and movement of the HMD and one recording movement of the eye:

- **Inside-out tracking:** camera and/or sensors are located on the HMD, no need for other external devices to do tracking.
- **Outside-in tracking:** external sensors, cameras, or markers are required (i.e., tracking constrained to specific area).
- **Eye tracking:** sensors (usually infra-red cameras) built into the device are used to compute the position and movements of the user's gaze and further adjust the scene more precisely than using head-movements alone.

Outside-in tracking has been used by most VR headsets in the past, but to reduce the need of external equipment, set-up time and calibration, inside-out tracking solutions are eventually required by all untethered HMD systems. Outside-in tracking solutions are commonly based on a combination of mechanical gyroscopes and accelerometers, structured lighthouse systems, a suite of interoperable sensors (e.g., ultra-sonic, magnetic, optical, near infra-red and thermal infra-red based sensors), GPS, and a communication network (e.g., WIFI, 5G) Tracking solutions integrating eye movement data in XR devices are starting to emerge (a notable example is the Microsoft HoloLens) and could be leveraged to improve the rendering of scenes through gaze-continent display techniques. In AR and MR applications, gaze information could be used to trigger or more precisely map virtual elements or relevant information in a manner not possible with head movement information alone.

Virtual Reality (VR) Goggles

VR devices are strapped onto the head of the user and make use of an integrated screen to display visuals, while occluding any other external imagery (see Figure 1-3). Usually, each eye is presented with a slightly different image to allow for stereoscopic perception. Furthermore, VR goggles usually employ head tracking, both inside-out and outside-in to directly translate head-movements into the corresponding shift in the picture. There are multiple manufacturers for the devices with varying specifications regarding field of view, screen resolution, refresh rates, tracking accuracy, etc. Some products also have eye tracking features implemented, see Figure 1-3.

Augmented Reality (AR) Goggles

Augmented reality goggles are also strapped onto the head of the user, but unlike VR goggles, these devices make use of an integrated transparent screen to display visuals superimposed on the real world. These goggles normally only employ inside-out tracking to translate head-movements into corresponding shift in overlayed information. The transparent display are normally a waveguide based display allowing full color and resolution imagery but lacks the immersive possibility to provide large field of views in comparison to VR goggles. There are multiple manufacturers for these devices as well with varying specifications regarding field of view, screen resolution, refresh rates, tracking accuracy, etc. and some products also have eye tracking features implemented, see Figure 1-4 and Figure 1-5.

(a) **(b)**

Figure 1-3: Promotional Picture of Commercial HMDs. (a) Consumer graded headset and (b) professional graded headsets used in various military graded synthetic training simulators.

Figure 1-4: Promotional Picture of Commercial Binocular and Monocular AR HMD Systems.

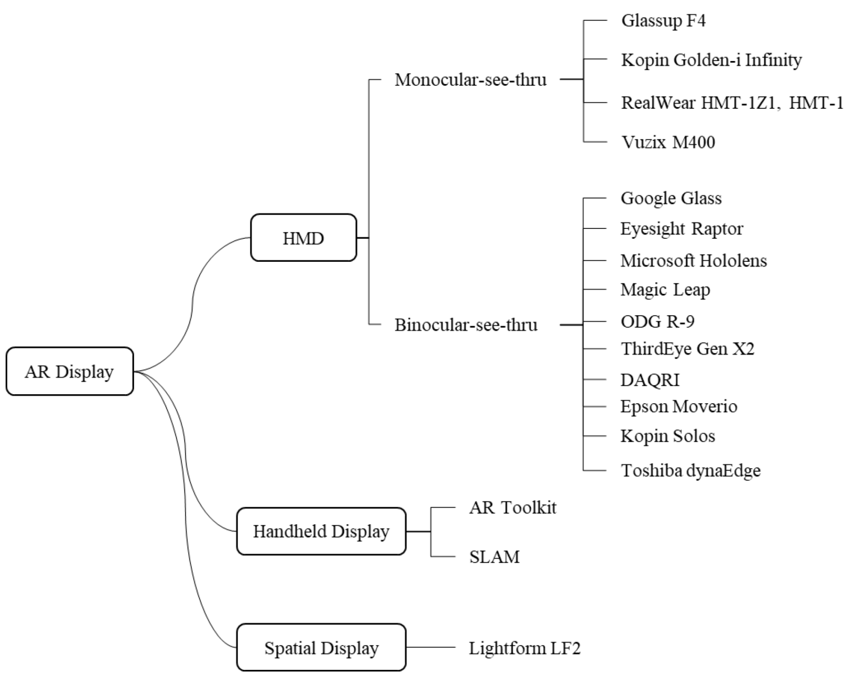


Figure 1-5: Categories of AR Devices [13]. This list is non-comprehensive and provides example tools and components within an established ontology.

Non-Wearable Visual Technologies

When it comes to non-wearable systems, domes use a spherical or hemi-spherical screen on which the image is projected to generate a wide field of view; it is comparable to the technique used in a planetarium [14]. The user in the middle of the dome is then usually wearing shutter glasses (or similar) to perceive the stereoscopic images as intended. Domes also possibly feature body and head-motion capturing systems, allowing to accommodate the user's movement within the given spatial limits of the device [15]. Domes can also feature interactive elements employed through means like handheld controllers.

It has been argued that the spherical shape of the surrounding makes for a more naturalistic and immersive user experience [16]. An example of a dome is provided in Figure 1-6(a) [17]. Cave Automatic Virtual Environments (CAVEs) are very similar to domes regarding the underlying concept; stereoscopic images are projected on screens surrounding the user. The main distinction is the arrangement of the screens, which are flat and usually arranged in a cubic shape, thus not providing a spherical view. An example of a CAVE is shown in Figure 1-6(b).

(a)

(b)

Figure 1-6: (a) Example of a Spherical Dome Projection System and (b) Example of a CAVE Setup.

Touch and Sound Technologies

Haptic interfacing is another perceptual interactive method applied to create an even richer enhanced world through touch and sensation. Haptic sensory aligns with touch and being able to replicate the physical properties of items and objects artificially created to support a realistic interaction in a virtually enhanced world. This involves replicating the shape, weight, feel and physical properties of the artificial world to create higher immersion through rich interaction fidelity.

There are multiple haptic devices that aim to enrich the presentation of the virtual environment in simulation settings. Dangxiao, et al [18] analyzed the development of haptic devices in the last 30 years and argue that currently the technology is following a paradigm they call 'wearable haptics' which includes devices that are worn by the user. According to them the focus lies on hand-worn devices and gloves that provide force feedback, allow for manual interaction and manipulation in a virtual environment and provide sufficient degrees of freedom to accommodate the motion of individual fingers. Examples of such devices are displayed in Figure 1-7.

Figure 1-7: Examples of Haptic Feedback Gloves.

Even though haptic feedback for manual operations seems to be a main area of interest, there are other devices that provide meaningful haptic feedback, like vests with multiple vibration motors, wrist-worn devices, or vibrating or poking surfaces on chairs. Furthermore, there has also been work on the induction of sensation when touching physical surfaces through friction modulation of touch screens [19].

Audio and voice can also be environmentally dependent. From an augmentation perspective, an audio system can leverage with audio perception best practices based on mono, stereo, spatial noise with Head-Related Transfer Function (HRTF) and Anatomical Transfer Function (ATF) creating natural audio experiences. This approach eliminates the effort for mounting peripheral audio device such as headphones, where external headphones require additional cables and can interfere with the ergonomic comfort within a helmet or head mounted display. Integrated audio technologies are further sub-categorized to earpieces that can block substantial amount of background sound, and open sound systems that do not block any real-world sounds. Without visualization aid, audio and voice can in a standalone setting augment various directional based sound sources effective for e.g., warning systems, audio based enhance local situation awareness etc., with direct links to cognitive deficits defined through the Multiple Resource Theory [20], [21].

1.4.1.2 Building Efficiencies

In the context of this report, building efficiencies is defined across two categories, performers and performance. While the individual chapters later in the report address these factors from a performance context, it is useful to briefly define common constructs that inform performance and diagnostic information linked to cognitive requirements to complete a task, and the metrics used to monitor quality of performance.

1.4.1.2.1 Performance Effects and Outcomes

As stated above, and reinforced throughout this report, performance is context driven and linked to a performer, a task, and the environmental and psychological conditions those tasks are performed under. While a task itself has specific criteria to dictate success, there are fundamental performance attributes that extend beyond the task constraints and align to the quality and effects of competency and skill application. This Report briefly presents a subset of these attributes, derived from the NIAG report [9], in the context of decision making, system and task expectations, and team effects. These can serve as high-level benchmarks linked to training and education strategies, along with technologies designed to operate as job aides.

OODA Loop Accuracy and Speed. A common model to align interaction and behavior against in the context of a task is the OODA (Observe, Orient, Decide and Act) Loop framework [22], [23] (Figure 1-8). It provides a structured approach to managing task analyses, and is used to explicitly define the stimuli and conditions that drive decision processes. From an efficiencies standpoint, the primary measures associated with this performance effect are speed and accuracy to a decision, as well as speed and precision of an action. Making sense of the environment through maintained situational awareness is critical to orienting towards a decision. This nuance is present prior to a task initiating, as well as during task execution, as the process is cyclical and continues until the mission in question is complete. There are a lot of factors that can impact an individual or team's OODA loop. The characteristics of the task, the distractors in the task environment (including competing tasks), the competencies possessed by the performer, the cognitive state of the performer, and the emotional state of the performer.

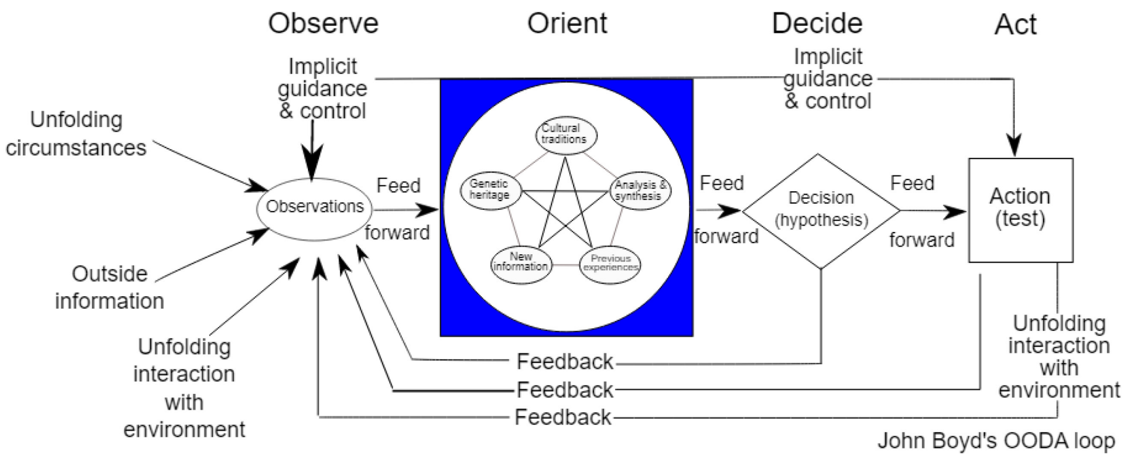


Figure 1-8: Diagram of the OODA Loop by Patrick Edwin Moran [24] Licensed Under CC BY 3.0.

Situational Awareness. A performance construct that aligns directly with OODA loop performance is Situational Awareness (SA) [25]. SA is well-studied in the human factors literature, and associates with an individual's understanding of what is happening around them, and is a critical causal factor of performance in a wide variety of domains, including aviation, air traffic control, military operations, driving, emergency management, healthcare, and power-grid operations [26]. A common and widely accepted framework of SA establishes three levels of performance: 1) Perception; 2) Understanding; and 3) Prediction. From a team perspective, Team SA goes beyond an aggregated representation of individual SA across members of a team, but rather is a critical contributing factor used to establish and maintain a common context within a team structure to make one's interaction and behavior reflected in another's mind [27]. This has been included as a performance indicator because SA will be dependent on the environment and context an individual or team are operating within, and there are specific events, triggers, and stimuli that can be embedded into training approaches and specifically monitored to measure performance.

Several approaches to measuring SA have been researched, with techniques involving self-report, situational judgment tests, and eye tracking. Under controlled and context defined conditions, a common approach to infer an individual's SA at the performance level is the SA Global Assess Technique (SAGAT; [27]), which is based on injected SA prompts and queries during a pause in task execution. Another approach called the Situation Present Assessment Technique (SPAT) differs in that it produces real-time queries to gauge an users understanding of their extended environment. These metrics have limitations in the domains the support, and inferring SA objectively within a dynamic operational setting.

Lethality and Survivability/Longevity. From a tactical standpoint, a link between lethality and survivability is established. This involves all performance metrics that contribute to neutralization and sometimes destruction of enemy components, while maintaining survivability across all friendly and neutrally defined factors. In this representation, a factor can take many forms, including people, structures, vehicles, weapons, and political, economic and health infrastructure. Some tasks have performance effects across both domains, while others are focused primarily on one or the other. These metrics of effectiveness are context dependent and should be explicitly defined for each augmentation supported task to monitor overarching impact of technology utilization.

Team Effects and Resource Allocation. Communication, human-machine teaming, distributed and cross-service and nation coordination. Regardless of the nature of the team members, the performance and effectiveness of teams can be understood as combining factors pertaining to teamwork and those pertaining to team performance or outcomes [28], [29]. The teamwork factors can further be broken down into factors related to communication, coordination, and cooperation [30], whereas team task factors often need to be assessed within a specific task context. In the special case of humans teaming with machines, trust in automation can also play a key role in team performance and effectiveness [31]. A number of approaches to assessing teamwork and team task metrics have been developed, including team SA metrics, team attribute models such as the "Teams Big Five" [32], and Social Network Analysis methods [33]. Such metrics and others need to be considered with, and adapted to, team performance contexts with augmentation technologies.

1.4.1.2.2 Performer Factors and Attributes

There are two essential elements to building efficiencies at the performer level. First, identify the performer factors and attributes that can be objectively monitored and that have a direct impact on performance effects. For this report, these are the cognitive constructs that impact learning and task outcomes and have variation in operationalized definition based on skill development versus skill application. Second, design, research and develop augmentation methodologies and interventions to mediate these cognitive constructs with the goal of influencing performance.

Cognitive Load. The relationship between cognitive load and cognitive functioning is well documented across decades of studies [34], [35], [36], with the common argument that too high of cognitive load can have a negative impact on the performance of working memory. This construct is well documented when considering dual-task environments, with a goal of modelling the human performer's limitations when managing multiple stimuli and task demands. This is an emphasis within the augmentation space. How can technology be applied to moderate cognitive load with a goal of extending the task to performer ratio, with human-machine teaming being the most directed domain of interest?

Fatigue. Fatigue is a measurable construct that aligns to sleepiness and low levels of alertness. This phenomenon has been studied under multiple operational contexts, with sustained effects on performance during continuous and sustained operations that require higher-order cognitive capacities to meet the mission objectives [37], [38]. Measuring fatigue to intervene before performance decrements are incurred is critical. Investigating methods to monitor and mitigate fatigue through augmentation tools and methods are also under consideration in this report.

Presence/Immersion. When engaging with technology and extended reality mediated interactions, immersion is a construct linked to cognitive load and is used to measure an individual's self-perception of engaging with non-physical elements within their physical world [39]. They become immersed in the experience as if it is a part of their natural environment, resulting in high cognitive effort [40]. According to Conkey [41] the concept of 'immersion' comes from situations where technology feeds the human senses with visual, audio, and tactile input through mediated interfaces, creating a perceptual sense of presence within the environment. The perceptual component associated with presence is that interaction invokes response from human senses, human cognition, and affective systems as if the user has a perceptual illusion of non-mediation [39]. This is an important human performer factor when considering desired interaction effects of information and intelligence delivered over augmentation derived channels. The task, the operational environment, and the role of a performer's surroundings in task execution will dictate the level of immersion and presence an augmentation method is designed to achieve. This can be aligned to level of cognitive load required to process and act on a form of information, with direct connections to Multiple Resource Theory to manage the theoretical requirements of balancing intake across perceptual faculties [41].

Signal Detection. In the context of this report, signal detection is a diagnostic attribute linked to situational awareness. This construct aligns to the detection of an anomaly, the identification of that anomaly, and the relation of that anomaly against task conditions and criteria, ultimately leading to a decision on what to do about that anomaly [42], [43]. In this instance, signal detection takes many different shapes and forms based on domain characteristics, with varying degrees of affordance when it comes to latency and accuracy for identifying anomalies in your environment. This performance moderator is also impacted by time, with a vigilance decrement highlighting a human's inability to perform effectively over extended periods of time due to attentional control deficiencies [43], [44]. Investigating assistive augmentation technologies in optimizing signal detection requirements across a broad spectrum of task types is important.

1.4.1.3 Automation

The augmentation category of Automation is generally defined within HFM297. It aligns primarily to the role of data, data visualization, Artificial Intelligence, and machine learning to drive augmentation models, tools, and methods. These processes are used at all levels of augmentation interaction, with an emphasis on managing, adapting, and personalizing the information and interventions a user experiences in support of the task they are executing. Rendering information across devices and human-machine interfaces, automating task allocation or prioritization for a robotic or agent asset, automating performance assessments to drive real-time instructional adaptations, and modelling physiology to manage injects and interventions are just a few of examples of how automation supports augmentation at the human performance level. Ultimately, it establishes the underlying capabilities and dependencies to enable the enhancement techniques and efficiencies described above.

These formalizations follow a common Context-Input-Process-Output-Outcome (CIPO) framework [45], with a reinforcement loop where feasible to drive automated model and policy updates based on the goals and measured effect of augmentation methods. This reinforcement loop is of critical importance to the automation of augmentation delivery and establishing evidence-centered probabilities of performance effects based on the variable a method is designed to mediate. For this purpose, it is critically important to recognize that all forms of automation are not well-suited for all human counterparts the methods were trained to support (e.g., a model that is 85% accurate will mis-diagnose 15% of its classifications, which can be detrimental if that model automates delivery or management of augmentation techniques). Humans are the wild card in this equation, with only probabilistic modelling well-suited to measure the accuracy across automation methodologies.

With that stated, effective automation in support of augmentation is dependent on data access and real-time capture. The operational environment a task is performed within will dictate the available data sources, with main dependencies on local and cloud-based network configurations and open system architectures that enable an interoperable Internet of Things paradigm. In the context of this report, automation is a critical element of all augmentation performance domains with the CIPO framework enabling the separation of automation functions. Whether it's automating the capture and processing of raw data; computing against data to auto generate metrics, features, models, and intelligence; synthesizing, layering and delivering information over enhancement UIs; applying metrics to monitor state and calculate automated assessments and classifications, and using those assessments and classifications to drive augmentation on adaption and personalization.

1.5 SUMMARY

This introductory chapter provided an overview of the HFM-297 RTG objectives. This included a brief overview on the types of technologies, tools and methods considered in this report, along with the measurement constructs that are used to measure impact on human performance. In the following chapters, we present an operational framework that will be applied to manage context alignments across each of the factors introduced above. With a framework in place based on task and human performance constraints, we provide individual chapters on each sector of the operational mission and the utility of augmentation technology to drive optimal performance outcomes.

1.6 REFERENCES

[1] Schlick, C.M., and Alexander, T. (2011). "Emerged/Emerging Disruptive Technologies 8: Virtual and Augmented Reality and Cognitive Interfaces." Position paper for RTO-Led NATO Long Term Capability Requirements. Fraunhofer Institute for Communications.

[2] Reding, D.F., and Eaton, J. (2020). Science and Technology Trends 2020 – 2040: Exploring the S&T Edge. NATO Science & Technology Organization, Neuilly-sur-Seine, France.

[3] Balogh, P. (2017). "Grey Zone Activities – With a Focus on the Social Domain." In Proceedings of Conflicts in the Grey Zone: A Challenge to Adapt, May 9 – 10 2017, Budapest, Hungary. Hungarian Defence Forces General Staff Scientific Research Centre, Budapest, Hungary.

[4] Ministry of Defence (2015). Future Operating Environment 2035. Ministry of Defence, UK.

[5] Scott, K.D. (2016). Joint Operating Environment 2035: The Joint Force in a Contested and Disordered World. Joint Chiefs of Staff Washington United States.

[6] Benzaghta, M.A., Elwalda, A., Mousa, M.M., Erkan, I., and Rahman, M. (2021). "SWOT Analysis Applications: An Integrative Literature Review." Journal of Global Business Insights, 6(1), 55-73. DOI: 10.5038/2640-6489.6.1.1148.

[7] Adlakha-Hutcheon. G., and Jarmasz, J. (2019). "Fostering Innovation in Training and Education through a Novel Technology Assessment Approach." Scientific Letter DRDC-RDDC-2019-L308, Defence Research and Development Canada.p

[8] Pedersen, I., and Duin, A.H. (2021). "Defining a Classification System for Augmentation Technology in Socio-Technical Terms." In Proceedings of the 2021 IEEE International Symposium on Technology and Society (ISTAS), DOI: 10.1109/ISTAS52410.2021.9629174.

[9] Proietti, P., Abate, A. and Sab-Möbus. (2021). Assessment of Human Augmentation Technologies for Exploitation in Battlefield. NATO Industrial Advisory Group (NIAG) Study Group 253.

[10] Warso, Z., and Gaskell, S. (2019). SIENNA D3.2: Analysis of the Legal and Human Rights Requirements for Human Enhancement Technologies In and Outside the EU (Version V2.0). Zenodo. DOI: 10.5281/zenodo.4066617.

[11] Skarbez, R., Smith, M., and Whitton, M.C. (2021). "Revisiting Milgram and Kishino's Reality-Virtuality Continuum." Frontiers in Virtual Reality, 2, 647997.

[12] M&S Glossary – Terms & Definitions. Modeling and Simulation Enterprise (MSE) (05 October 2015) NATO Modelling and Simulation Related Terms for NATO Term, The Official NATO Terminology Database Ver 05, 1. https://www.sto.nato.int/publications/Pages/ModellingandSimulationGlossary.aspx

[13] Xu, J., Doyle, D. and Moreu, F. (2021). State of the Art of Augmented Reality (AR) Capabilities for Civil Infrastructure Applications. Arxiv Human-Computer Interaction. Cornell University. Retrieved 30 May 2022 from https://arxiv.org/abs/2110.08698

[14] Chen, J. (2002). A Virtual Environment System for the Comparative Study of Dome and HMD. Doctoral dissertation, University of Houston.

[15] Grogorick, S., Überheide, M., Tauscher, J.P., Bittner, P.M., and Magnor, M. (2019). "Gaze and Motion-Aware Real-Time Dome Projection System." In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 1780-1783.

[16] Lantz, E. (1996). "The Future of Virtual Reality: Head Mounted Displays versus Spatially Immersive Displays (Panel)." In Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, pp. 485-486.

[17] Grogorick, S., Überheide, M., Tauscher, J.P., Bittner, P.M., and Magnor, M. (2019). "Gaze and Motion-Aware Real-Time Dome Projection System." In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 1780-1783.

[18] Dangxiao, W., Yuan, G.U.O., Shiyi, L.I.U., ZHANG, Y., Weiliang, X.U., and Jing, X. (2019). "Haptic Display for Virtual Reality: Progress and Challenges. Virtual Reality & Intelligent Hardware, 1(2), pp. 136-162.

[19] Bernard, C., Monnoyer, J., Ystad, S., and Wiertlewski, M. (April 2022). Eyes-Off Your Fingers: Gradual Surface Haptic Feedback Improves Eyes-Free Touchscreen Interaction. In CHI Conference on Human Factors in Computing Systems, pp. 1-10.

[20] Wickens, C.D. (2008). "Multiple Resources and Mental Workload." Human Factors, 50(3), pp. 449-455.

[21] Jeffri, N.F.S., and Rambli, D.R.A. (2021). "A Review of Augmented Reality Systems and their Effects on Mental Workload and Task Performance." Heliyon, 7(3), e06277.

[22] Vettorello, M., Eisenbart, B. and Ranscombe, C. (2019). "Toward Better Design-Related Decision Making: A Proposal of an Advanced OODA Loop." In Proceedings of International Conference on Engineering Design, 05 – 08 August, Delft, the Netherlands.

[23] Mednikarov, B. and Lutzkanova, S. (2021). "Applicability of the OODA Loop Theory for Analysis of Human Error Processes in the Maritime Safety." Pedagogy, 9(7), 241-250.

[24] OODA Loop (n.d.). Wikipedia. https://en.wikipedia.org/wiki/OODA_loop (Accessed 01 May 2022) from

[25] Endsley, M. (2021). "Situational Awareness." In G. Salvendy and W. Karwowski (Eds.), Handbook of Human Factors and Ergonomics. 5th Edition, Wiley Online Library. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119636113.ch17 (Accessed 02 June 2022)

[26] Shu, Y. and Furuta, K. (2005). "An Inference Method of Team Situation Awareness Based on Mutual Awareness." Cognition, Technology & Work, 7, pp. 272-287. https://link.springer.com/article/10.1007/s10111-005-0012-x

[27] Endsley, M. (2019). "A Systematic Review and Meta-Analysis of Direct Objective Measures of Situation Awareness: A Comparison of SAGAT and SPAM." Human Factors: The Journal of the Human Factors and Ergonomics Society, 63(1), pp. 124-150. https://journals.sagepub.com/doi/full/10.1177/0018720819875376

[28] Duncan, M. (2015). "Team Effectiveness in Collaborative Operations: Final Project Report." DRDC Toronto Scientific Report DRDC-RDDC-2015-R120, Defence Research and Development Canada.

[29] Huddlestone, J., and Pike, J. (2016). "Team and Collective Training Needs Analysis: Defining Requirements and Specifying Training Systems." Ashgate Publishing Ltd.: Farnham, UK.

[30] Wilson, K.A., Salas, E., Priest, H.A., and Andrews, D. (2007). "Errors in the Heat of Battle: Taking a Closer Look at Shared Cognition Breakdowns through Teamwork." Human Factors, 49(2), pp. 243-265.

[31] McNeese, N.J., Demir, M., Chiou, E.K., and Cooke, N.J. (2021). "Trust and Team Performance in Human-Autonomy Teaming." International Journal of Electronic Commerce, 25(1), pp. 51-72. DOI: 10.1080/10864415.2021.1846854

[32] Salas, E., Sims, D. E., and Burke, C. S. (2005). "Is there a "Big Five" in Teamwork?" Small Group Research, 36(5), pp. 555-599. DOI: /10.1177/1046496405277134

[33] Houghton, R.J., Baber, C., McMaster, R., Stanton, N.A., Salmon, P., Stewart, R., and Walker, G. (2006). "Command and Control in Emergency Services Operations: A Social Network Analysis." Ergonomics, 49(12-13), pp. 1204-1225.

[34] Vanneste, P., Raes, A., Morton, J., Bombeke, K., Van Acker, B.B., Larmuseau, C., Depaepe, F., and Van den Noortgate, W. (2021). "Towards Measuring Cognitive Load through Multimodal Physiological Data." Cognition, Technology & Work, 23, pp. 567-585. DOI: /10.1007/s10111-020-00641-0

[35] Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S., Khawaji, A., and Conway, D. (2016). "Robust Multimodal Cognitive Load Measurement." Hum Comput Interaction Ser Cham. DOI: /10.1007/978-3-319-31700-7

[36] Johannsen, G. (1979). "Workload and Workload Measurement." In N. Moray (Ed.) Mental Workload: Its Theory and Measurement. Springer, Boston, pp. 3-11.

[37] Durmer, J.S., and Dinges, D.F. (2005). "Neurocognitive Consequences of Sleep Deprivation." Semin Neurol. Mar 25(1), pp. 117-29. DOI: 10.1055/s-2005-867080. PMID: 15798944

[38] Killgore, W.D. (2010). “Effects of Sleep Deprivation on Cognition.” Prog Brain Res. 185, pp. 105-29. DOI: 10.1016/B978-0-444-53702-7.00007-5. PMID: 21075236.

[39] Donghee, S. (2019). “How Does Immersion Work in Augmented Reality Games? A User-Centric View of Immersion and Engagement.” Information, Communication & Society, 22(9), pp. 1212-1229. DOI: 10.1080/1369118X.2017.1411519

[40] Murphy, C., Chertoff, D., Guerrero, M., and Moffitt, K. (2011). “Creating Flow, Motivation, and Fun in Learning Games.” In The Design of Learning Games, Springer-Verlag.

[41] Conkey, C.A. (2011). Machinima and Video-Based Soft Skills Training. University of Central Florida.

[42] Wixted, J.T. (2020). “The Forgotten History of Signal Detection Theory.” Journal of Experimental Psychology: Learning, Memory, and Cognition, 46(2), p. 201.

[43] Carpenter, J., Sherman, M.T., Kievit, R.A., Seth, A.K., Lau, H., and Fleming, S.M. (2019). “Domain-General Enhancements of Metacognitive Ability through Adaptive Training.” Journal of Experimental Psychology: General, 148(1), p. 51.

[44] Al-Shargie, F., Tariq, U., Mir, H., Alawar, H., Babiloni, F., and Al-Nashash, H. (2019). “Vigilance Decrement and Enhancement Techniques: A Review.” Brain Sciences, 9(8), p. 178.

[45] Gao, Y., and Janssen, M. (2020). “Generating Value from Government Data Using AI: An Exploratory Study.” In G. Viale Pereira, M. Janssen, H. Lee, I. Lindgren, M. P. Rodríguez Bolívar, H. J. Scholl, and A. Zuiderwijk (Eds.), Electronic Government – 19th IFIP WG 8.5 International Conference, EGOV 2020, Proceedings, pp. 319-331. (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 12219 LNCS). Springer. DOI: /10.1007/978-3-030-57599-1_24.

Chapter 2 – EVALUATING TECHNOLOGY AGAINST A HUMAN PERFORMANCE TEMPORAL CONSTRUCT

Benjamin Goldberg

Establishing a one-size-fits all definition of human performance is not realistic. Performance is a context driven measurable construct that aligns around a task, a performer, and underlying criteria to satisfy the objectives of a task or mission. Regardless of the augmentation technologies and methods discussed throughout this report, the context driving its application is everything. From a reporting standpoint, deriving common task characteristics across contexts can assist in evaluating enhancement techniques across augmentation modes and creating context-general insights on strengths and weaknesses from a human performance improvement standpoint. In this instance, we establish common performance dimensions highlighted in Bloom's four domains of learning and doing:

1) Cognitive performance;
2) Physical/psychomotor performance;
3) Affective performance; and
4) Performance in the social domain [1].

Each dimension provides further diagnostics on performance as it aligns against specific context representations of task types and the conditions and standards under which those task types are completed. In this chapter, a set of generalized performance contexts are defined against a high-level operational timeline. This will serve as the framework to organize Strengths, Weaknesses, Opportunities, Threats (SWOT) Analyses in the chapters to follow. The established contexts are aligned against a performance timeline that manages performance requirements before, during and after the execution of a task or job function while differentiating what aspects of performance are monitored and influenced within each explicit phase on that timeline.

2.1 PERFORMANCE DOMAINS LEFT AND RIGHT OF A "BANG" EVENT

While a task and mission has its own performance definition, the performer has measurable insights that serve as diagnostics and causal factors around the performance outcomes produced during execution of a task. It is this latter representation of performance that is of most interest to HFM-297, in addition to evaluating the effects various augmentation technologies have on those skills, processes and behaviors. To provide an evaluation of augmentation technology on human performance, a strategy is required to define generalized context categories that influence the factors of skill development and application, and to use those categories as a means to analyze the impact of augmentation tools and methods in a controlled and objective manner.

For this report, an Operational Timeline Framework has built to represent the dynamic nature of human performance in relation to so called a 'Bang Event.' The associations presented are extensions of the construct presented by Vanhorne and Riley [2] in their book 'Left of the Bang: How the Marine Corps' Combat Hunter Program Can Save Your Life'. In our implementation, we established a full temporal timeline of common Human Performance Domains, with each domain serving as a mechanism to the preparation and execution of competencies as they relate to a specific task or mission (i.e., the Bang Event). In the following sub-sections, a description of the Proactive and Reactive relationship of time around a designated Bang Event, and introduction of the domains that drive the interactions and outcomes of individual and team related task and performance functions are provided. These associations will then be used to organize a series of chapters that review the current state of augmentation technologies applied around established contexts.

2.1.1 What Is the BANG Event?

What does a Bang Event mean in the study context? A bang event references an explicit moment in time when an individual or team are initiating the execution of a task based on the conditions of their operational environment and job duties. For the purposes of this report, the bang event is a very general term that highlights the shift from Proactive preparatory work to Reactive human performance skill activation based on the requirements to satisfy the task's established objectives and standards.

Regardless of an individual's job function, occupational specialty and/or team's assignment, the Bang Event has general associations that align to a task environment, the events and triggers that initiate task execution, and the standards under which those tasks will be assessed and how the performers will be evaluated. In other words, Context is King. These variables are used to drive front-end task analyses to systematically define the cognitive, physical, social and affective performance demands required by the human performer to meet those task criteria. Whether you're a pilot flying an aircraft, a mechanic repairing a vehicle, a combat medic treating a casualty, a first-responder coordinating with a command center, an operator controlling several unmanned assets, or an infantryman clearing a structure, the Bang Event can have dramatically different definitions. However, there are discrete and severable performance domains that align left and right of the Bang, each providing a set of contexts and boundaries for evaluating how, and to what effect, technology impacts those performance characteristics and performer objectives (see Figure 2-1).

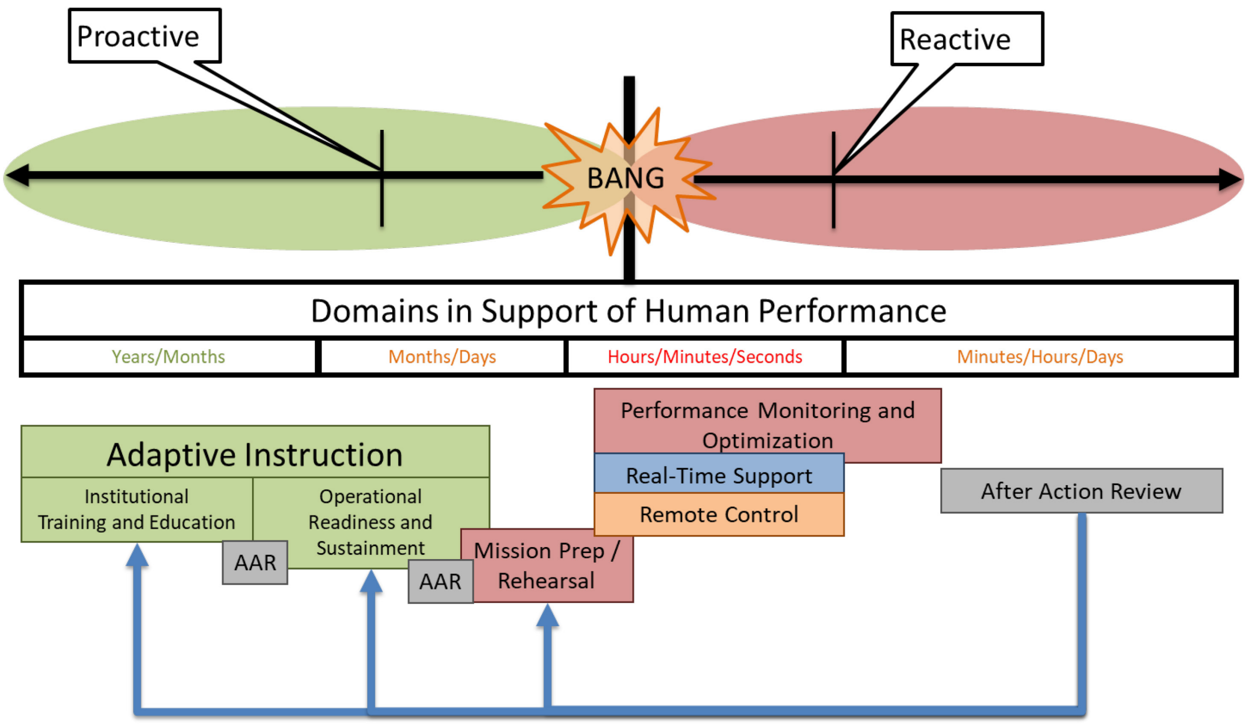


Figure 2-1: The Domains of Human Performance Across an Operational 'Bang Event' Timeline.

2.1.2 Left of the BANG Event

Considering human performance Left of the Bang Event, the Proactive categories and activities that influence performance outcomes Right of the Bang are examined. These overlapping phases are established to produce the highest probability of success when a Bang Event is experienced. This is reliant on the development, sustainment, and preparation of knowledge, skills and behaviors to drive success when executing real-time job functions. Each high-level domain is based around an extended calendar perspective (i.e., years and months vs weeks and days vs hours, minutes, and seconds/milliseconds), with timeline categories serving as explicit differentiators of variations in time constraints and the role they serve in preparing an individual or team for their job function.

The proactive domains reported on in focused subsequent chapters include:

1) Adaptive instruction supporting institutional training and education;
2) Adaptive training supporting operational readiness and sustainment;
3) Mission preparation and rehearsal functions near-term to a Bang Event; and
4) The utilization of After Action Review processes to optimize the iterative refinement within and across each performance phase.

The following sub-sections provide brief descriptions of each phase without the consideration of technology use and impact. These definitions have been established by the HFM-297 team to associate directly with the operational timeline represented in Figure 2-1.

2.1.2.1 Institutional Training and Education

The institutional training and education performance domain is focused on the delivery of controlled Programs of Instruction that target and develop basic and advanced occupational and foundational competencies. It accounts for the largest timeline on the 'Bang Event' framework and utilizes best practices in instructional design and cognitive science to guide the development and retention of specialized knowledge, skills and behaviors.

2.1.2.2 Operational Readiness and Sustainment

Following completion of schoolhouse requirements, human performance objectives shift towards preparation and training activities prior to a mission or task assignment. These account for Proactive Left of Bang engagements used to train and evaluate readiness at the team and collective level and to target and sustain core competencies. This phase of performance is managed through organizationally defined criteria for certification and re-certification. Training requirements are aligned to essential tasks and competencies, and are prioritized across several factors, including recency, recognized skill deficiencies, and leadership guidance.

2.1.2.3 Mission Preparation and Rehearsal

In accordance with the operational timeline presented in Figure 2-1, mission prep and rehearsal are associated with activities in close time proximity to a defined Bang Event. The operational environment is better defined, and planning and preparation tasks and workflows align to an established context. Dependent on the assignment and roles involved, these efforts can begin days, hours, or even minutes before a mission is initiated. This involves building plans and courses of action from existing situational awareness and underlying intelligence reporting. From a tactical view, wargaming is conducted to test and validate a plan through rehearsal procedures. The amount of effort dedicated to planning, preparation, and rehearsal is dependent on the task, job, and team type, along with the tolerance for error in the execution of a task. This serves as the final pathway before a bang event is experienced and Proactive skill application is required.

2.1.2.4 After Action Review

After Action Review (AAR) serves a critical role across all interconnected human performance domains. At each relevant opportunity, AAR practices are defined to encourage and stimulate focused reflection on experiences, outcomes, and overarching objectives, with a goal to drive behavioral change and performance improvement. The performance domain as it relates to time constraints and performance impact goals (i.e., building competency, sustaining competency, preparing for application of competency) will dictate the AAR best practices and evaluation criteria (i.e., assessing if an AAR was effective). The AAR has relevance at many different points of the performance timeline; however, this report focusses on its proactive, "Left of Bang" (e.g., training or mission rehearsal) applications.

2.1.3 Right of the BANG Event

In the context of this report, the 'Right of the Bang Event' is considered as the explicit execution of tasks and procedures based on operational events, task triggers, and shifting conditions and contexts. While Left of the Bang Event is preparatory in nature, Right of the Bang establishes a Reactive application of human performance to drive objective outcomes. In this instance, the focus on augmentation technology is to ensure overmatch, no matter how Proactive an individual or team was in preparing for any Bang Event they can experience while performing their job. In the Proactive category, the performance domains are initiated just before a task's Bang Event trigger, with certain human performance characteristics requiring optimization to better identify and respond to Bang Event stimuli and conditions. Based on the necessity to impact performance metrics and increase probability of desired outcome, the sub-sections for Right of the Bang Event align to capability enhancement categories that dictate the role and user requirements of augmentation requirements. These sub-sections include Real-time Support, Remote Control, and Monitoring and Optimization.

2.1.3.1 Real-Time Support

Real-time Support and Remote Control enable soldiers to deploy capabilities supporting missions and operations on strategic, tactical and technical functions during operation. In the context of this report, the timeline conditions associated with human performance are defined in the activities just prior to a Bang Event, and all of the subsequent reactions based on operational and environmental context. The time dependencies and constraints linked to real-time support can be hours, minutes, or even seconds and microseconds, with the task and job function dictating those relationships.

2.1.3.2 Remote Control

A remotely controlled system allows an operator to operate a system without having to be in close proximity to it. Broadly speaking there are two types of remote-controlled systems. The first of which is unmanned systems, which may or may not have a degree of autonomy. The second type of system is a Telexistence system that allows a user to project their presence to another environment to control a system. This relies on the integration of telepresence which allows the user to see and hear in the remote environment, robotics which enables the user to move in the remote environment and haptics which enables to user feel the remote environment. Telexistence had been enhanced by rapid advances in immersive technologies such as virtual and augmented reality haptics, robotics and computer vision.

2.1.3.3 Monitoring and Optimization

To achieve mission success consistently and under vary conditions, no matter the task, it is critical to maintain cognitive and physical readiness. In the context of this reporting activity, there is interest in the monitoring and optimization of performance constructs and their interacting behavioral and physiological dependencies. A specifical interest is in the variable that can be modelled through viable data sources and managed through augmentation methods. The tools and methods to support this performance domain begin shortly before a bang event with a goal of maintaining alertness and situational awareness to manage the stimuli in an individual's or team's surroundings and right of bang to moderate stress response and distractors while maintaining cognitive engagement on associated task performance and effects.

2.2 ASSESSMENT FRAMEWORK DEFINITIONS

To assist in this reporting effort, the RTG utilized the Strengths, Weaknesses, Opportunities, Threats (SWOT) analysis framework as a guiding structure to describe the state of the art in augmentation technology. A SWOT analysis is presented for each topic area, with a high-level emphasis on how these capabilities can be applied today, and what weaknesses and opportunities should prioritize research investments. serves as a fundamental tool for organizations to assess their position on a market and is applied to evaluate the internal and external factors during times of indecision [3], [4].

2.2.1 The SWOT (Strengths, Weaknesses, Opportunities, Threats) Analysis

A SWOT Analysis provides a methodology to evaluate the internal characteristics of a technology's application against a context within a human performance domain and the external factors that impact effectiveness, adoption and overall maturation of a technology sector. This a controlled approach to compare and contrast the quality of tools and methods applied against different contexts and use cases within a performance domain. Internal characteristics align to the observable and measurable strengths and weaknesses of a technology, and the external factors are linked to opportunities and threats at the organizational and technological level. Figure 2-2 presents a typical tabular framework used to present SWOT analyses.

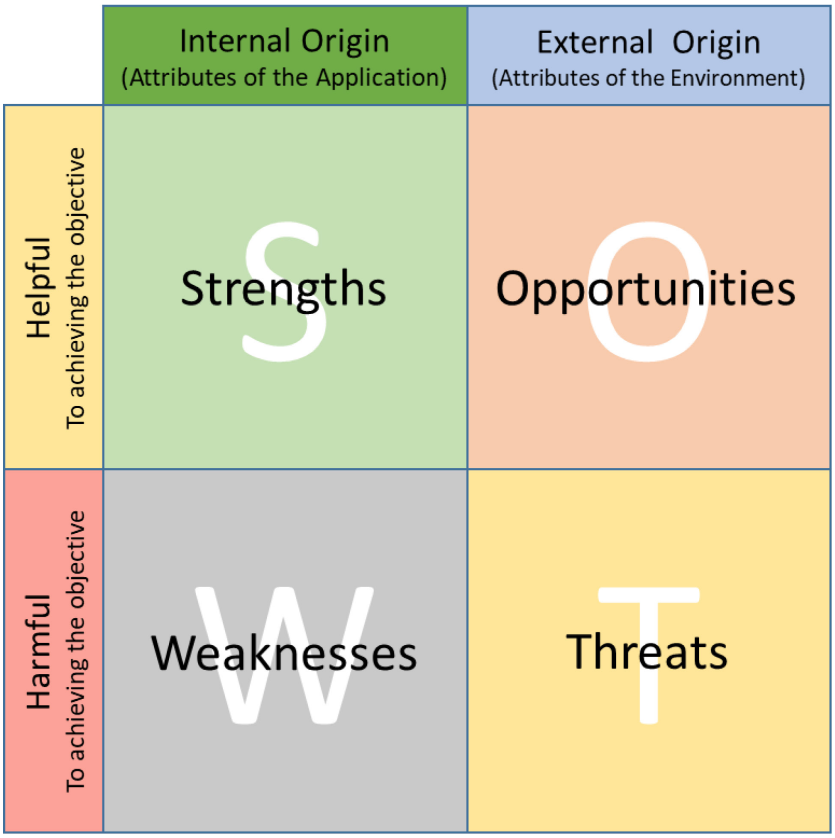


Figure 2-2: Internal and External Organizational Factors Aligned to a SWOT Framework.

For this report, strengths and weaknesses are associated with metrics identified in chapter one linked to the human performer and the task characteristics driving performance measurement identified in in this chapter. This approach is applied to recognize the task types and conditions an augmentation technology is measured against, with a goal of establishing cross-task recommendations based on the characteristics that define their application. This can be used be decision makers to identify the task types a technology is best suited from an acquisition standpoint, while identifying specific capability gaps that can be addressed through research and development investments. Ideally, these claims are based on empirical evidence collected through controlled studies, but an overarching review to into account claims void of that consideration.

The external factors are a little more abstract in definition than strengths and weaknesses. Opportunities and Threats are dictated by infrastructure, organization, market trends, and customer needs that can positively or negatively impact the maturation of capability. This is especially critical in defining in tracking in relation to research and development investments. In addition, this category of external attributes can serve as direct causal factors that promote acceptance or serve as a barrier to technology transition and wide scale user adoption.

2.3 TOPICS TO BE COVERED

The first two chapters in this report provide a definition of augmentation technologies and an operational timeline by which the technologies will be evaluated. This was accomplished by defining sub-categories of augmentation associated with enhancement categories, hardware and interfacing requirements, data dependencies, analytics and Artificial Intelligence, and intervention types. In addition, by creating a temporal relationship across human performance domains in the framework of a mission, the sub-categories will be examined to determine what contexts and use cases are best suited for an augmentation tool or method. The chapters to follow provide an overview of technology considerations for each performance domain in Figure 2-1. This will include a SWOT analysis to assist in determining the impact of augmentation technology. The topics covered across the impact evaluation chapters are briefly introduced below.

2.3.1 Augmentation Technology for Adaptive Instruction for Accelerated Readiness (Chapter 3)

Adaptive Instructional Systems (AISs) provide a means of augmented human performance by accelerating the training timeline [5], improving mastery of levels of learned material [6] and enhancing rate of retention of learned material. AISs are defined as “computer-based systems that guide learning experiences by tailoring instruction and recommendations based on the goals, needs, and preferences of each learner in the context of domain learning objectives” [7]. They accelerate knowledge and skill acquisition through individualized instruction, balanced feedback, and coaching; and adjust the level of interaction based on the needs of each individual learner or team. These adjustments vary as a function of instructional method, performance, retention, and the need to transfer newly-learned skills to the operational setting. AISs are domain agnostic, enabling implementation in a wide range of training domains. With that said, AISs are most readily developed when the domain context is well-defined. Chapter 3 will provide an overview of AISs, applications that have been developed and conclude with a review of their strengths and opportunities as well as weaknesses and threats.

2.3.2 Augmentation Technology for Mission Prep and Rehearsal (Chapter 4)

Historically, mission planning was an activity that unfolded on white boards and using physical models of the operational environment at various levels of fidelity. However, with the increased emphasis on joint all-domain operations and the startling degree of complexity involved in planning how to bring integrated effects to bear on a peer adversary in a contested environment, these traditional tools are no longer adequate for the planning and rehearsal challenges faced by today’s militaries. Mission planning must bring together the individuals who will execute the mission to identify mission priorities, consider threats and alternative scenarios, and establish ‘contracts’ that represent a shared understanding of how individuals and teams will react to deviations to the plan to maximize effectiveness.

2.3.3 Augmentation Technology for Real-time Support and Remote Control (Chapter 5)

Real-time Support and Remote Control offers soldiers to deploy functionalities supporting missions and operations on strategic, tactical and technical functional level during operation in the timeframe domain supporting human performance during bang. A real-time system is a system software that is designed to carry out different tasks simultaneously with real-time output. The execution of tasks is implemented in a fixed time basis without any delay. Even if the system undergoes the same failure in different occasions, there will be no difference in the results.

The effects, dependencies and human performance construct is summarized to:

- Effects:
  - Decision support and validation of decision making.
  - Aid on daily work.
  - Tools to gather information and operate in complex environments.
- Dependencies:
  - Data network – speed, bandwidth, coverage, reliability and redundancy.
  - Cyber security and privacy.
  - Local and central compute.
  - Interoperability within NATO organizations and to external resources.
  - Commercial off the shelf technology.
  - Artificial Intelligence and Machine Learning (AI/ML).
- Human Performance Constructs:
  - Cognitive load.
  - Working memory / multi-task paradigm.
  - Cyber sickness.
  - Physical load.
  - Situation awareness.

2.3.4 Augmentation Technology for Performance Monitoring and Optimization (Chapter 6)

Cognitive performance is a function of many factors including level of training/expertise, motivation level, amount and quality of feedback, etc. However, the upper limit of cognitive performance at any given timepoint depends on the brain's physiological capacity to productively engage in mental work, and this is largely determined by two factors: a) extant sleepiness level (the product of sleep debt level x circadian rhythm phase); and b) extant fatigue level (the product of 'time on task' x cognitive load). Subjective (e.g., rating scales) and objective (e.g., psychophysiological) monitoring methods each have advantages and drawbacks, as does monitoring of actual operational performance. Multimodal monitoring (i.e., systems that include each type of monitoring method) can optimize sensitivity and specificity. However, accurate interpretation and optimal utilization of these data (e.g., to accurately set red, yellow, and green thresholds; to identify meaningful trends in performance over time; and to specify optimal timing and dosing of interventions to prevent operationally-relevant declines in performance [e.g., rest, sleep, caffeine, etc.]) can best be achieved via real-time application of Mathematical Performance Prediction Models (MPPMs). The type of intervention applied to facilitate cognitive performance depends on the underlying cause of the cognitive performance deficit (or MPPM-predicted deficit) and the operational situation/exigencies. Nootropic (cognitive performance enhancing) pharmaceuticals are stimulants (e.g., d-amphetamine, modafinil, caffeine), which can be useful for short-term enhancement of performance, but can disrupt subsequent sleep (thus exacerbating the problem). Likewise, hypnotic medications can facilitate the recuperative effects of sleep on cognitive performance, but can also directly impair performance (e.g., via drug "hangover" effects). Next-generation interventions (e.g., transcranial electrical stimulation to enhance both sleep and alertness) that potentially have no (or at least fewer) downsides are currently under investigation.

2.3.5 Augmentation Technology for After Action Review (Chapter 7)

The After Action Review (AAR), also known as the debrief, team huddle, or other names, is a well-established performance improvement practice. From its origins in US Army collective training, it has become a widely-used intervention across military services, in non-military training, and for reviewing operational activities. Despite its many names and formats, its basic logic remains the same: following a performance event, a review of specific activities is prepared whereby they are assessed against the performance objectives for the event; then the review is conducted as a discussion of the event with the participants, to increase their own self-reflection about their performance and set goals for improving future performance. While this process is technology-agnostic, augmentation technologies have always been at the core of the AAR process, starting with the advent of instrumented training ranges in the 1970s. With the development of synthetic and distributed task and training environments, the range of augmentation technologies applied to support the human participants in the conduct of AARs has grown: performance monitoring technologies and automated metrics have emerged to support the observation and assessment of the performance event itself, while virtual and augmented visualization technologies have been applied to support the performance discussion, particularly in synthetic and distributed environments. The chapter on AARs in this report will focus on AAR as a training intervention and will leverage recent meta-analyses on AAR effectiveness to discuss the user limitations that have driven the application of augmentation technologies in AARs and give an overview of the technologies themselves. The strengths and weaknesses of augmentation technologies for AAR, as well as the opportunities and threats facing them, will be discussed, with a view to providing recommendations on the near-term use of augmentation technologies for AARs and identifying future research investment areas.

2.4 REFERENCES

[1] Dettmer, P. (2005). "New Blooms in Established Fields: Four Domains of Learning and Doing." Roeper Review, 28(2), pp. 70-78.

[2] Riley, P., and Van Horne, P.V. (2014). Left of Bang. New York, Black Irish Entertainment.

[3] Benzaghta, M.A., Elwalda, A., Mousa, M.M., Erkan, I., and Rahman, M. (2021). "SWOT Analysis Applications: An Integrative Literature Review." Journal of Global Business Insights, 6(1), pp. 55-73. DOI: 10.5038/2640-6489.6.1.1148

[4] Rozmi, A.N.A., Nordin, A., and Bakar, M.I.A. (2018). "The Perception of ICT Adoption in Small Medium Enterprise: A SWOT Analysis." International Journal of Innovation Business Strategy, 19(1), pp. 69-79.

[5] Bloom, B.S. (1984). "The Two Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring." Educational Researcher 13(4-6).

[6] Champney, R., Milham, L., Parrish, T., Kokini, C. and Stanney, K. (2010). "Automating Instruction Support: Insights from Field Research." SpringSim '10: Proceedings of the 2010 Spring Simulation Multiconference, April 2010, Article 31, pp. 1-6. DOI:10.1145/1878537.1878570

[7] Craven, P.L. (2021). "Blended and Adaptive Learning in Ground School Instruction for Aviators." 2021 Interservice / Industry, Training Simulation and Education Conference Proceedings, 29 Nov – 2 Dec 2021. https://www.xcdsystem.com/iitsec/proceedings/

Chapter 3 – ADAPTIVE INSTRUCTION AND ACCELERATED READINESS

Elizabeth Biddle and Thomas Herzig

3.1 INTRODUCTION

Adaptive Instructional Systems (AISs) provide a means of augmented human performance by accelerating the training timeline [1], improving mastery levels of learned material [2] and increasing rate of retention of learned material. Adaptive Instructional Systems are defined as "computer-based systems that guide learning experiences by tailoring instruction and recommendations based on the goals, needs, and preferences of each learner in the context of domain learning objectives" [3]. AISs are derived applications that accelerate knowledge and skill acquisition through individualized instruction, balanced feedback, and coaching. AISs adjust the level of interaction based on the needs of each individual learner or team. These adjustments vary as a function of instructional method, performance, retention, and the need to transfer newly-learned skills to the operational setting. AISs can refer to the learning system used by the student or to aids that are solely used by the instructor, such as Instructor Operating Stations (IOSs) and Debrief / After Action Review (AAR) tools. Based on a review of learning focused technologies [4], AISs are comprised of multiple tools and methods used to optimize technology-delivered instruction. These include personal learning assistants, Intelligent Tutoring Systems (ITSs), adaptive learning systems, and brain-computer interfaces.

AISs are learning applications that are data-driven, with the particulars of the AIS learning domain continuously updated and shaped by the student's interactions with the AIS, to produce a learning experience that is maximally tailored to the individual student. AISs are domain agnostic and can be applied to a variety of training needs including those relevant to operators, maintainers and support personnel. Likewise, they can be adapted for instruction at the individual, team and/or collective training levels. These systems can be utilized in both academic settings (where the goal is acquisition of fundamental knowledge) and in more applied settings (in which the training is focused on the development of skills in the context of a specific mission or task). AISs are the product of multidisciplinary integrative efforts involving the learning sciences, artificial intelligence (AI), computer science, psychology, and pedagogy. For example, early AISs were based on computational representations of foundational cognitive/learning/memory principles discovered by research psychologists. However, it is important to note that although AISs are adept at teaching facts and training well-defined skill sets (i.e., declarative and procedural knowledge), they are not very useful for teaching more abstruse types of knowledge (e.g., tacit and implicit knowledge that is difficult to put into words [e.g., cultural mores]), and is generally obtained via first-hand experience.

Not only are AISs good for imparting declarative and procedural knowledge, but they frequently impart these types of knowledge in a blended manner – i.e., via a seamless integration of both knowledge and skill acquisition that is provided in a way that blurs the distinction between the acquisition of skills and the acquisition of fundamental knowledge.

The ultimate purpose of using AISs is to achieve the benefits that accrue from one-on-one human instruction compared to standard classroom instruction (i.e., a 2 sigma gain in learning outcomes, see Ref. [5]), albeit without the requirement of having a human instructor. In the simplest implementation, a pre-test is administered to the student to determine the baseline level of proficiency on each topic, or sub-section, which constitutes the course curriculum. Based on this information, the AIS adjusts its emphasis to focus on the specific content that the student has not yet mastered. In other implementations, a knowledge pre-check is used to recommend student self-study topics prior to instructor-led learning activities [6]. More advanced AISs provide instructorless digital education (e.g., distance learning) or training systems (e.g., flight training) that individualize the learning experience by tailoring instructional feedback, sequencing of events, and/or difficulty levels. These systems are sometimes called "Adaptive Learning Systems," "Adaptive Training Systems," "Cognitive Tutors," or "Intelligent Tutoring Systems." For a detailed review of Intelligent Tutoring Systems see Sottilare et al. [7].

AISs for individual training are typically comprised of:

1) A student model that represents the student's knowledge, skills, abilities, etc.;
2) An expert model that represents expert competency of the subject and/or domain; and
3) An instructor model that represents the optimal learning intervention.

The Instructor Model uses discrepancies between the student's performance (Student Model) and the Expected Performance (Expert Model) to select an appropriate instructional intervention or sequencing of curriculum. (Note: For detailed description of these components see R. Nkambou et al. [8]). These models are integrated within an individualized Learning Management System (LMS) that can guide and manage the student's learning activities from that point forward.

Although extremely promising, the full potential of AISs has not yet been realized. For a variety of reasons discussed later in this chapter, current implementations are typically either micro-adaptations (i.e., tailored learning within a single learning session) or macro-adaptations (tailored of the sequencing of learning sessions) [6], [9]. Finally, there are also a range of adaptation types, varying by modality, type of information conveyed, and timing of feedback [10], and this list is rapidly increasing. However, to date, new AISs tend to be "black boxes" – lacking in the details needed for users to seamlessly adapt these programs for their own, specific needs.

3.2 REQUIREMENTS FOR IMPLEMENTING AIS

3.2.1 Human Performer

Two primary active user roles are central to the development and implementation of an AIS: the student and the instructor.

The student is any trainee using an AIS to achieve his or her learning objectives. Requirements for the students include adequate time for each student to prepare for and participate in the training activity and access to the equipment required for the learning activity per the Technology Requirements section. Connectivity to a network to access distance learning or other server based content is often required. Depending on the AIS implementation hardware (e.g., Augmented [AR] or Virtual Reality [VR]) training to orient the student to the learning environment may be required. In addition, depending on the design of the AIS, the student may be required to wear or setup behavioral and/or physiological monitoring devices.

The instructor is the focus of the student's attention and conducts the training. Depending on the type of AIS employed, the instructor role can range from minimal (e.g., involving only review / monitoring of learning session results) to maximally interactive (e.g., when adaptive IOS or debrief/AAR capabilities are employed). Given the currently low level of maturity of AIS (i.e., relative to its potential), development of an AIS still typically requires input from human factors specialists and/or learning/cognitive scientists, software engineers, and, of course, subject matter experts. At present, relatively few AISs have sophisticated authoring tools – a feature that facilitates tailoring by an Instructional Systems Designer (ISD) or Instructor to support a unique learning activity.

3.2.2 Technology and Hardware Requirements

Given the wide range of potential uses of AISs, the exact technology configuration for a given AIS is dependent on the domain / context of the intended application of instruction, learning goals or objectives for the educational or training activity, location / setting of the instructional activity (e.g., schoolhouse, home, field) and the student's learning needs (e.g., the purpose or intent of the adaptation).

All distance learning applications – including those that impart declarative or procedural knowledge, as well as those that are training-focused (i.e., involving acquisition of skills) – can utilize a variety of technologies to complete lessons. The technological requirements will depend upon the learning domain, the specific task being taught, and the target level of expertise to be achieved. Technology can be as simple as a PC/laptop interface to a gaming environment using a keyboard and mouse – or even a joystick. For blended and training-focused learning activities, simulators of varying levels of fidelity, depending on the learning activity, may be employed. Immersive reality (e.g., VR and AR) is used to provide realistic domain environments via headsets/goggles, haptic interaction devices, specialized chairs and other hardware to simulate real world environments [11].

3.2.3 Data Requirements

The primary data requirement for AISs is access to outcomes related to the learning activity and information that is used for performance assessment and implementation of adaptations for the individual student, such as responses or other behaviors exhibited during a lesson. Distance learning / educational AISs data requirements may be as simple as a test score or as complex as behavioral data (e.g., keyboard strokes, websites accessed) or even a priori responses to personality, motivation, self-efficacy or learning style questionnaires. Part-task and full-task simulator training sessions may vary from simple and exiguous (e.g., an instructor score for an entire session) to complex and detailed (e.g., simulator data that includes multiple measures of the student's performance like throttle use, speed, banking angle, etc.). Several systems incorporate wearable devices to monitor student physiological and/or behavioral responses [12]. For some types of training, mobile systems are sometimes deployed. The Generalized Intelligent Framework for Tutoring (GIFT) was applied to the development of an AIS deployed on a dismounted infantry's mobile phone to provide live, adaptive training for land navigation [13].

3.3 STRENGTHS, WEAKNESSES, OPPORTUNITIES, THREATS (SWOT) ANALYSIS

Table 3-1: Adaptive Instruction and Accelerated Readiness SWOT Analysis.

Adaptive Instruction and Accelerated Readiness	
Strengths	**Opportunities**
• Optimized learning time and resources • Enhanced learning compared to non-AIS methods • Established techniques for implementation • Flexibility: can be tailored to all kinds of training and education contexts • Especially well-suited to well-defined knowledge domains	• Emerging applications for instructor-in-the-loop training and collaborative learning • Governments (e.g., US DoD) investing in adaptive learning ecosystems • Resilience to factors that mitigate against in-person training (e.g., COVID) • Provide tailored training at point of need • Enhanced connectivity (e.g., 5G) and distributed computation • Enhancements immersive environment technologies to increase fidelity of training

Adaptive Instruction and Accelerated Readiness	
Weaknesses	**Threats**
• Overall low maturity, with various instructional models / adaptation types • Investment in institutional resources still considerable due to ecosystem required • Content authoring and system management require specialized skills and knowledge • Authoring content is labor intensive • Current state of Natural Language Processing (NLP) does not yet support open dialogue, varying accents • Challenges in applying to abstruse (e.g., tacit and implicit) knowledge domains • Requires investment in instrumented training environment • Limited by limitations of synthetic environments	• Data/cyber security • Pace of technical evolution/obsolescence and operational concepts outstrip speed of development of instructional systems • High upfront costs can weigh heavily in a 'return on investment' analysis • Cannot seamlessly adapt to changes in operational and training concepts and technology • Cognitive "crutch" potential in the case of on-the-job support • Organizational culture and user-buy in Privacy and GDPR issues with user data

3.3.1 Strengths

The increased focus and adoption of AISs in military training is due to potential reduction in training time and costs, superior learning outcomes, applicability to a wide range of knowledge domains, flexibility, and compatibility with technologies ranging from basic online learning systems to high-fidelity simulations. AISs are best suited to well-defined domains in which the expert / domain model and learning tasks follow a logical flow, with unambiguous definitions of correct and incorrect performance (i.e., success vs. failure), that occur in the context of predictable events and environments.

3.3.1.1 Optimized Training Time and Learning Outcomes

The primary benefits of AISs are the reduction in training time required to attain mastery (e.g., time to train) and superior learning outcomes. Van Buskirk et al. [14] reported a 46% reduction in missed reports and 49% improvement in accuracy during submarine electronic warfare training when the Submarine Electronic Warfare Adaptive Trainer (SEW-AT) was used to train contact classification tasks. When embedded non-player characters that provided immediate feedback were integrated into SEW-AT, report timeliness was further improved [15]. Likewise, Craven [6] reported that retention of aviator ground school material was significantly greater for those trained with adaptive self-directed study prior to instructor-led training than compared to students who had completed the traditional instructor-led training with the students responsible for self-study (i.e., loss of 14.5% vs. 27.3% of the material, respectively).

3.3.1.2 Established Implementation Techniques

Carnegie-Mellon has implemented an open-source tool for creating cognitive tutors called Cognitive Tutor Authoring Tools (CTAT [16]). This tool allows educators with limited programming experience to build cognitive tutors for teaching simple and complex problem-solving strategies to students based on the Adaptive Control of Thought – Rationale cognitive modelling architecture, better known as ACT-R. Access to the tool is available on the CTAT's website (see https://www.cmu.edu/simon/open-simon/toolkit/tools/learning-tools/ctat.html).

The Generalized Intelligent Framework for Tutoring (GIFT) was implemented by the US Army to provide government, industry and academia a common platform for developing adaptive computer-based training capabilities based on Merrill's Multiple Resource Theory. GIFT provides a set of tools for specifying performance measures, including physiological and behavioral measures via the Sensor Module, Learner Module, Pedagogical Module, and Domain module [17].

The Advanced Distributed Learning (ADL) Total Learning Architecture (TLA) is an active R&D project with the goal of developing technical specifications, standards and policy guidance for adaptive learning ecosystems [18].

3.3.1.3 Domain Agnostic

AISs provide a methodology and capability for providing training tailored to the students' needs. They are flexible and can be tailored to all kinds of training and education contexts using a variety of training mediums (e.g., distance learning, simulation based training). Adaptive learning has been used in educational military settings, such as the Air Education and Training Command's Basic Military Training curriculum [19] and the US Navy's My Navy Learning adaptive learning platform to provide content for personalized distance learning. AISs have also been used with a variety of training domains, including: Undergraduate Pilot Training under US Air Force Air Education and Training Command's (AETC) Pilot Training NEXT program [6], [11], US Air Force maintenance training [20] and Navy submarine training applications [14].

3.3.2 Opportunities

3.3.2.1 Nascent Applications to Instructor-in-the-Loop Training and Collaborative Learning

Although AIS development efforts have primarily been focused on instructorless learning, this technology can also be applied to adjunctively support instructor-led training. Such systems typically include real-time performance assessment indicators that are accessed at dedicated Instructor Operator Stations [21] and Debrief tools.

Collaborative learning – sometimes referred to as peer-to-peer learning or social learning – is known to be highly effective. Accordingly, the Flow Driven Experiential Learning (FLXD) instructional methodology developed in support of the US Air Force recommends that peer learning be incorporated into the design of AISs [20].

3.3.2.2 Learning Ecosystems Investments

AISs require a large ecosystem to support implementation and deployment. They also require institutional flexibility – e.g., a willingness to modify long-held policies and practices to reflect, and take full advantage of, the new capabilities provided by AISs. (For example, AIS may prompt reconsideration of 'fixed duration' training of pilots.) It is a positive sign that attention and resources are increasingly be applied to development of AIS learning ecosystems.

A focus of the ADL Initiative is the implementation of the TLA to operationalize the design, development, deployment, and maintenance of distributed, personalized learning [18]. To reduce training time and the cost of training their 174,000 acquisition professionals, and to enable these professionals to start their careers with a greater knowledge base, the Defense Acquisition University may shift to an adaptive learning environment that responds in real time to student needs [22]. AIS capabilities that are being developed in support of this possibility include analytics to predict 'time to mastery' and 'decay rates' -outcome variables that will help address challenges associated with personnel selection and assignment [23].

3.3.2.3 Distributed Implementation Potential

AISs are implemented via distributed learning or remote learning mediums, rendering the physical location of the student immaterial. In 2020, during the COVID-19 pandemic, it became apparent that remote learning was more than a convenience – for millions of school children it was the only way for them to safely continue their studies. Likewise, remote learning makes possible continued learning by warfighters regardless of assignment or location – even during deployment.

3.3.2.4 Adaptive Performance Aiding

As AISs mature, the possibilities for applications of these technologies expands. For example, many of the AIS advancements in human-machine interaction for instructional purposes can (and most likely will) be applied to human-machine teaming efforts in both civilian and military operational environments. There are a wide variety of potential applications that could benefit from AIS-derived tools for assessing human performance in real time, identifying deficiencies at the individual operator level (e.g., due to the accrual of fatigue effects), and facilitating delivery of (or actually providing) the correct type and amount of support needed to the individual operator to sustain performance efficacy.

3.3.2.5 Enhanced Connectivity

Enhanced connectivity and the increased availability of 5G networks, which facilitate the ability of AISs to access increased processing power, is making the deployment of AISs for geographically distributed warfighters a reality distributed computing, further expands the ability of AISs to swiftly process data from multiple sources and maximize the fidelity with which the learner's needs are assessed and met.

3.3.2.6 Continued Advancement in Synthetic Environments

Every advancement of Synthetic Environment (SE) technologies, especially AR and VR, constitutes a potential advancement for AISs – especially for those AISs that require, or are enhanced by, hi-fidelity simulations. Immersive reality expands the potential to create interactive learning environments that can provide relevance to the operational environment to promote transfer as well as enable interactive instructional methods to reinforce instruction [11]. The access to human performance data enabled by these systems (many of which are increasingly instrumented with hardware used in actual operations) will facilitate efforts to identify meaningful outcome measures during live training, and potentiate acceptance of, and the integration of, machine-human teaming technologies in actual operations.

3.3.3 Weaknesses

3.3.3.1 Low-to-Moderate Maturity

While the use of AISs is increasing rapidly, this technology is still at a low-to-moderate level of maturity, with little standardization of instructional models and adaptation processes. Multi-modal feedback and adaptation techniques continue to evolve along with complementary technologies that are especially important for AISs that involve immersion in simulated environments. It is difficult to conduct head-to-head comparisons of AISs because of the wide variety of instructional theories and implementation methods that they employ. Consequently, there is danger that this area will be increasingly characterized by "stove piping" with inefficient, parallel development of application-specific methods and approaches. While efforts to standardize some aspects of AIS do exist (e.g., the US Army's Generalized Intelligent Framework for Tutoring [GIFT]) [17], the ADL Initiative's Total Learning Architecture (TLA) [18], and the IEEE AIS standards development effort [24], AIS design and instructional methods are still largely researched and developed independently.

While the services are supporting some pilot studies to assess the utility of adaptive learning, the number of programs of record that are actively and comprehensively assessing adaptive learning remain limited.

3.3.3.2 Ecosystem Investment Requirements

While initial investments of institutional resources to implement AISs are increasing, considerably more is needed to provide the infrastructure and personnel needed for full implementation of AISs. Technology requirements include instrumentation to collect, protect and distribute learner data. Personnel issues are potentially more complicated because the time required for training, or retraining, of personnel during the transition to a new or modified AIS; and subsequently, the time and effort required to manage and periodically maintain and update the AIS will likely vary across institutions and applications.

3.3.3.3 Labor Intensive Development

New content for authoring applications is continually being developed. However, the design and development of AISs requires quality data (the maxim “garbage in, garbage out” applies to these efforts), as well as considerable time and resources to collect and analyze those data. Even for those AISs that include authoring tools that allow the instructor to tailor content, there is often a need for software engineers to tweak the system and/or create applications to support performance data requirements or instructional intervention implementations. These issues exist even for well-defined knowledge domains. For less well-defined knowledge domains, the challenges associated with implementing an effective AIS can increase exponentially.

3.3.3.4 Maturity Level of Natural Language Processing

The current state of Natural Language Processing (NLP) does not yet support open dialogue with students. Often interactions between human and machine are constrained by both a limited lexicon, and a limited mode of interaction (e.g., a keyboard and computer screen). Even when the lexicon is severely restricted, verbal human-machine interactions can be problematic because the technology does not typically support a range of accents.

3.3.4 Threats

3.3.4.1 Data/Cyber Security

Multiple types of data, including, but not limited to, lessons learned from prior experience, personality test scores, motivation level assessments, and physiological measures, can be useful for assessing the learning needs of individual students. However, the more personal the data collected, the greater the potential harm associated with a breach of system security. Regulations to protect individual data privacy rights are increasing, a trend that may limit future opportunities for AISs to collect and utilize such data. Cyber threats to data protection are a growing reality that also make the use of AISs operating with sensitive data a challenge.

3.3.4.2 Relatively Fast Pace of Technology Evolution

The pace of technical evolution and obsolescence, and associated changes to operational concepts is outstripping the speed of development of instructional systems. Adaptive instruction systems are not yet agile enough to keep up with technical, operational, and training advancements. In other words, the possibility exists that for some rapidly advancing and changing knowledge areas, a lag in updating the AIS could result in the student being provided with outdated (and thus incorrect) information.

3.3.4.3 High Initial Investment

The high upfront investments required for AIS implementation can make the ROI proposition unattractive, especially when leadership rotation cycles outpace potential benefits to be gained from the results produced by AIS. Upfront investments include the time to design, develop and implement the AIS for a specific lesson or course, which can be lengthy and expensive due to lack of automated processes for developing the expert and domain models.

3.3.4.4 Skill Erosion

Similar to the advent of other support technologies such as the GPS, point of need AIS has the potential to serve as a cognitive crutch and lead to skill erosion. If students depend too much on the AIS to support their needs, there is the potential that the student will not invest the time and energy into the learning required to sustain readiness, if, for example, the capability being trained is translated into a machine-human teaming system. From this consideration, it will be critical to link to Research Task Group outcomes for HFM-292 titled "Understanding and Reducing Skill Decay", with an emphasis on establishing best practices and ethical guidelines for embedding augmentation technologies across a human performance aligned operational timeline.

3.3.4.5 User Adoption

Organizational culture and user-buy in is a challenge. With the advent of any new technology, there is always the possibility that those affected by that new technology will resist implementing it. With AIS, there is a logical but ultimately incorrect concern that human instructors will be replaced – although the reality is that this technology will aid them and allow them the time to provide more focused instruction. Also, it appears that instructors will always be needed to impart the complex types of knowledge (e.g., tacit and implicit lessons learned from prior experience) that cannot be effectively imparted by an AIS. Given the current low-moderate maturity of AISs, a bad experience within AIS (e.g., due to inappropriate diagnosis of learning needs and/or provision of inappropriate feedback), could result in the stigmatization of the technology and a delay of its implementation.

3.4 CONCLUSION

AISs have the potential to augment human performance by providing training tailored to the individual's learning needs, thereby improving learning outcomes and retention rates. Additionally, in some cases, adaptive instruction can shorten the time to mastery. As a result, the use of AISs can enhance or improve warfighter readiness. Since AISs can be deployed anytime, anywhere, targeted learning can be provided at the point of need, thus enhancing human performance.

3.5 REFERENCES

[1] Foster, R.E., and Fletcher, D.J. (2002). "Computer-Based Aids for Learning, Job Performance, and Decision Making in Military Applications: Emergent Technology and Challenges." Presented at the RTO HFM Symposium on The Role of Humans in Intelligent and Automated Systems, Warsaw, Poland, 7 – 9 October 2002. NATO STO Meeting Proceedings Paper RTO-MP-88-06.

[2] Perrin, B.M., Buck, B.J., and Gehr, S.E. (2010). "Intelligent Tutoring: Bridging the Gap from Knowing to Doing." Paper No. 10040 Presented at the 2010 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC, Orlando, FL).

[3] Brawner, K., and Sottilare, R. (June 2018). "Proposing Module Level Interoperability for Intelligent Tutoring Systems." In the Exploring Opportunities to Standardize Adaptive Instructional Systems (AIS) Workshop of the 19th International Conference of the Artificial Intelligence in Education (AIED) Conference, London, England, United Kingdom, June 2018.

[4] Richins, Caldeira-Hankey, A., Deighton, C., Mundy, D., Grimshaw, T., Newell, K., and Snook, A. (2021). "Personalisation of Learning: Developing the Case for Implementation within Defence Learning Establishments." In 2021 Interservice/Industry, Training Simulation and Education Conference Proceedings, Nov 29 – Dec 2. https://www.xcdsystem.com/iitsec/proceedings/

[5] Bloom, B.S. (1984). “The Two Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring.” Educational Researcher, 13(4-6).

[6] Craven, P.L. (2021). “Blended and Adaptive Learning in Ground School Instruction for Aviators.” In 2021 Interservice/Industry, Training Simulation and Education Conference Proceedings, Nov 29 – Dec 2. https://www.xcdsystem.com/iitsec/proceedings/

[7] Sottilare, R.A., Fletcher, J.D., Skinner, A., Perez, R., Durlach, P., Zotov, V., and Alexander, T. (2018). “Assessment of Intelligent Tutoring System Technologies and Opportunities.” NATO STO Technical Report STO-TR-HFM-237), NATO Science & Technology Organization, Neuilly-sur-Seine, France. DOI: 10.14339/STO-TR-HFM-237. https://www.gifttutoring.org/attachments/download/2520/$$TR-HFM-237-ALL.pdf

[8] Nkambou, R., Bourdeau, J., and Mizoguchi, R. (Eds.). Advances in Intelligent Tutoring Systems, SCI 308, pp. 33-62. Berlin/Heidelberg, Springer-Verlag.

[9] VanLehn, K. (2006). “The Behavior of Tutoring Systems.” International Journal of Artificial Intelligence in Education, 16(3), pp. 227-265.

[10] Johnson, J., Smith, K., Diaz, R., Dennis, T. and Jimenez, G. (2021). “Adaptive Assessment Feedback in Competency Based Learning Ecosystems.” In 2021 Interservice/Industry, Training Simulation and Education Conference Proceedings, Nov 29 – Dec 2. https://www.xcdsystem.com/iitsec/proceedings/

[11] Sheets, T.H., and Elmore, M.P. (2018). “Abstract to Action: Targeted Learning System Theory Applied to Adaptive Flight Training.” Research Report. Maxwell Air Force Base, Alabama. https://apps.dtic.mil/sti/pdfs/AD1053015.pdf (Accessed 8 February 8, 2022)

[12] Finseth, T., Dorneich, M., Keren, N., Franke, W. and Vardeman, S. (2021). “Training for Stressful Operations Using Adaptive Systems: Conceptual Approaches and Applications.” In 2021 Interservice/Industry, Training Simulation and Education Conference Proceedings , Nov 29 – Dec 2. https://www.xcdsystem.com/iitsec/proceedings/

[13] Davis, F., Riley, J., and Goldberg, B. (2019). “Extending GIFT Wrap to Live Training.” In Proceedings of the 7th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym7), Orlando, FL. https://gifttutoring.org/attachments/download/3271/2019_05_GIFTsym7_proceedings.pdf

[14] Van Buskirk, W.L., Fraulini, N.W., Schroeder, B.L., Johnson C.I., and Marraffino, M.D. (2019) “Application of Theory to the Development of an Adaptive Training System for a Submarine Electronic Warfare Task.” In R. Sottilare and J. Schwarz (Eds.), Adaptive Instructional Systems, First International Conference, AIS 2019, held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26 – 31, 2019. Lecture Notes in Computer Science, 11597. Springer, Cham. DOI: /10.1007/978-3-030-22341-0_28.

[15] Schroeder, B.L., Fraulini, N.W., Van Buskirk, W.L., and Johnson, C.L. (2020). Using a Non-Player Character to Improve Training Outcomes for Submarine Electronic Warfare Operators. Adaptive Instructional Systems, Second International Conference, AIS 2020, held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19 – 24, 2020, pp. 531-542. DOI: /10.1007/978-3-030-50788-6_39.

[16] Carnegie-Mellon University (n.d.). Cognitive Authoring Tools (CTAT) Website. https://www.cmu.edu/simon/open-simon/toolkit/tools/learning-tools/ctat.html

[17] Sottilare, R.A., Goldberg, B.S., Brawner, K.W., and Holden, H.K. (2012). "A Modular Framework to Support the Authoring and Assessment of Adaptive Computer-Based Tutoring Systems (CBTS)." In 2021 Interservice/Industry, Training Simulation and Education Conference Proceedings, Nov 29 – Dec 2. https://www.xcdsystem.com/iitsec/proceedings/

[18] Smith, B., Johnson, A., Hayden, T., and Tolk, F. (2021). TLA Functional Requirements Document. Advanced Distributed Learning Initiative: Orlando. https://adlnet.gov/projects/tla/

[19] Hawkins, D. (14 January 2020). "Air Force Beta-Tests Adaptive Learning Platform in Basic Military Training." Air Force News Service." Joint Base San Antonio-Lackland, Texas: Air Education and Training Command Public Affairs.

[20] Taylor, N.E., and Clayton, A.S. (2021). "Form from Function: Applying Flow Driven Experiential Learning to The Integration of Immersive Technology in Formal Military Aviation Training Programs." In 2021 Interservice/Industry, Training Simulation and Education Conference Proceedings, Nov 29 – Dec 2. https://www.xcdsystem.com/iitsec/proceedings/

[21] Champney, R., Milham, L., Parrish, T., Kokini, C. and Stanney, K. (2010). "Automating Instruction Support: Insights from Field Research." SpringSim '10: Proceedings of the 2010 Spring Simulation Multiconference, April 2010, Article 31, pp. 1-6. DOI:10.1145/1878537.1878570.

[22] Miller, S. (30 January 2020). "The Defense Acquisition University and the Air Force Are Experimenting with New Ways to Speed Up and Improve Learning." Defense One.

[23] Jastrzembski, T. (2020). "Predictive Analytics and Personalized Learning." Presentation at Air Force Agency for Modeling and Simulation and the 4th Training Squadron, 2020 Training Innovation Workshop.

[24] Sottilare, R., Barr, A., Graesser, A., Hu, X., and Robson, R. (Eds.), (March 2018). Proceedings of the First Adaptive Instructional Systems Standards Workshop. US Army Research Laboratory, Orlando, Florida. ISBN: 978-0-9977257-3-5.

Chapter 4 – MISSION PREPARATION AND REHEARSAL

Glenn Gunzelmann and Benjamin Goldberg

4.1 INTRODUCTION

Military operations are complex, involving numerous weapons systems and a multitude of individuals to accomplish complicated objectives in a decidedly adversarial environment. To achieve mission objectives requires careful coordination and planning to build shared understanding of priorities and contingencies, to identify risks to mission success, and to work through the complicated logistics of achieving integrated effects across time and space. Mission Planning is the process that militaries use to translate a set of objectives into specific actions to achieve them. Just as sports teams plan and rehearse specific strategies, plays, and scenarios to prepare for specific opponents, the military engages in targeted planning activities to prepare for specific operations.

Historically, mission planning was an activity that unfolded on white boards and using physical models of the operational environment at various levels of fidelity. However, with the increased emphasis on joint all-domain operations and the startling degree of complexity involved in planning how to bring integrated effects to bear on a peer adversary in a contested environment, these traditional tools are no longer adequate for the planning and rehearsal challenges faced by today's militaries. Mission planning must bring together the individuals who will execute the mission to identify mission priorities, consider threats and alternative scenarios, and establish 'contracts' that represent a shared understanding of how individuals and teams will react to deviations to the plan to maximize effectiveness.

This complexity increases the need for technological sophistication in planning. In future operations, large collections of highly varied weapons systems and cross-service and coalition warfighters will need to coordinate to achieve specific mission objectives over long periods of time in a warfighting theatre that may span vast geographic areas. At the same time, the plan must be flexible enough to achieve the objectives within a Volatile, Uncertain, Complex, and Adversarial (VUCA) operational environment. Finally, each plan is unique. Although there are opportunities to inform current mission planning by leveraging After Action Review (AAR; see Chapter 7), identical situations never exist in military operations.

Mission planning activities are contextualized in terms of a specific set of objectives in a specific environment. Missions are established through formal taskings that are derived from operational plans, campaign objectives, threat assessments, and resource availability. The mission planning challenge is to define how to use the assigned resources to achieve a prioritized list of mission objectives. This is basically a constraint satisfaction problem but is significantly more complex than just assigning assets to targets. Considerations of fuel, range, payload, objective priority, and other factors play heavily into decision making.

Because operations are dynamic and evolving the process must also consider a range of contingencies that influence the overall plan. For instance:

- What if a critical enemy defence is not eliminated as planned?
- What if there are additional defences that were not identified?
- What if the locations of enemy assets differ from expectations?
- What if specific assets are eliminated before completing their missions?

These considerations lead to discussions about how to adapt to dynamics changes, which serves to set overall expectations, or 'contracts,' that define how to react in various potential circumstances. Importantly, however, those discussions also create more general expectations about the expertise and tendencies of the operators in the mission. This can aid in understanding and predicting reactions for deviations during the operation that weren't explicitly planned for.

4.2 HUMAN PERFORMER WITHIN THE MISSION PREP AND REHEARSAL DOMAIN

Establishing an agreed upon general definition of the human performer in the mission preparation and rehearsal context is challenging. This is a critical set of activities and procedures that are executed across multiple levels within an organizational hierarchy, as well as across different personnel, task, and operational environment constraints. Regardless, there are some common characteristics that can be used to establish human performer roles and interrelated performance attributes that impact task assignments and operational orders that are output at the end of this tactical timeline phase. These performance attributes will also drive technology development with a focus on augmenting the task interaction space for the purpose of mitigating or enhancing performance effects.

Mission planning is executed within a team of interconnected roles. From an operational perspective, these are the planning and preparation activities closest to a "bang" event when considering the tactical timeline. The context is better established and linked to multiple forms of information/intelligence that drive an overarching mission plan used to initiate rehearsal and course of action analyses. However, timelines are constrained and the primary performance moderators are cognitive in nature; in other words, mission preparation is context driven synthesis of near-time and real-time information into actionable intelligence that leads to a tactical plan of action centered around a notional "bang" event.

Interestingly, as the operational environment and the tasks performed within become more complex and technologically dependent, processes and procedures within a mission prep and rehearsal workflow remain relatively static. Despite the profoundly different character of current missions, their procedures, and the used technology they use have not changed with the same speed. Even more complex and dynamic military missions are planned and conducted with old procedures and slightly improved technology [1] At the Operational level, tasks generally include planning, and the processes of acquiring, analyzing and interpreting information in order to allocate available resources. Advances in AI-assisted analysis, and machine and deep learning are leading to the creation of completely new capabilities [2]. As an example, capabilities leveraging advancements in AI and computer vision would allow automated feature extraction and classification from imagery captured remotely by unmanned airborne systems. As a result, the performance requirements of those engaged in surveillance and reconnaissance and using operational analysis would therefore be shifted further toward the cognitive [3].

4.3 AUGMENTATION TECHNOLOGIES FOR MISSION PLANNING

Given the complexity of the mission planning activity and the breadth of augmentation technologies that have emerged and matured in recent years, it is not surprising that there are many opportunities to leverage technology to improve the efficiency and effectiveness of mission planning for military operations.

4.3.1 Virtual Presence

Beginning in the 2000s, but certainly accelerated by the COVID-19 pandemic, technologies for virtual presence and interaction have matured to the point where they provide a reliable capability for personal interaction at great distances. The military has long leveraged networking for distributed mission training through large-force virtual exercises, even if numerous challenges remain in establishing a seamless and persistent capability. However, applying those same technologies for mission planning has lagged. Given the increasing scale of military operations, and the diversity in assets and personnel that may be involved in them, it is unlikely that traditional "white board" approaches to mission planning will remain viable. Simply gathering all of the critical personnel in a single location to engage in the process is likely to be an insurmountable challenge. As a result, taking advantage of virtual communication technologies and data-sharing capabilities will be essential to planning and executing JAD missions. A number of new capabilities in this technology space are emerging and maturing, including improvements to shared "white board" spaces and immersive technologies that enable shared perspective taking, gesturing, and other non-verbal communication strategies that are common in mission planning activities. Other opportunities afforded by these technologies are discussed next.

4.3.2 Immersive Technologies

Augmented, Virtual, and Mixed Reality (AR/VR/MR) technologies are a driver for this group's work to consider how augmentation technologies may be leveraged in military contexts. Mission planning is an obvious application alternative. As noted, it can be used in conjunction with virtual interaction technology to bring geographically distributed individuals together to engage in large-scale mission planning activities using mostly normal communication strategies [4]. In addition, however, these technologies can allow leaders and operators to build a more comprehensive understanding of the operational environments, explore implications of tactical decisions, and better evaluate alternative courses of action during the planning process. Using these technologies to adopt a first-person perspective within the engagement can provide insight that would be difficult to derive from 2-dimensional maps. Simulating portions of the mission to better visualize how the operation unfolds under various contingencies can help with risk reduction and improve mission effectiveness. In addition, immersive technologies allow an opportunity to visualize abstract features of operations like weapons engagement zones, surface-to-air missile site ranges, natural camouflage, choke points, etc. Although these features may be surmised in many cases using traditional planning materials, it will also generally require greater cognitive effort to keep everything in mind simultaneously. As a result, immersive technologies can free up cognitive bandwidth to reason through engagement strategies and Concepts of Operations (CONOPS), which will ultimately benefit mission effectiveness.

Importantly, immersive technologies can be used in conjunction with other capabilities to create environments that are tailored to specific needs. For instance, augmented reality 'sand tables' that include elevation deformation that can be manipulated to match specific terrain, can instantiate a realistic 3-dimensional model of an operational area, while also allowing operational systems, personnel, and adversary capabilities to be incorporated into the view. Rich immersion of this sort can facilitate understanding line of sight, natural cover, and likely navigation paths that can provide a deeper understanding of the implications of various tactical decisions than would otherwise be possible. These hybrid simulation environments can also serve as bridge between more traditional mission planning environments and entirely virtual representations.

4.3.3 Artificial Intelligence

Like immersive and other augmentation technologies, Artificial Intelligence (AI) has broad implications for military operations. In the mission planning space, AI can improve the efficiency of planning by performing some tasks in seconds that currently take hours for humans to perform. This includes many aspects of what is referred to as 'admin mission planning.' These activities are crucial to the successful execution of the mission, such as logistics to ensure all of the required assets are available and in position to initiate the mission. The complexity in many of these activities derives from the intricate constraint satisfaction challenges that emerge when so many systems need to be coordinated, rather than from the uncertainty and ambiguity of the environment. By performing these roles, however, bandwidth is created for human operators to grapple with those tactical decisions and to plan for a larger number of potential contingencies.

In addition to supporting some of the mundane – if also critical – planning activities, AI can also be used to evaluate alternative courses of action to support tactical decision making. By creating high-fidelity simulation environments and populating them with cognitively realistic synthetic operators, risks can be identified and mitigated through more detailed planning than humans operating in isolation could achieve. This teaming of human operators with AI-based machine systems allows the planning process to leverage the deep expertise and creativity of the people while exploiting the processing efficiency and scalability of artificial systems to create mission plans that are more likely to be successful.

4.4 STRENGTHS, WEAKNESSES, OPPORTUNITIES, THREATS (SWOT) ANALYSIS

Table 4-1: Mission Preparation and Rehearsal SWOT Analysis.

Mission Preparation and Rehearsal	
Strengths • AR/VR offer potentially greater immersion and shared perspective • Aspects of the planning process can be automated – reduce planning time/cycle • Reduced errors by maintaining information in a digital form throughout • Deeper immersion to better understand implications of plans and options • Increased opportunity to consider alternative strategies and tactics • Improvements to shared awareness and trust • Multiple industry solutions in support of digital battlespace visualization	**Opportunities** • Distributed and collaborative planning opportunities • Enhanced team SA • Enhanced common operating picture through multi-modal interfacing (e.g., 2D satellite imagery, 3D VR, AR layered on satellite, etc.) • Layer multi/joint domains and effects on operational space • Leverage AI for simulated Course of Action analysis • Longer-term opportunity to augment mission planning with AI-based reasoning to further improve strategy and tactics • Leverage synthetic training resources to manage iterative planning and rehearsal
Weaknesses • No defined concept of operation for AR VR in mission planning • Interoperability issues and data standards • Exposure time in visual space has cognitive effects (cyber- or simulator sickness) • Physical constraints for hardware • Most done on white boards and ppt slides… rocks and sticks on the floor • Multiple, time-intensive, workflows • Inefficient • Data entry by people (error prone) • Doesn’t leverage modern tech	**Threats** • Culture to avoid risk (reluctance) • Network and power dependencies • Security requirements; introduces cyber risks • Zero tolerance for technology issues (i.e., propensity to revert back to ways of the old) • Over reliance on AI

4.5 REFERENCES

[1] Knobeloch, B. (2020). Military Mission Planning Using Augmented Reality Technology. Doctoral dissertation. Naval Postgraduate School, Monterey, CA.

[2] Billing, D.C., Fordy, G.R., Friedl, K.E., and Hasselstrøm, H. (2021). "The Implications of Emerging Technology on Military Human Performance Research Priorities." Journal of Science and Medicine in Sport, 24(10), pp. 947-953.

[3] Kott, A., Alberts, D., Zalman, A., Shakarian, P., Maymi, F., Wang, C., and Qu, G. (2015). "Visualizing the Tactical Ground Battlefield in the Year 2050: Workshop Report." Army Research Lab, Adelphi, MD. Computational and Information Sciences Directorate.

[4] Amburn, C.R., Vey, N.L., Boyce, M.W., and Mize, J.R. (2015). "The Augmented Reality Sandtable (ARES)." Army Research Lab, Aberdeen Proving Grounds, MD. Human Research and Engineering Directorate.

Chapter 5 – REAL-TIME SUPPORT AND REMOTE CONTROL

Peder Sjölund and Mayowa Olonilua

5.1 INTRODUCTION

Modern organizations cannot resort to the decision-making process without relying on information and communication technologies available. Besides information as an important input of this process, the tools and techniques used by decision-makers and military personnel are equally important in the support and validation of their decisions and aid on daily work. Military commanders face some of the most difficult and high-stake decision issues meaningful not only at the level of the military, but also for the humankind. Under these circumstances and as a result of an increase in the diversity and complexity of conflict situations, means that there is a need to support military decision making and operations by providing commanders with the tools to gather information and operate in complex environments. This could be enabled by remote systems that provide real-time support and remote-control systems.

As well as providing military commanders with enhanced situational awareness and decision-making capabilities, these technologies provide numerous human benefits for the operators on the ground. For example, real-time support technologies can provide just in time training [1] which would reduce training required earlier in the training pipeline. Additionally, remote-control systems can enhance the physical and cognitive capabilities of the human operator for example allowing the operator to move heavy objects or to comprehend more information than they would be able to normally due to the extra information overlaid on to their visual field.

5.1.1 Scope

A real-time system is a system of software that is designed to carry out different tasks simultaneously with real-time output. The execution of tasks is implemented in a fixed time basis without any delay. Even if the system undergoes the same failure in different occasions, there will be no difference in the results. The scope of real-time support and remote-control technologies is delimited to only cover operational aspects “during bang” in the domains in support of human performance with a timing window between end points occurring within seconds down to a few milliseconds.

Being able to perform real-time support and remote control involves both big data transfer and low latency between communicating endpoints. This demands sophisticated and redundant data network architectures and network services to be realized.

The scope of remote-control technologies will only cover operational use cases during bang and due to the wide range of potential remote-control use cases within the military, the time latency can be as low as milliseconds for systems that require rapid response times for real-time feedback such as remote surgery or in seconds for a use case where slightly longer response times are acceptable such as operating a satellite.

Example of a real-time application “prior bang” in the domains in support of human performance are proactive functions such as adaptive instructions and mission preparation and rehearsal while after e.g., action review is a reactive function. These aspects will not be covered in this section.

5.2 DIGITAL BACKBONE OF DATA NETWORK INFRASTRUCTURE

A data network infrastructure with certain capabilities is key to realize different applications augmenting end-user functions based on real-time support and remote control. Different types of data networks offer different capabilities in data transport in terms of both speed and bandwidth. Modern mobile cellular technology such as the Long-Term Evolution (LTE/4G) and the fifth generation 5G networks also offers Internet Protocol (IP) control of priority, quality of service, multi-cast streaming, etc. and to store and process large amount of data in real time. The type of data network used determines to a large extent the type of application that can be built and deployed on top of the network. Traditional combat network radio solutions are in comparison to 5GNR (5G New Radio) offering less data put-through. Tactical 5G based mission-critical networks can provide high-capacity bandwidth in flexible end-to-end solutions scalable both in terms of user numbers and system range with multiple frequency bands available. This is broadband networks that can deliver high throughput of data and enable real-time video streaming and push-to-talk functionality and are based on the open mobile broadband standard 3GPP (www.3gpp.org) which covers cellular telecommunications technologies, including radio access, core network and service capabilities, which provide a complete system description for mobile telecommunications. The 3GPP standards and their global scale are quintessential to LTE and 5G in providing cost-effective communications for both commercial and mission-critical users. The continuous development of the 3GPP standard, and its gradual evolution to new generations technologies, will ensure that these standards remain the best option to meet the upcoming demands of critical communications.

Besides the data transport functionality, local and central compute is a crucial part of a modern network to realize new end-user services. In combination with functionalities such as edge infrastructure, edge-user plane, edge-routing and orchestration, and edge exposure helps the network owner to add value beyond connectivity. Exposure through Application Programming Interfaces (APIs) on the edge is getting increasingly important for network owners to enable new services, increase their relevance in the 5G ecosystem and become more attractive for both private and hyperscale cloud providers, application ecosystems, public safety, military and other players. The main benefits edge compute solutions provide include low latency, high bandwidth and device processing and will unlock many real-time and remote-control use cases for future mission-critical deployment (see Figure 5-1).

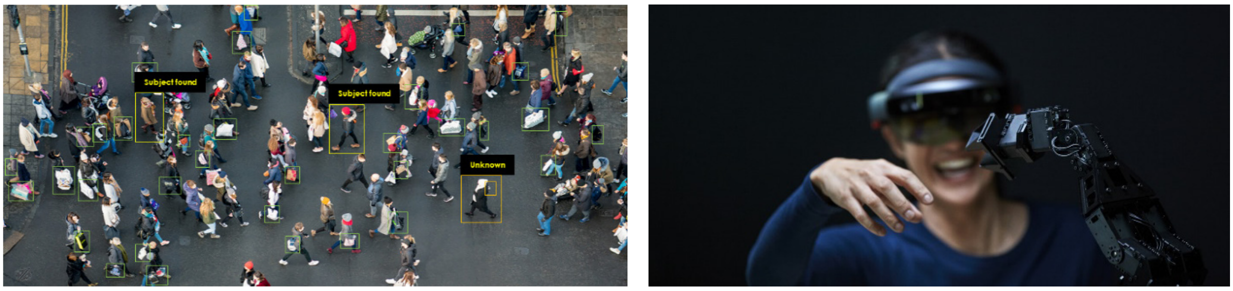


Figure 5-1: Edge Compute Real-Time AI and Video Analytics Supporting Search and Rescue, Object Localization, Image Enhancement, Behavioral Analytics, Predictions, etc. to Low Latency Interactive Human Machine Applications.

5.3 REAL-TIME SUPPORT

Real-time support can be divided into operational support of strategic, tactical and technical functions. In a real-time perspective, strategic support refers to strategic real-time aggregation of intelligence. This function defined as intelligence required for the current formation of policy and military plans and corresponds to the strategic level of warfare in the short time frame where situation awareness is one of the main supporting intelligence components. In the instant time frame, strategic support merges into both operational and tactical support. However, operational and tactical real-time support hosts additional functions such as task specific aiding functions and human augmentation performance enhancement functions where remote collaboration, reduction of human exposure, field support and logistic assistance are a few examples, see Figure 5-2.

Figure 5-2: Example of Real-Time Applications Used in Strategic, Operational and Tactical Functions.

Real-time support also incorporates assistance on soldier combat value, in both internal and external affects, specifically related to e.g., support of psychological restrains by improving awareness on local situation and team localization helping reduce stress and burden in terms of user role, user knowledge, skill, ability requirements, user limitations based on task characteristics.

During the COVID pandemic, technologies such as remote communication and remote control has paved the way to a rapid worldwide transition from face-to-face to mediated communication in both private and professional settings. At the same time, companies and organizations face several developments, such as personnel shortage, flexible work schedules and increased specialization that make it increasingly difficult to get the required people together physically when this is needed, e.g., for complex problem solving or when tasks must be trained as a team. And particularly relevant for defence organizations, the technology will introduce possibly critical security and privacy risks when personnel and private networks are exposed to public networks and Internet. Depending on information sharing classification, mediated communication will put high demand on cyber security and network architectures.

Due to current investments from big tech companies, the expectation is that improved tools for mediated communication will soon become available, based on a combination of immersive technologies such as Extended Reality (XR) which is used as overarching term for virtual, mixed and augmented reality technologies (i.e., establishing a metaverse of interoperable immersive technologies). The advent of XR technology for mediated communication creates many opportunities for the military and raises a number of concerns. Just like civil organizations, defence organizations are currently struggling with the consequences of the COVID pandemic and with the question to which degree the current transition to working at a distance will last. They also suffer, in many cases more than the civil organizations, from personnel shortage and scarcity of specialists. Furthermore, the military are routinely required (more than most civil professionals) to work at dispersed, sometimes remote locations, away from their home front. While these conditions will often only require ‘standard’ audio or audiovisual ways of communication, there are several settings that could benefit substantially from more advanced communication network, edge compute capabilities and XR communication tools. Some of the relevant application examples for remote collaboration during bang are:

- Expertise at a distance (for remote maintenance, medical care, etc.).
- Specific types of meetings (brainstorm sessions, problem solving, strategic decision making, etc.).
- Communication with the home front.
- Just in time training applications (teaching of complex skills, training of members of a command center, training at a distance).
- Command and Control.

5.4 REMOTE CONTROL

A remotely controlled system allows an operator to operate a system without having to be in close proximity. Broadly speaking there are two types of remote-controlled systems. The first of which is unmanned systems, which may or may not have a degree of autonomy. The second type of system is a Telexistence system that allows a user to project their presence to another environment to control a system. This relies on the integration of telepresence which allows the user to see and hear in the remote environment, robotics which enables the user to move in the remote environment and haptics which enables to user feel the remote environment. Telexistence had been enhanced by rapid advances in immersive technologies such as virtual and augmented reality haptics, robotics and computer vision.

Both types of remote-control systems use a range of human machine interface methods ranging from relatively simple throttle and joystick inputs to more complex motion capture-based control to enable a human operator to control a remote system. The key distinction between unmanned and a Telexistence system, is that the Telexistence systems provide real-time feedback to the user via the haptic, visual auditory feedback sensor. This essentially allows the Telexistence to augment the skills and experience of the user while reducing the risk to the user by removing them from a hazardous environment. Whereas an unmanned systems augments human performance by adding degrees of autonomy, potentially reducing task difficulty.

The main advantage of remote-control systems is that they allow operators to project their effort to another physical location while mutating the risks associated with humans operating in dangerous environments.

5.4.1 Human Performance Constructs

Real-time support and remote-control systems are designed (or are being developed) to support a number of human performance constructs such as cognitive/physical loading, and situational awareness, as well as mitigating risk and skill fade. While some of the technology is relatively mature and well understood especially in the remote-control domain where it has been shown to enhance human performance constructs such as situational awareness e.g., operators are now able to manage swarms of unmanned aerial vehicles while maintaining a high level of situational awareness [2]. However, there is a large body of evidence which demonstrates that the addition of real-time and remote-control systems can have detrimental effects on human performance [3], so the implementation of these technologies needs to be carefully considered. The advent of low-cost commercial off the shelf immersive technologies may potentially bring benefits to real-time and remote-control technologies. In the case of real-time support, augmented reality head mounted displays such as the Microsoft HoloLens 2®, be operated hands free reducing the physical load as they do not need to hold a device while completing their task. For remote control, virtual reality and haptic controls can possibly allow the user to operate systems in an intuitive way, however many of these systems are still in early development and require significant investment before they are ready for regular use. The potential benefits of using immersive technology with remote support and real-time systems could be tempered by our lack of understanding of how they affect the user. Both virtual and augmented reality are known to cause cybersickness in certain users [4] and users of virtual reality head mounted displays are consistently shown to underestimate distance [5] which could have an effect when completing fine motor tasks in a telexsistence task, so further research is required to fully understand how to mitigate these effects on the performer. Further insights into cybersickness effects and mitigation techniques, see Ref. [6].

5.5 STRENGTHS, WEAKNESSES, OPPORTUNITIES AND THREATS (SWOT) ANALYSIS

Table 5-1: Real-Time Support and Remote Control SWOT Analysis.

<table>
<tr><th colspan="2">Real-Time Support and Remote Control</th></tr>
<tr><td>Strengths

Remote collaboration
Remote warfare with support of strategic and tactical UAV, UGV, USV providing SA and RWS target engagement
Decision support by real-time local and global SA
Remote healthcare and telemedicine
Less demand on logistics “human needs” by utilization of logistics drones
Improved in-field training and education
Remote control

Enhances human capability
Increases the ability of humans to operate in hazardous conditions i.e., Explosive Ordnance Device (EOD) disposal
Allows the user to operate in multiple locations, so expertise can be accessed at the point of need immediately

</td>
<td>Opportunities

Improved capabilities along with the evolution of the 3GPP standards
Virtual operations reducing cost for travels and space requirements
Improved support for field training and real-time tutoring
Deep integration of network compute blurring the line between the device, the edge of the network and in the cloud
Compute intensive encryption schemes possible on powerful networks such as Homomorphic encryption
Local and central compute can be a single unified, integrated execution environment for distributed applications
Remote control

Increased ability to operate in EOD, Chemical, Biological, Radiological, Nuclear (CBRN) Disaster Response and Decontamination, battlefield evacuation operations
Reduced logistical burden
Armed forces could consist of fewer, more highly skilled personnel. Reducing overall personnel requirements

</td></tr>
<tr><td>Weaknesses

High demands on data network enabled capabilities
Downtime
Could impact the human performance construct if not properly designed or utilized such as cognitive load, working memory, simulation/cyber sickness and multi-task paradigm
Software driver requirements and updates
</td>
<td>Threats

Cyber threats and jamming causing network failure and application downtime
Reliability based on data and network infrastructure instability
Power failure causing network shutdown
Complexity
Program crashes
</td></tr>
</table>

Weaknesses (cont'd)	Threats (cont'd)

• Remote control
  • High network demand required for Telexistence systems
  • High power demands for remote systems
  • High cognitive loads demands placed on the operator
  • Skill fade of less experienced operators
	• Remote control
	  • Cyber Vulnerabilities
	  • Reliance on commercial off the shelf technologies
	  • Poorly designed human machine interfaces

5.6 REFERENCES

[1] Karasinski, J.A., Joyce, R., Carroll, C., Gale, J., and Hillenius, S. (2017). "An Augmented Reality/Internet of Things Prototype for Just-in-Time Astronaut Training." International Conference on Virtual, Augmented and Mixed Reality, pp. 248-260. Springer.

[2] Hocraffer, A., and Nam, C.S. (2017). "A Meta-Analysis of Human-System Interfaces in Unmanned Aerial Vehicle (UAV) Swarm Management." Applied Ergonomics, 58, pp. 66-80.

[3] Parasuraman, R., and Riley, V. (1997). "Humans and Automation: Use, Misuse, Disuse, Abuse." Human Factors, 39(2), pp. 230-253.

[4] Tian, N., Lopes, P., and Boulic, R. (2022). "A Review of Cybersickness in Head-Mounted Displays: Raising Attention to Individual Susceptibility." Virtual Reality 26, pp. 1409-1441.

[5] El Jamiy, F., and Marsh, R. (2019). "Distance Estimation in Virtual Reality and Augmented Reality: A Survey." In 2019 IEEE International Conference on Electro Information Technology (EIT), pp. 063-068. IEEE.

[6] NATO Science and Technology Office (2021). "Guidelines for Mitigating Cybersickness in Virtual Reality Systems." Technical Report STO-TR-HFM-MSG-323. NATO Science & Technology, Neuilly-sur-Seine, France.

Chapter 6 – COGNITIVE MONITORING AND OPTIMIZATION

Vincent Capaldi and Thomas Balkin

6.1 INTRODUCTION

Sleep loss, fatigue, and other performance-decrementing factors are unavoidable during most military operations. As detailed in the other chapters of the present report, augmentation strategies and technologies that sensitively detect human operator cognitive performance deficits, and that effectively compensate for those deficits in real time, can extend and enhance the military effectiveness of sleepy, fatigued, and/or otherwise cognitively impaired operators. Military effectiveness depends upon the extent to which both halves of the human/machine dyad are operating optimally. This is achieved by sustaining cognitive performance for as long as possible, and incrementally applying cutting-edge, finely-tuned augmentation technologies as needed.

Military operator effectiveness increasingly requires cognitive as well as physical prowess – i.e., the ability to understand an evolving battlespace and recognize/respond appropriately to both emergent threats and opportunities in real time. Cognitive performance is a function of many factors including level of training/expertise, motivation level, amount and quality of feedback, etc. However, the upper limit of cognitive performance at any given timepoint depends on the brain's physiological capacity to productively engage in mental work, and this is largely determined by the interaction of two factors:

a) Sleepiness level (the product of sleep debt level × circadian rhythm phase).

b) Fatigue level (the product of 'time on task' × cognitive load).

Over the past several decades, knowledge regarding the physiological basis of sleepiness and its effects on performance has expanded greatly [1], [2], [3]. In contrast, relatively little is currently known about the physiological basis of cognitive fatigue, a knowledge gap that is currently being addressed by HFM-331 RTG: "Biomedical Bases of Mental Fatigue and Military Fatigue Countermeasures."

During both civilian and military operations – but especially during continuous and sustained military operations – higher-order cognitive abilities that are critical to mission performance (i.e., "executive functions" including situational awareness, judgment, problem solving, decision making, memory, response time, etc.) are invariably decremented [2], [4], [5], [6], [7].

Fatigue Risk Management Systems (FRMSs) are comprehensive programs designed to maximize performance and safety in operational environments in which sleepiness and fatigue are potentially present. Although the details of each FRMS varies by operation or industry, all FRMSs generally include the following components [8]:

a) A set of fatigue management-relevant policies that reflect evidence-based industry regulations;

b) Appropriate education and training programs for individuals at all operational levels;

c) A straightforward process for self-reporting subjective fatigue without fear of reprisal; and

d) Procedures for investigating, reporting, and recording possible fatigue-related events.

But most importantly, and central to the ultimate utility of a FRMS is:

e) The ability to actually monitor and measure operationally relevant levels of fatigue and sleepiness, and then intervene in a timely and effective manner.

This is also the component of FRMSs that is most challenging – largely because the neurophysiological underpinnings of sleepiness, fatigue, and sleepiness x fatigue interactions (as well as the neurophysiological basis of individual differences in chronotype, susceptibility to sleep loss, and 'time on task') – are not yet fully understood. Nevertheless, fatigue monitoring technologies – which are critical components of modern FRMSs – have advanced significantly over the past two decades, with progress on understanding, quantifying, monitoring and counteracting sleepiness and fatigue [9]. The following sections provide an overview of the current 'state of the science.'

6.2 MONITORING

6.2.1 Subjective Sleepiness and Fatigue

The simplest way to assess sleepiness and fatigue in operational environments is to periodically ask the operators how they feel. Not surprisingly, prior research shows that operators' generally report increasing fatigue and/or sleepiness as the number of hours on the job increases [10]. And although other factors such as perceived levels of job-related stress can influence these subjective ratings [10], they nevertheless tend to correlate reasonably well with objective measures of sleepiness – especially with measures of chronic, trait-like sleepiness (i.e., the type of sleepiness measured by the Epworth sleepiness scale [11]). Of course, although chronic trait-like sleepiness is clearly relevant to operational performance and safety, identification of the more acute fluctuations in sleepiness that occur during work/duty shifts are critical for effectively managing performance and safety in real time. For a variety of reasons [12], [13] the correlations between subjective and objective measures of short-term (state) sleepiness are not strong, with objective measures generally more sensitive than subjective measures [14]. This is especially true for individuals suffering from sleep disorders such as obstructive sleep apnea [15] and narcolepsy [16] or for non-sleep-disordered individuals (such as shift workers) whose sleep has been chronically restricted [17]. This is because chronically sleepy individuals, regardless of whether that sleepiness is due to a sleep disorder or a suboptimal sleep/wake schedule – tend to subjectively habituate to reduced alertness over time, although objective measures of sleepiness reveal no evidence of actual adaptation.

Therefore, it can be concluded that objective measures of sleepiness are preferable to subjective measures of sleepiness in operational environments – with the caveat that, as pointed out by Balkin [9] self-assessments of excessive sleepiness or fatigue should always be taken seriously, regardless of what may be indicated by objective measures.

6.2.2 Objective Performance

Monitoring systems that provide direct, automatically-collected, real time information on operationally relevant performance (e.g., piloting drones, driving trucks) can be invaluable for identifying negative trends in operator sleepiness and fatigue – and unlike subjective ratings, they are immune to habituation. Another advantage of such systems is their self-evident operational relevance (i.e., all questions regarding ecological and construct validity are obviated) – a feature that can promote acceptance and utilization within the relevant operational communities.

Of course, the metrics employed by such systems vary by industry and job. And like subjective ratings, there are some strengths and weaknesses that, to some extent, all performance monitoring systems share. Take for example the performance monitoring system developed by Mollicone et al. in 2007 [18] for the trucking industry: Based on prior findings showing that sleep loss results in lapses in attention, and reasoning that such lapses increasingly result in 'near misses' (incidents in which collisions are narrowly averted due to swiftly executed evasive action(s) by the driver), Mollicone et al. [18] developed a system in which "hard braking" events are monitored and recorded. Hard braking constitutes an evasive action, and it is reasonable to presume that most instances that are not associated with an actual accident constitute near-miss events. Because near-miss events occur at a much higher rate than actual traffic accidents, implementation of this system could clearly improve driver safety by prompting interventions (e.g., ingesting caffeine and/or taking a nap at a nearby rest stop before continuing to drive) when such events are detected by sensors connected to the on-board computer.

However, the usefulness of this system is reduced when there is relatively little traffic (i.e., less opportunity to rear-end other vehicles) – for example during the long, straight drives on highways connecting distant cities in Australia. This means that the sensitivity of this performance monitoring system logically varies not only as a function of driver alertness, but also as a function of the extant traffic conditions. That being the case, the absence of hard braking events cannot be interpreted as evidence for the absence of fatigue or sleepiness. Additionally, it should be noted that although hard braking events almost always indicate a near-miss event, they sometimes actually reflect the action of a maximally alert driver who has been unexpectedly "cut off" by another vehicle. Thus, as this example illustrates, performance monitoring does not necessarily provide a highly sensitive and specific measure of operator fatigue or sleepiness, even when the ecological validity of the performance measure being monitored is high.

In some cases, and to some extent, these problems might be mitigated by incorporating additional embedded performance measures into the monitoring system. For example, it is possible that improved sensitivity and specificity of Mollicone et al.'s system [18] might be achieved by adding automatically-detected lane deviations, which have been shown to increase in frequency and duration with increasing sleepiness [19] and/or number of small steering adjustments per minute, which decline with increasing sleepiness [20], [21]. But because there are a large number of possible causes for variability in driving performance [22] – or, for that matter, performance of any operational task – no operational performance metric (or combination of metrics) provides a direct window into an operator's level of sleepiness and/or fatigue. In addition, the relationship between sleepiness/fatigue level and operational performance is nonlinear because motivated operators can usually maintain nominally adequate operational performance during early stages of sleepiness and/or fatigue by application of increased effort – at least for a while [23]. This suggests that performance-based indicators of drowsiness may be absent when sleepiness/fatigue is mild (i.e., when the window of opportunity for administering countermeasures is relatively optimal), and manifest only after moderate to severe levels of impairment have been reached [24] – i.e., the point at which an operator's ability to sustain safe performance has already been overwhelmed.

6.2.3 Psychophysiological Indicators of Sleepiness and Fatigue

Given the aforementioned unreliability of self-ratings, and the pitfalls associated with tracking operational performance as a means of monitoring sleepiness and fatigue in the operational environment, it is logical to surmise that a better approach might be to monitor sleepiness and fatigue more directly – i.e., to monitor the brain state itself, rather than the downstream manifestations of the brain state (i.e., performance and/or subjective ratings). Potentially, this approach offers two major advantages:

a) To the extent that a measure taps directly into the neurophysiological processes that underlie sleepiness and fatigue, that measure will be unaffected by factors like motivation and habituation.

b) Such measures would also make possible the recognition of sleepiness and fatigue in their early stages, facilitating the ability to apply countermeasures in a timely manner (i.e., before meaningful performance deficits manifest).

Recognizing the potential utility of this approach, several psychophysiological monitoring technologies have been developed, tested, and (to varying extents) validated. These have included devices to monitor EEG, eye movements, percent eyelid closure, respiration rate, heart rate, and galvanic skin response, to name but a few.

However, these technologies all have a similar problem: determination of the threshold at which alarms and interventions should be initiated and/or the point at which an operator's ability to sustain nominally adequate operational performance has been meaningfully impacted [9]. Stated simply: the problem is that if the alarm threshold is set too low (i.e., the sensitivity is set to sound the alarm when extremely mild levels of sleepiness and/or fatigue are detected) the risk of "false positive" alarms will be increased, and the difference between the alarm threshold and the operators' subjective perception will be large – both of which would likely result in a tendency for the operator to ignore the alarms. In contrast, if the alarm threshold is set too high (i.e., the sensitivity is set to sound the alarm only when clear signs of significant sleepiness and/or fatigue are detected, then the risk of "false negatives" would be increased (i.e., the alarm would fail to sound at moderate levels of sleepiness and fatigue that actually reflect a meaningful level of impairment). In addition to the obvious problem that higher thresholds increase the risk that significant levels of impairment could be missed, a secondary problem is the possibility that operators would rely too heavily on the alarms and discount their subjective perceptions ("Although I feel pretty tired, the alarm hasn't sounded yet so I must be okay"). To some extent, these problems could be mitigated if the psychophysiological monitoring system utilizes a validated, evidence-based "graded" alarm system (e.g., red, yellow, green indicators), with, for example, "yellow" indicating that current alertness is within normal limits but is predicted (see section on mathematical performance prediction modeling) to degrade to a significant extent within the next ~2 hours.

6.2.4 Monitoring Summary and Conclusions

Because there are different strengths and weaknesses associated with each monitoring modality, the overall sensitivity and specificity of a monitoring system can generally be improved via multi-modal monitoring (i.e., monitoring of subjective perceptions, objective performance, and psychophysiological indicators of sleepiness and fatigue). It can also sometimes be improved by monitoring additional channels within a single modality (e.g., multiple aspects of driving performance). There are, of course, practical limitations, with the logistical feasibility of operator monitoring (and the appropriateness of each monitoring modality) varying as a function of the nature of the operational environment.

But even an ideal monitoring system (i.e., one that identifies and quantifies extant, operationally relevant cognitive impairment level with perfect precision) would not provide enough information to optimize operational performance and safety. What is also needed is the ability to accurately interpret data produced by the monitoring system. What, for example, does the operator's current level of sleepiness and/or fatigue mean for his/her current and near-term future performance? If the operator's status is currently "green" how much longer can it be expected that he/she can continue to work until his/her status turns to "yellow" and then "red"? What interventions (e.g., naps, rest breaks, caffeine) can be prophylactically applied either singly or in combination, and at what time and at what dose levels, to maximize subsequent alertness and safety? And for how much longer will such interventions extend nominally safe and effective operator performance? In the next section, the Unified Model of Performance (UMP) [25] – an evidence-based mathematical performance prediction model that has been (and continues to be) developed for the purpose of providing input to decision makers who are faced with such questions – is described.

6.3 MATHEMATICAL PERFORMANCE PREDICTION MODELING

Several Mathematical Performance Prediction Models (MPPMs) have been developed for the purpose of facilitating fatigue management in operational environments [26]. In most of these models, the concept of 'fatigue' is fundamentally identical to that of 'sleepiness', and most are based primarily (if not exclusively) on the "two process model of sleep regulation" proposed by Borbély in 1982 [27].

According to Borbély's model [27], the likelihood of entering Slow Wave Sleep (SWS) at any given time varies as a function of the interaction of two processes (factors): the sleep homeostat ("process S") and the circadian rhythm of alertness ("process C"). Because the likelihood of entering SWS is itself a direct reflection of both objective and subjective sleepiness, and because sleepiness accounts for much of the variance in operational performance, it was (and continues to be) logical to utilize Borbély's model as the foundation upon which MPPMs are built. To date, these models have been used primarily as scheduling aids – e.g., to inform development of work-rest schedules that help ensure that operators are maximally alert during periods when optimal performance is critical [28].

However, the potential of MPPMs will not be fully realized until they are integrated into FRMSs in a manner that informs decision making in real time. Currently, the most advanced MPPM developed specifically for this purpose is the Unified Model of Performance (UMP) [25]. The UMP was developed collaboratively by researchers at the US Army's Biotechnology High Performance Computing Software Applications Institute (BHSAI) and the Walter Reed Army Institute of Research (WRAIR) and has been validated reasonably well [29]. In part, the utility of this model derives from the fact that it represents the integration of two MPPMs: a Borbély (1982)-based model [27] that accounts for the effects of sleep loss and the circadian rhythm of alertness [30] and a model that predicts the performance effects of caffeine [31]. The latter capability is important because caffeine is an effective fatigue countermeasure that is widely available and often consumed in many operational environments. Therefore, given the ubiquitousness of caffeine consumption, a performance prediction model that fails to account for the performance-enhancing effects of caffeine is of limited real-world utility.

Another significant advantage of the UMP as instantiated by the 2B-Alert app [32] is that the predictions are individualized – i.e., the algorithm can essentially "learn" each individual user's sensitivity to variations in nightly sleep amounts [33]. This is accomplished by inputting nightly sleep data (e.g., via automatic wireless transmission of wrist-actigraphy scored sleep data) and comparing it to the user's behavioral data (occasional performance a 5-minute psychomotor vigilance test [34] administered on a smartphone). There is also a free, web-based version of 2B-Alert [35] (https://2b-alert-web.bhsai.org/2b-alert-web/login.xhtml) although this version does not include the ability to individualize performance predictions and the site currently works only when accessed with a PC.

Utilization of the 2B-Alert app in operational environments can improve performance and safety by:

a) Predicting when and how long an individual operator's performance will remain within whatever performance range is deemed acceptable (e.g., at the performance level associated with a blood alcohol level of .05 or lower); and

b) By informing decision making regarding the optimal dosing and timing of interventions like caffeine and/or naps.

Typically, such interventions are administered only after subjective and/or objective evidence of impairment is manifest. But 2B-Alert can predict dangerous sleepiness-induced dips in performance well in advance and can recommend prophylactic interventions to help ensure that such dips in performance do not occur. Currently, 2B-Alert (both the app and the web-based version) provide recommendations for optimal administration of caffeine [36] but it is possible that advanced development efforts will include additional interventions (e.g., modafinil) when and if sufficient data become available.

Among the virtues of the UMP/2B-Alert is the fact that its predictions are not counterintuitive (i.e., as one would expect, it predicts greater performance deficits as sleep loss accrues across days of sleep restriction, with performance within days mediated by the circadian rhythm of alertness). And because they are quantified, the predictions provide a level of precision that facilitates decision making. However, the precision of the 2B-Alert predictions can be a double-edged sword: No MPPM can take into account all of the relevant factors that determine individual operator performance (e.g., motivation, experience, personality, physical dexterity and endurance, to name but a few). The danger is that the apparent precision of its predictions will lull the human decision-maker into complacent over-reliance on the MPPM – using it as an ultimate and final arbiter rather than the 'decision aid' that it is meant to be.

6.3.1 Countermeasures

As indicated in the previous section, the need for administration of countermeasures is typically determined on the basis of the subjective experience of fatigue and/or sleepiness, although MPPMs can be used to anticipate deficits, facilitating the precision with which interventions can be administered prophylactically. Of course, the type of intervention that is most appropriate depends upon the underlying cause of the deficit: If the problem is fatigue from extended 'time on task,' then the optimal countermeasure would be 'time off task' (rest). If the problem is sleepiness, then the optimal countermeasure is sleep. When such interventions are not feasible (e.g., during continuous combat operations) then it can become necessary to administer interventions that help sustain performance by "flogging" the physiological processes that sustain alertness and performance (i.e., without actually restoring those processes).

In general, there are two types of pharmacological intervention strategies that can be employed: administration of stimulants to sustain alertness and performance acutely and/or administration of hypnotic medications to facilitate restorative sleep (e.g., when sleep would otherwise be short or disrupted – e.g., during the ascending phase of the circadian rhythm of alertness, in an environment that is not conducive to sleep (well-lit and/or noisy), and/or following rapid deployment across multiple time zones).

It is beyond the scope of the present report to review all of the available options for pharmacological interventions in operational environments, but for a recent, relevant review see Alger et al. (2021) [37]. It is, however, important to note that there are costs associated with administration of any pharmaceutical for the purpose of sustaining operational performance. For example, although dextroamphetamine, modafinil, and caffeine are each effective for counteracting the effects of sleep loss [38] there are potential downsides associated with each. For example, it is well known that dextroamphetamine has high abuse potential, and discontinuation results in severe rebound deficits in alertness and performance [39]. Modafinil has much less abuse potential than d-amphetamine [40] but it may not be appropriate for some operational environments (e.g., aviation) because it may increase vertigo, nausea, and dizziness [41] (Caldwell et al., 2000). Caffeine is generally well-tolerated and effective and has the advantage that most operators have a significant amount of experience with caffeine (i.e., they know how sensitive they are to its effects, and whether they experience any side effects such as jitteriness) so it is the stimulant of choice in most military environments [42]. However, as is well known, ingestion of caffeine within a few hours of bedtime results in sleep disruption, which can reduce next-day alertness and prompt increased intake of caffeine, in a vicious circle manner [37].

For all sleep-inducing (hypnotic) agents, the primary downside is the drug "hangover" effect – deficits in performance that are produced by the hypnotic itself. For this reason, use of hypnotic medications by military personnel who might need to awaken rapidly (while the drug is still active) and start performing a task (e.g., respond to an early morning attack by enemy forces) may be precluded. In general, the level of performance deficit produced by a hypnotic is tightly and positively correlated with the efficacy of that hypnotic.

Another approach is "sleep banking" – i.e., extending nightly sleep duration (to ~10 hours) for several days prior to an anticipated period of sleep loss. This has been shown to reduce the rate at which performance declines during a period of sleep loss, and to improve the speed with which full recovery from that sleep loss is subsequently achieved [43], [44], [45]. Of course, the downside of this approach is twofold: a) It is not always possible to predict upcoming periods of sleep loss (which can occur in response to spontaneous emergency situations), and b) To the extent that nightly time in bed is extended, less waking time is available for performance of other duties (i.e., it can be a "zero sum game").

New techniques such as Transcranial Electrical Stimulation (TES) to enhance the recuperative value of sleep and/or to enhance waking alertness and performance are currently being explored. However, the efficacy and potential risks/costs associated with these technologies are currently unknown, and several more years of testing and validation will be required before they are ready to be transitioned to the operational environment [46].

6.4 CONCLUSION AND SWOT ANALYSIS

Military efficacy is enhanced when both halves of the human/machine dyad are operating with optimal effectiveness. Human cognitive performance capacity is largely determined by operator sleepiness and fatigue levels. Comprehensive Fatigue Risk Management Systems (FRMSs) include monitoring (subjective and objective indicators of indicators of sleepiness and fatigue); an armamentarium of fatigue/sleepiness countermeasures (both pharmacological and non-pharmacological); and a mathematical performance prediction model (to inform decision making regarding the optimal countermeasure type, timing and "dosage"). As indicated in the SWOT analysis, each FMRS component currently exists at an implementable level of technical readiness, but full integration among the components has not yet been achieved, and improvements to each of the components are possible (and expected). Threats primarily consist of the possibility that components of a comprehensive FRMS (especially a MPPM) could be used inappropriately (e.g., replace rather than inform decision making by the commander).

6.4.1 Strengths, Weaknesses, Opportunities, Threats (SWOT) Analysis

Table 6-1: Cognitive Monitoring and Optimization SWOT Analysis.

Cognitive Monitoring and Optimization	
Strengths	**Opportunities**
• Overall: State of the science for each separate component (monitoring, enhancement, mathematic performance prediction modeling) is at an implementable level (at least for some tasks) • Monitoring: A plethora of technologies exist for monitoring psychophysiological variables in real time • Enhancement: Several pharmaceuticals available to improve cognition directly (i.e., nootropics) and indirectly (e.g., hypnotics to improve sleep) • Modeling: 2B-Alert provides individualized predictions based on sleep/wake schedule, caffeine consumption – and has "optimize" function – predictions are not counterintuitive	• Overall: Interest from SOCOM and other military groups is high – especially with currently increased general awareness of the importance of sleep for performance and health • Monitoring: New technologies (e.g., fieldable, single dry electrode EEG devices) are being developed at a rapid pace • Enhancement: Non-pharmacological methods (e.g., tCDS) may prove effective without side effects – currently being investigated • Modeling: could it benefit from application of/merging with AI (e.g., to enhance the process of individualizing performance predictions)
Weaknesses	**Threats**
• Overall: Components have not yet been fused into a comprehensive management system. Not a simple task (e.g., relative weighting of monitoring and predictions?) • Monitoring: is the "weakest link": no "sleepiness breathalyzer" exists (yet). Interpretation of psychophysiological (and even performance) variables is not always straightforward • Enhancement: All pharmaceuticals exact some sort of "cost" (i.e., have a downside) • Modeling: Not all relevant mediators of performance have been modeled and/or integrated with 2B-Alert (e.g., "time on task effects" – what HFM-331 calls "fatigue")	• Overall: Over-dependence on the system, which should be utilized by commanders as a decision aid, not a decision-maker • Monitoring: inappropriate threshold setting (where is green, yellow, red?) • Enhancement: Advantages conferred could be offset by improper application, (e.g., reducing sleep time to the same extent that TES enhances the recuperative value of sleep resulting in no net gain in alertness and performance) • Modeling: Liability implications if model recommendations are ignored (?) • Privacy and cyber security issues

6.5 REFERENCES

[1] Reynolds, A.C., and Banks, S. (2010). "Total Sleep Deprivation, Chronic Sleep Restriction and Sleep Disruption." Prog Brain Res. 185, pp. 91-103. DOI: 10.1016/B978-0-444-53702-7.00006-3

[2] Durmer, J.S., and Dinges, D.F. (2005). "Neurocognitive Consequences of Sleep Deprivation." Semin Neurol. Mar 25(1), pp. 117-29. DOI: 10.1055/s-2005-867080

[3] Goel, N., Basner, M., Rao, H., and Dinges, D.F. (2013). "Circadian Rhythms, Sleep Deprivation, and Human Performance. Prog Mol Biol Transl Sci. 119, pp. 155-90. DOI: 10.1016/B978-0-12-396971-2.00007-5

[4] Killgore, W.D. (2010). "Effects of Sleep Deprivation on Cognition." Prog Brain Res. 185, pp. 105-29. DOI: 10.1016/B978-0-444-53702-7.00007-5

[5] Lowe, C.J., Safati, A., and Hall, P.A. (2017). "The Neurocognitive Consequences of Sleep Restriction: A Meta-Analytic Review." Neurosci Biobehav Rev. Sep, 80, pp. 586-604. DOI: 10.1016/j.neubiorev.2017.07.010

[6] Diamond, A. (2013). "Executive Functions." Annu Rev Psychol. 64, pp. 135-68. DOI: 10.1146/annurev-psych-113011-143750

[7] Dinges, D. F., and Powell, J. W. (1985). "Microcomputer Analyses of Performance on a Portable, Simple Visual RT Task During Sustained Operations." Behavior Research Methods, instruments, & Computers, 17(6), pp. 652-655.

[8] Gander, P., Hartley, H., Powell, D., Cabon, P., Hitchcock, E., Mills, A., and Popkin, S. (2011). "Fatigue Risk Management: Organizational Factors at the Regulatory and Industry/Company Level." Accid Anal Prev 43, pp. 573-90.

[9] Balkin, T.J. (2022). "Assessment of the Occupational Sleep Medicine Field." In M. Kryger, C. Goldstein, T. Roth and W. Dement (Eds.), Principles and Practice of Sleep Medicine, 7th Edition, pp. 814-820. Elsevier.

[10] Park, J., Kim, Y., Chung, H.K., and Hisanaga, N. (2001). "Long Working Hours and Subjective Fatigue Symptoms." Ind Health 39(3), pp. 250-4.

[11] Johns, M.W. (1991). "A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. Sleep 14(6), pp. 540-5.

[12] Saletin, J.M., Hilditch, C.J., Dement, W.C., and Carskadon, M.A. (2017). "Short Daytime Naps Briefly Attenuate Objectively Measured Sleepiness Under Chronic Sleep Restriction." Sleep 40(9). DOI: 10.1093/sleep/sleep/zsx118.

[13] Tremaine, R., Dorrian, J., Lack, L., Ferguson, S., Zhou, Z., and Roach, G. (2010). "The Relationship between Subjective and Objective Sleepiness and Performance during a Simulated Night-Shift with a Nap Countermeasure." Appl Ergon 42(1), pp. 52-61.

[14] Kosmadopoulos, A., Sargent, C., Zhou, X., Darwent, D., Matthews, R.W., Dawon, D., and Roach, G.D. (2017). "The Efficacy of Objective and Subjective Predictors of Driving Performance Impairment During Sleep Restriction and Circadian Misalignment." Accid Anal Prev 99(Pt B), pp. 445-51.

[15] Mehra, R., Wang, L., Andrews, N., Wilson Tang, W.H., Young, J.B., Javaheri, S., and Foldvary-Schaefer, N. (2017). "Dissociation of Objective and Subjective Daytime Sleepiness and Biomarkers of Systemic Inflammation in Sleep Disordered Breathing and Systolic Heart Failure." J Clin Sleep Med 13(12), pp. 1411-22.

[16] Sangal, R.B., Mitler, M.M., and Sangal, J.M. (1999). "Subjective Sleepiness Ratings (Epworth Sleepiness Scale) Do Not Reflect the Same Parameter of Sleepiness as Objective Sleepiness (Maintenance of Wakefulness Test) in Patients with Narcolepsy." Clin Neurophysiol, 110(12), pp. 2131-5.

[17] Philip, P., Sagaspe, P., Prague, M., Tassi, P., Capelli, A., Bioulac, B., Commenges, D., and Taillard, J. (2012). "Acute versus Chronic Partial Sleep Deprivation in Middle-Aged People: Differential Effect on Performance and Sleepiness. Sleep. Jul 1, 35(7), pp. 997-1002. DOI: 10.5665/sleep.1968. PMID: 22754046; PMCID: PMC3369235.

[18] Mollicone, D., Kan, K., and Mott, C., Bartels, R., Bruneau, S., van Wollen, M., et al. (2019). "Predicting Performance and Safety Based on Driver Fatigue." Accid Anal Prev. 126, pp. 142-145.

[19] Caponecchia, C., and Williamson, A. (2018). "Drowsiness and Driving Performance on Commuter Trips." J Safety Res. Sep, 66, pp. 179-186. DOI: 10.1016/j.jsr.2018.07.003.

[20] Lawoyin, S., Fei, D-Y., and Bai, O. (2015). "Accelerometer-Based Steering-Wheel Movement Monitoring for Drowsy-Driving Detection." Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering 229(2) ,pp. 163-173. DOI:10.1177/0954407014536148.

[21] Li, Z., Li, S.E., Li, R., Cheng, B., and Shi, J. (2017). "Online Detection of Driver Fatigue Using Steering Wheel Angles for Real Driving Conditions. Sensors (Basel)." Mar 2, 17(3)p. 495. DOI: 10.3390/s17030495.

[22] Sparrow, A.R., LaJambe, C.M., and Van Dongen, H.P.A. (2019). "Drowsiness Measures for Commercial Motor Vehicle Operations." Accid Anal Prev, May, 126, pp. 146-159.

[23] Hockey, R. (2013). The Psychology of Fatigue: Work, Effort and Control. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139015394.

[24] Schwarz, C., Gaspar, J., Miller, T., and Yousefian, R. (2019). "The Detection of Drowsiness Using a Driver Monitoring System." Traffic Inj Prev. 20(sup1), pp. S157-S161. DOI: 10.1080/15389588. 2019.1622005.

[25] Ramakrishnan, S., Wesensten, N.J., Kamimori, G.H., Moon, J.E., Balkin, T.J., and Reifman, J.A. (2016). "Unified Model of Performance for Predicting the Effects of Sleep and Caffeine." Sleep Oct 1, 39(10), pp. 1827-1841. DOI: 10.5665/sleep.6164

[26] Dawson, D., Ian Noy, Y., Härmä, M., Akerstedt, T., and Belenky, G. (2011). "Modelling Fatigue and the Use of Fatigue Models in Work Settings." Accid Anal Prev., Mar, 43(2), pp. 549-64. DOI: 10.1016/j.aap.2009.12.030

[27] Borbély, A.A. (1982). "A Two Process Model of Sleep Regulation." Hum Neurobiol. 1(3), pp. 195-204. PMID: 7185792

[28] Hursh, S.R., Redmond, D.P., Johnson, M.L., Thorne, D.R., Belenky, G., Balkin, T.J., Storm, W.F., Miller, J.C., and Eddy, D.R. (2004). "Fatigue Models for Applied Research in Warfighting." Aviat Space Environ Med, Mar, 75(3 Suppl), pp. A44-53; discussion A54-60. PMID: 15018265.

[29] Ramakrishnan, S., Wesensten, N.J., Balkin, T.J., Reifman, J. (2016). "A Unified Model of Performance: Validation of its Predictions Across Different Sleep/Wake Schedules." Sleep, Jan 1, 39(1), pp. 249-62. DOI: 10.5665/sleep.5358

[30] Rajdev, P., Thorsley, D., Rajaraman, S., Rupp, T.L., Wesensten, N.J., Balkin, T.J., and Reifman, J. (2013). "A Unified Mathematical Model to Quantify Performance Impairment for both Chronic Sleep Restriction and Total Sleep Deprivation." J Theor Biol. Aug 21, 331, pp. 66-77. DOI: 10.1016/j.jtbi.2013.04.013. Epub 2013 Apr 24. PMID: 23623949.

[31] Ramakrishnan, S., Laxminarayan, S., Wesensten, N.J., Kamimori, G.H., Balkin, T.J., and Reifman, J. (2014). "Dose-Dependent Model of Caffeine Effects on Human Vigilance during Total Sleep Deprivation." J Theor Biol. Oct 7, 358, pp. 11-24. DOI: 10.1016/j.jtbi.2014.05.017

[32] Reifman, J., Ramakrishnan, S., Liu, J., Kapela, A., Doty, T.J., Balkin, T.J., Kumar, K., and Khitrov, M.Y. (2019). "2B-Alert App: A Mobile Application for Real-Time Individualized Prediction of Alertness." J Sleep Res, Apr, 28(2), e12725. DOI: 10.1111/jsr.12725

[33] Liu, J., Ramakrishnan, S., Laxminarayan, S., Balkin, T.J., and Reifman, J. Real-Time Individualization of the Unified Model of Performance." J Sleep Res. Dec, 26(6), pp. 820-831. DOI: 10.1111/jsr.12535

[34] Dinges, D.F., and Powell, J.W. (1985). "Microcomputer Analyses of Performance on a Portable, Simple Visual RT Task During Sustained Operations." Behavior Research Methods, Instruments, & Computers, 17, pp. 652-655.

[35] Reifman, J., Kumar, K., Wesensten, N.J., Tountas, N.A., Balkin, T.J., and Ramakrishnan, S. (2016). "2B-AlertWeb: An Open-Access Tool for Predicting the Effects of Sleep/Wake Schedules and Caffeine Consumption on Neurobehavioral Performance." Sleep, Dec 1, 39(12), pp. 2157-2159. DOI: 10.5665/sleep.6318.

[36] Vital-Lopez, F.G., Ramakrishnan, S., Doty, T.J., Balkin, T.J., and Reifman, J. (2018). "Caffeine Dosing Strategies to Optimize Alertness During Sleep Loss." J Sleep Res, Oct, 27(5), e12711. DOI: 10.1111/jsr.12711

[37] Alger, S.E., Stekl, E., Allotey-Addo, S., Klosterman, G., and Balkin, T.J. (2021). "The Role of Pharmacological Interventions for Sleep Deprivation and Restriction." In Reference Module in Neuroscience and Biobehavioral Psychology, Elsevier. DOI: 10.1016/B978-0-12-822963-7.00009-8.

[38] Wesensten, N.J., Killgore, W.D., and Balkin, T.J. (2005). "Performance and Alertness Effects of Caffeine, Dextroamphetamine, and Modafinil During Sleep Deprivation." J Sleep Res, Sep, 14(3), pp. 255-66. DOI: 10.1111/j.1365-2869.2005.00468.x

[39] Hurst, P.M., Chubb, N.C., Bagley, S.K., and Ross, S. (1971). "Rebound from d-Amphetamine." Psychol Rep, Dec, 29(3), pp. 1023-33. DOI: 10.2466/pr0.1971.29.3f.1023

[40] Jasinski, D.R., Kovacević-Ristanović, R. (2000). "Evaluation of the Abuse Liability of Modafinil and Other Drugs for Excessive Daytime Sleepiness Associated with Narcolepsy. Clin Neuropharmacol. May-Jun, 23(3), pp. 149-56. DOI: 10.1097/00002826-200005000-00004.

[41] Caldwell, J.A. Jr, Caldwell, J.L., Smythe, N.K. 3rd, and Hall, K.K. (2000). "A Double-Blind, Placebo-Controlled Investigation of the Efficacy of Modafinil for Sustaining the Alertness and Performance of Aviators: A Helicopter Simulator Study. Psychopharmacology (Berl). Jun, 150(3), pp. 272-82. DOI: 10.1007/s002130000450.

[42] Rupp, T.L., Wesensten, N.J., Bliese, P.D., and Balkin, T.J. (2009). "Banking Sleep: Realization of Benefits During Subsequent Sleep Restriction and Recovery." Sleep, Mar, 32(3), pp. 311-21. DOI: 10.1093/sleep/32.3.311. PMID: 19294951; PMCID: PMC2647785.

[43] Arnal, P.J., Lapole, T., Erblang, M., Guillard, M., Bourrilhon, C., Léger, D., Chennaoui, M., and Millet, G.Y. (2016). "Sleep Extension Before Sleep Loss: Effects on Performance and Neuromuscular Function." Med Sci Sports Exerc. Aug, 48(8), pp. 1595-603. DOI: 10.1249/MSS.0000000000000925. PMID: 27015382.

[44] Arnal, P.J., Sauvet, F., Leger, D., van Beers, P., Bayon, V., Bougard, C., Rabat, A., Millet, G.Y., and Chennaoui, M. (2015). "Benefits of Sleep Extension on Sustained Attention and Sleep Pressure Before and During Total Sleep Deprivation and Recovery." Sleep, Dec 1, 38(12), pp. 1935-43. DOI: 10.5665/sleep.5244. PMID: 26194565; PMCID: PMC4667385.

[45] Balkin, T.J., Bliese, P.D., Belenky, G., Sing, H., Thorne, D.R., Thomas, M. et al. (2004). "Comparative Utility of Instruments for Monitoring Sleepiness-Related Performance Decrements in the Operational Environment." J Sleep Res. 13(3), pp. 219-227.

[46] NATO Science and Technology Office (2021). "Guidelines for Mitigating Cybersickness in Virtual Reality Systems." Technical Report STO-TR-HFM-MSG-323. NATO Science & Technology Organization, Neuilly-Sur-Seine, France.

Chapter 7 – AFTER ACTION REVIEW (AAR)

Jerzy Jarmasz and J.D. Fletcher

7.1 INTRODUCTION

The AAR is a well-established performance improvement practice. In its original conception within the US Army, an AAR "is a professional discussion of an event, focused on performance standards, which enables soldiers to discover for themselves what happened, why it happened, and how to sustain strengths and improve on weaknesses" [1] (from a US Army "training circular" still widely-cited [2]). While its original application was to Army collective training, the definition above suggests that AARs can be applied to a wide array of events to support performance, so long as some kind of observations from the event are available and there is some means of assessing them against meaningful standards or benchmarks. Accordingly, since its development in the 1970s, it has spread to other professions (notably medical) as well as to non-training domains (e.g., review of operational events), and has accordingly acquired different names (e.g., debrief, team huddle) while maintaining the same basic elements [3]. As such, the AAR can occur in many places, both "Left of Bang" and "Right of Bang," in the augmentation technologies timeline described in Chapter 2 of this report. However, the effectiveness of AARs has not been systematically researched outside of its use in training [3], [4], thus the present chapter will focus on AAR as a training intervention (i.e., as a "Left of Bang" domain), while acknowledging its application in work and operational settings.

Keiser and Arthur [3] note that many different "models" of the AAR process have been given by different researchers. Two somewhat contrasting examples are Villado and Arthur's [5] model (which was used as a basis for Keiser and Arthur's meta-analyses), shown in Table 7-1, and the Canadian Army AAR process, as described in Ref. [6] (see Figure 7-1). However, in Ref. [3] Keiser and Arthur stress that all of these characterizations follow the same basic logic: a performance event occurs, a review of the event is prepared (often concurrently with the event), and the review itself is performed; the review itself includes a presentation or recollection of specific actions (effective and ineffective) during the event, as well as a discussion to generate insights, feedback and plans to improve future performance.

Keiser and Arthur [3], [4] also noted that while the effectiveness of AARs as a general training intervention is well-established empirically, its effectiveness is not unqualified. That is, AARs can vary greatly in their effectiveness, or even fail to be of any use at all, depending on a variety of factors. One of the factors that influence AAR effectiveness is the technology used to conduct them. While AARs can be conducted with minimal-to-no augmentation technologies (e.g., a group of trainees discussing a training event with a facilitator, aided only by a notepad or whiteboard), it has been noted that the development of systematic AARs did not really take off until the advent of instrumented live Army training ranges in the 1970s, using technologies that allowed for the simulation of weapons effects and tracking of participant actions [7].

Table 7-1: AAR Phases as per Villado and Arthur (2013).

AAR Phase	Psychological Domain Theory
1) Review event objectives	Feedback theory, goal setting
2) Review event outcome	Feedback theory, goal setting, observational learning theory
3) Review of effective and ineffective actions	Observational learning theory
4) Discuss future objectives	Goal setting
5) Discuss strategy	Goal setting

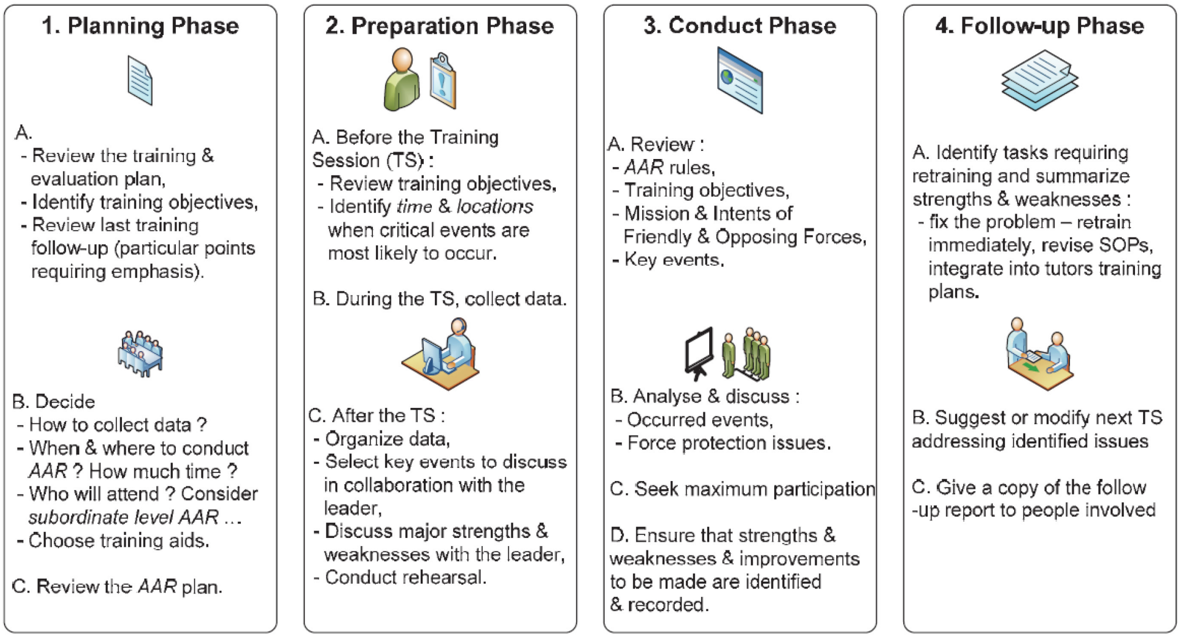


Figure 7-1: Typical Canadian Army AAR Phases [6].

Thus, augmentation technologies are integral the AAR process, and the scope for applying augmentation technologies to AARs is large. Technological supports to the AAR process are most pertinent in the provision of feedback (e.g., replays of specific events) during Phases 1 and 2 of the AAR process (Table 7-1). However, technological augmentation of the AAR process can occur across all phases of the AAR and can vary significantly as a function of the technological complexity of the training environment (e.g., live vs synthetic) and the task (e.g., small co-located team task vs large distributed collective task). One may even speak of a "technology gradient" in AARs: the more a training environment or task conditions are technologically complex, and involve synthetic or instrumented environments, the more these technologies will need to be relied on in order to support effective AARs.

With these introductory considerations in mind, the use of augmentation technologies in AARs is considered. This examination starts by characterizing the user requirements within the AAR process and the limitations user face which drive the application of augmentation technologies for AARs. Following this, specific AAR technology applications are described in more detail, and submitted these to a SWOT analysis. The chapter concludes with a consideration of the implications of the SWOT analysis for next steps in research on technologies for AARs.

7.2 USER REQUIREMENTS AND LIMITATIONS

7.2.1 User Roles and Skills

There are fundamentally three user roles in the AAR: trainees, AAR facilitators (who may be the trainees themselves), and support personnel who may be required to assist the first two groups of users.

Trainees: apart from performing the task under review, trainees must be able to understand objectives of the AAR, recall relevant activities of the event under review, and be open to discussing them in a critical but constructive way and learning from the discussion. Note that the task under review may can be either an individual or collective (team) task, and that the AAR itself may be performed as a team discussion or one-on-one between individual trainees and a facilitator (discussed next), regardless of the format of the task (individual vs team) itself.

Facilitator(s): The burden of conducting the AAR itself falls to the facilitator (or team thereof) [2], [6]. Nevertheless, exactly who the facilitator is and what they do varies widely across AAR applications. The AAR might be facilitated by SMEs (e.g., instructors), or by the trainees themselves (with or without an external facilitator for support [3]). The facilitator(s) may focus on effective behaviors by the trainees or may choose to also highlight ineffective behaviors. The facilitator(s) may also choose to engage trainees in a dialogue, allowing the discussion topics to emerge organically from the trainees' responses, or conversely may prefer to tightly direct the discussion onto pre-determined discussion points. In any case, the facilitator will need to understand relevant performance and training objectives, recall and present to trainees relevant activities from the event under review, correct trainees mistaken recall of these events if needed, assess these events with respect to the relevant objectives or standards and provide constructive feedback based on this comparison. The facilitator will need skills relevant to the conduct of group discussions, including managing potentially difficult conversation topics and eliciting participation from reluctant trainees. An ideal facilitator would be adept at fostering self-reflection and self-awareness in the trainees. If the trainees are facilitating the AAR themselves, they will need to be able to foster a collegial discussion among themselves, managing the trade-off between honesty (accuracy) and maintaining trust bonds and camaraderie (especially if they are a formed team or unit). The facilitator role is complex, and requires strong cognitive (recall, assessment and judgment), didactic (understanding of instructional methods and standards) and interpersonal (group dynamics management) skills.

Support personnel: the role and activities of support personnel in AARs is highly variable and difficult to characterize accurately without delving into the technologies used to support AARs. At this point, suffice it to say that the activities of the trainees and facilitators described above may require various kinds of support, from the simple logistics of setting up a space for the AAR discussion, to collating and preparing notes made by observers during the event for presentation by the facilitator, to the operation of simulator systems and range instrumentation to gather and process simulation data for a "digital" AAR [2], [6], [7]. Some of these roles may be performed by the facilitator, or may require dedicated personnel, depending on the size and complexity of the training and AAR facilities. An important but sometimes unappreciated role is played by the exercise planning and control staff, who create the conditions for an effective AAR by generating training objectives, scenario events and performance metrics conducive to review of the training event by the facilitator [2]. That is, without the support of the planning staff, the AAR process itself will not have the "raw materials" required for its conduct.

All users: the AAR is fundamentally an interactive and collaborative process, even in the case of a facilitator debriefing a trainee one-on-one (some attempts at fully automated AAR systems not withstanding). Though this may seem obvious, it does not hurt to underline the requirement for all the personnel involved in an AAR to be able to communicate and collaborate. This is particularly relevant (and particularly challenging) in distributed simulation task environments. Fittingly, the attention will be next turned to the challenges with the AAR process that drive the need for augmentation technologies.

7.2.2 User Limitations and Challenges

Regardless of who is performing the roles outlined above, most of the challenges with conducting AARs reside with the facilitation process, and the activities required to support the facilitation process. Notably, significant effort may be required to prepare an AAR, including make observations, collecting and analyzing data, and formatting data to present at the AAR itself. Accordingly, this section organizes user limitations and challenges thus: those related to the preparation of the AAR (data collection and analysis), those related to the review of performance and feedback during the AAR, and those related to the facilitation of group discussion and insights during the AAR. In general terms, the individual and collective cognitive processes involved in preparing and conducting the AAR come under increasing strain (and are increasingly prone to inaccuracy and failure) the more complex and distributed the task environment, and the larger set of trainees are. These are further described below.

7.2.2.1 Limitations at the Level of AAR Preparation: Performance Observation, Metrics and Analysis

Generally speaking, observing and recording events for recall is easier with individuals or small teams performing simple tasks in confined spaces, and becomes more challenging as the size and complexity of the team, task and environment grow. Live environments allow for direct observation but may be too large/complex for a single observer, in which case a new challenge arises with the team of observers needing to "compile" their observations into a coherent "common picture," which serves as a basis for the AAR. The team of observers then have to either manage the AAR as a team, or designate a single facilitator, and convey their common picture to this individual [6]. Instrumented live environments add the challenge of:

1) Collecting (the appropriate) additional data reliably;
2) Interpreting it (i.e., translating data into information and insight);
3) Combining it with "unaided" observations; and
4) Presenting/visualizing it in a useful way to the trainees.

Synthetic Environments (SE) create the challenge of the observers now needing a "presence" in the SE to observe and collect data [7]. That is, an SE provides increased opportunities for generating and capturing event data, while making it harder for those data to be observed and understood by human observers. Technological supports and automated measures/analytics become imperative and must be suited to the participants' knowledge and abilities. Distributed environments, which can be wholly SE or combined "Live-Virtual-Constructive" or LVC environments [8], present additional challenges of facilitating (human) observations, collecting and reconciling data from a variety of networked environments (which might have different timing systems, terrain databases, rendering engines, etc.) which need to be compiled into a "common picture" for the AAR.

Generating feedback comes down to interpreting relevant observations in light of relevant standards and objectives. The AAR literature indicates that providing feedback on a combination of successful and unsuccessful actions during an event leads to more effective AARs [3], [4]. The types of feedback that can be generated from observed performance will be highly constrained by the type of observations recorded in the first place. Simple observation of events, perhaps aided by hand-written notes, will at best support narrative, qualitative feedback. Collection of simple, objective metrics for well-defined (e.g., task completion times, personnel or unit positions, key event markers) are usually easily comparable to objective standards or cut-offs. However, these metrics must be planned so as to support the desired analysis (e.g., the events to be flagged must be pre-determined and observable). The quantitative analysis of performance in ill-defined (e.g., open-ended) tasks and team performance task metrics is notably more difficult, and often involves subjective judgment by human observers as to the success or adequacy of such performance [2], [6], [7]. The recording of large quantities of observations in "free-form" media formats (e.g., continuous video, audio, logs of all possible events in a simulation) may help address some of the limitations on human observation described above but will impose additional burden on the processing and interpretation of performance to generate feedback. These challenges are magnified as the size of the team and training environment grow, multiplying the sheer quantity of observations to be managed. Human limitations on integrating multiple observations from various players in a large space over a long period of time create significant challenges for "rolling up" observations from an exercise into over-arching themes, in order to generate targeted and relevant feedback for trainees in an AAR.

7.2.2.2 Challenges in AAR Conduct: Reviewing Events and Presenting Feedback to Trainees

The challenge here resides in managing the trade-off between quantity of feedback (avoiding overload) and its relevance (ensuring the feedback is meaningful with respect to training objectives). Predictably, this will be a function of the quantity and types of metrics collected, and the analysis to which they were subjected to generate feedback. One key aspect here is the amount of interpretation required by specific metrics [7].

For instance, position data might be fairly self-explanatory, especially for relatively few entities, whereas audio recordings of team communications might require some more interpretation. Another consideration is the degree to which the facilitator(s) wish to "prescribe" interpretations of events, rather than allow an understanding of the events to emerge from the group discussion; the former, more "top down" approach to facilitation may require more processing and analysis of the observations to extract objective metrics (which then need to be visualized in meaningful ways), whereas the latter, more "bottom up" may only require the bare essentials of specific events (timings, positions, perhaps video recordings of the event if available) to be presented for discussion. In both cases, the chosen facilitation style will impose constraints on how much and what kinds of post-processing of event data are required.

The presentation and visualization of feedback to trainees during the AAR requires careful selection and planning of data collection and an adequate consideration of performance standards and objectives prior to the AAR itself [6], [7]. As with the challenges related to performance observation and analysis, challenges with feedback presentation grow with the size and complexity the task and environment. Integrating multiple data sources and metrics from a large number of players so as to present a coherent and meaningful picture of the event as a whole involves both judgements by the AAR personnel about what observations are relevant and a significant effort in terms of processing the data in a practical way within the timeframe allocated to AAR preparation. AAR support personnel may only have a few hours to "pull it all together" for an AAR following a large collective training event. As mentioned above, the intended AAR format itself (top-down or bottom-up) also constrains the requirements for processing and visualizing performance feedback. Thus, the study briefly examines the group discussion facilitation process itself to conclude this discussion of user challenges with the AAR.

7.2.2.3 Challenges in AAR Conduct: Group Discussion Facilitation

AARs are typically conducted in-person, in a co-located space, sometimes the training environment itself, sometimes in a dedicated meeting room [2], [6]; AARs are sometimes conducted this way event when the event itself was conducted in a virtual environment with some participants in remote locations. Thus, the main challenges are those related to managing group discussions meant foster honest appraisal of performance and self-reflection [6], [7]. Some of these rest with the facilitator's interpersonal and didactic skills, whereas some challenges pertain to the dynamics of group discussions. The top-down vs bottom-up approaches to facilitation may require different technical solutions (and levels of technical preparation), and also different skills in eliciting insights from participants. Ensuring that individual trainees gets their say, and feel comfortable with honest discussions of collective performance, can be challenging in both large and small groups, for different reasons: large groups may lead to individual contributions getting "lost" or ignored, but may also temper strong personalities that could come to dominate a small group discussion. If the training event took place in a SE, a choice needs to be made about whether the AAR can be performed in the SE itself (i.e., all participants "re-entering" the training environment for the AAR), or in a dedicated meeting space (as with some live training events). Performing the AAR in the SE may make it easier to reproduce relevant event activities for discussion but may create challenges for the interaction between the facilitator and the trainees, and also for presenting performance data in formats that differ from those supported by the SE (e.g., graphs of task completion times, rather than replays of task events). If the event was conducted in a distributed SE, an in-person, co-located AAR may become highly impractical or impossible, in which case it will need to be conducted through virtual means (the distributed SE itself or a dedicated remote meeting system), which again may create challenges for effective interaction between the facilitator and trainees, and visualization of performance data and feedback. In particular, creating a common, shared understanding of the event review and feedback thereon might prove particularly challenging over distributed, virtual means.

7.3 AAR AUGMENTATION TECHNOLOGIES CONTEXT

Having laid out the human user challenges and limitations in the AAR process, the report now examines how these challenges can be addressed with augmentation technologies. It is noteworthy that, while some sort of augmentation technology (i.e., instrumentation at training ranges to track trainee performance and actions) has been a feature of AARs since the concept was formalized in the 1970s [7], the potential for (and actual use of) augmentation technologies to support user performance in AARs has increased dramatically with the development of SEs, and continues to evolve [2], [7]. Figure 7-2 provides an overview of AAR technologies and tools across the LVC spectrum of training environments.

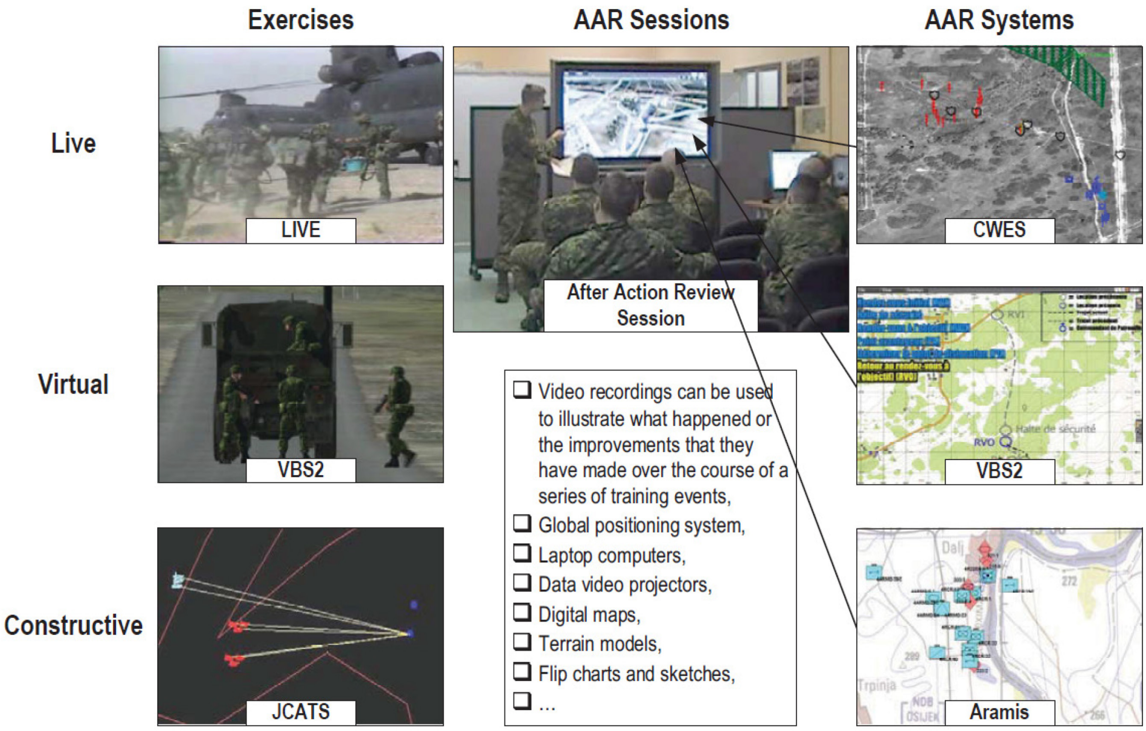


Figure 7-2: Sampling of AAR Technologies in Live, Virtual and Constructive Environments [6]. VBS3 = Virtual Battle Space 3 (Bohemia Inc.); JCATS = Joint Conflict and Tactical Simulation (US DoD); CWES = Canadian Weapons Effect Simulation (Canadian Army).

In the following sections the role of these technologies in supporting the three categories of user limitations and challenges described above will be discussed.

7.3.1 Augmentation Supports for AAR Preparation

The most relevant supports to the AAR preparation phase are performance tracking technologies and metrics. In live environments, instrumented ranges that track entity location and simulated weapons effects (e.g., via laser engagement systems) provide basic performance data for AARs. Over the years, these have been augmented with the ability for support personnel to timestamp significant events, record audio and video of the live events for further review, and user interfaces to access, manipulate and extract these data for AARs. Some degree of standardization for instrumented live training ranges, including their AAR supports, are provided in standards such as NATO's Urban Combat Advanced Training Technology (UCATT) architecture [9]. These instrumented ranges could be augmented with the use of performance monitoring technologies of the type discussed in Chapter 6 on performance monitoring in this report. However, these are not yet suited to field conditions with large teams. Finally, the activities by observer during live events may be augmented by digitizing their note-taking tools (e.g., using tablets to log observations or timestamp events) or providing them the ability to perform or trigger the recording of events of interest (e.g., via a camcorder or body camera). Note that such data collection capabilities borne by the observers would need to be integrated into the main data collection system, either during the event or afterwards.

In SEs (virtual or constructive), support personnel may also directly observe events and flag them for subsequent discussion. However, SEs create the opportunity for every event and entity state in an event to be recorded for future analysis, effectively providing a "built in" data capture capability [2], [7]. At a most basic level, this allows for easy "replaying" of actions during the event during an AAR [7]. This also allows automated performance measures of the type discussed in Chapter 4 on adaptive instructional systems in this report to be implemented for AARs. In both synthetic and live environments, the analysis of performance metrics must also consider performance standards and objectives. Therefore, technologies to augment the AAR preparation process much support the documentation of relevant training or performance standards and must include them in some way in the analysis (e.g., performance metrics must encode standards in their algorithms). It has been also noted that automated, objective measures are much easier to implement for well-defined tasks than open-ended tasks or team (collaboration) processes; for the latter, given the current state of technology, it may be advisable to simply replay the events and leave the interpretation to human SMEs and analysts. However, advances in Machine Learning (ML) and Data Mining (DM) hold the promise of being able to extract meaningful performance patterns and metrics from complex and/or unlabeled data sets (even potentially from audio and video recordings), thus adding to the arsenal of augmentation technologies for AARs in the future. Automation of event flagging is already a feature of a number of performance collection systems, both in live and Synthetic Environments (SE) [2], [6].

7.3.2 Augmentation Supports to AAR Conduct: Event Review and Visualization

The visualization of performance and event data relies largely on conventional display and data visualization technologies, namely: computer monitors and various data graph formats (e.g., bar graphs to compare task completion times). Such visualization methods are applicable regardless of the data source (i.e., regardless of whether the performance data were generated in a live or SE). The replay of event in "rich media" (e.g., video recordings or animated renderings of events captured from a SE) is also well-established and understood from a technical point of view. However, the effective visualization of performance metrics for AARs is still evolving [6], [10], especially with respect to avoiding information overload or visually integrating different data sources into coherent, meaningful insights. When the performance data of interest are highly abstracted from the events themselves (e.g., average task completion times of casualty rates in a team), "dashboard" type data presentations may be useful to provide "at a glance" overviews of multiple interrelated metrics.

While it is relatively simple (and common practice) to extract data from SEs for subsequent analysis and visualization via conventional methods as described above, another option with SEs is to visualize event data and metrics in the SE itself. Many SE environments have built-in "replay" functions that log the simulated events for future review, and support event flagging/timestamping. Additional capabilities for displaying data abstracted from the event logs themselves (e.g., average task parameters per team) may need to be built into the SE platform expressly to support AAR objectives, but this is a matter of design choice rather than technical limitations. This opens up additional, 3-dimensional and more interactive data visualizations, integrated with the SE in "smart" ways (e.g., graphs that can be viewed from many angles, integrated into a relevant virtual landmark). The leveraging of synthetic platforms for AAR visualization is of particular interest in distributed virtual environments, where the option of gathering participants around a conventional monitor may not be available and is a domain of active exploration [7], [10].

Finally, VR and AR technologies may be useful in augmenting AARs for live events. The data collected from live instrumented ranges could be used to generate virtual recreations that can be explored from multiple angles, unlike real video recordings, for instance. In addition, AR technologies used in a live environment could be used to visualize abstracted metrics or event virtual replays of key events overlaid over physical landmarks, providing additional context to these [6]. This could become a more relevant option for AARs as AR technology becomes a viable option for augmenting live training (e.g., providing a virtual inject into a live training event).

7.3.3 Augmentation Supports to AAR Conduct: Facilitation of Group Discussion

The most obvious role for augmentation technologies in supporting group discussions may be in the case of distributed SE events where the participants may not have the option of joining an in-person AAR. In this case, either the distributed SE itself (as suggested above) or separate means for remote/virtual meetings (e.g., current remote meeting platforms) could be used to facilitate a group discussion. Data/file sharing options in these platforms, as well as meeting facilitation features (e.g., "raise hand" functions, virtual whiteboards) and virtual meeting best practices can be used to ensure a successful distributed AAR (Table 7-2).

Numerous technologies and methods are also available to facilitate discussion and interaction in group in-person meetings (e.g., see discussion in Ref. [11]). Whiteboards (including virtual implementations) can support collaboration, while audience response technologies (e.g., so-called "clickers") can help manage group dynamics and ensure that large groups can provide responses to discussion points in an efficient manner. Audience response technologies in particular (which recently have evolved to be accessible via personal mobile devices), which add the possibility of anonymizing responses, may help more reluctant participants to contribute to the discussion (although the negative effects of being able to express oneself anonymously, often seen in social media, should be monitored in such cases). While these technologies may not obviously seem to fit the description of "augmentation technologies" as used elsewhere in this report, they are nonetheless technological supports that augment the "social space" that users are engaged in.

Finally, the use of objective performance measures themselves may be considered as an augmentation to the facilitator's managing of the group discussion, in particular regarding sensitivities around giving feedback. Objective measures mitigate bias and incorrect recall by observers and trainees, potentially increasing the perceived fairness of the feedback. It is noteworthy that in their meta-analysis, Keiser and Arthur [3] found that the use of objective measures for feedback increases the effectiveness of AARs, thereby supporting its value (note that none of the other augmentation technologies discussed in this report figure explicitly in their meta-analyses).

Table 7-2: Summary of User Challenges and Augmentation Technology Supports for Different AAR Phases.

AAR Phase	User Challenges	Augmentation Supports
AAR preparation	• Observing/filtering all relevant performance events/data, regardless of size/complexity/type of environment • Storing/recalling performance data • Analyzing or interpreting said data against performance standards • Fusing large volume of disparate data sources into meaningful metrics, possibly from multiple locations/systems (esp. for distributed SE)	• Live ranges instrumentation • Note-taking tools for OCTs (video, tablets) • Use of SEs • Logging of events and entity states in SEs • Automated performance metrics • Digital storage of performance data (observation, metrics) • ML/DM methods for discovering patterns in large or unlabeled performance data sets

AAR Phase	User Challenges	Augmentation Supports
AAR delivery: performance review	• Presenting performance data in meaningful ways • Avoiding information overload • Interpreting metrics with varying degrees of abstraction • Synthetizing an overall performance "narrative" from individual observations • Adapting data formats and visualizations to facilitation approaches	• Data visualization techniques (e.g., 2D, 3D graphs) • Data "dashboards" to synthesize data sets into meaningful summary metrics • Using virtual environments to enhance standard desktop computer visualizations • Using AR to provide data visualization in the original (live) training environments • Use of "rich media" (audio, video) for event replay
AAR delivery: group discussion facilitation	• Fostering open discussion and self-reflection in participants • Managing group dynamics in co-located environments • Fostering effective interaction in distributed environments • Maintaining a shared picture of performance events and feedback in distributed environments	• Using objective performance metrics to minimize perceived bias (increase perceived fairness?) • Audience participation technologies for large groups • Distributed (virtual) meeting technologies for distributed environments (or even to manage personalities) • Synthetic environments to enhance shared understanding (may also help with group dynamics)

7.4 STRENGTHS, WEAKNESSES, OPPORTUNITIES, THREATS (SWOT) ANALYSIS

Table 7-3: SWOT Analysis for AAR Augmentation Domain.

After Action Review (AAR)	
Strengths	**Opportunities**
• Well-established, understood, accepted process; AAR is entrenched as an essential part of the training process in most militaries [2] and is being increasingly used in other areas (e.g., medical training; [3], [4]) • Proven effectiveness: in recent meta-analyses examining wide range of AAR factors [3], [4], average effect sizes were found for AARs of d = 0.79 and d = 0.92 respectively (effect size varied by specific AAR context)	• "Lock down" events such as COVID-19 have forced big organizations and vendors to step up their game WRT facilitating distributed events and collaboration, as well as interest from militaries in investing in virtual/remote training solutions • AI and Machine Learning are advancing the potential for automated performance measurement • The increased use of SEs and LVC integration for training opens up opportunities for data capture and playback, possibly for (distributed) virtual collaboration spaces for AARs

After Action Review (AAR)	
Strengths (cont'd) • Flexible technological requirements; human-driven collaborative process that isn't intrinsically dependent on technology • Basic AAR technologies, especially those that rely on (and facilitate) human interpretation (e.g., playback of events in simple virtual environments or video from live events) are well-established [2], [6]. • Basic measures (positions, events times **and** labels) well-established • Synthetic Environments (SEs) and live instrumented ranges provide built-in data sources to support AARs [2], [7].	**Opportunities (cont'd)** • Improvements in performance monitoring technology and AR (coupled with AI for processing performance data) create opportunities for new, more comprehensive performance review in live (instrumented) training ranges. • The development of dedicated AAR standards, perhaps building on the NATO UCATT architecture [9], provide opportunities for advancing common AAR practices. • The fact that AAR requires human facilitation might provide an opportunity for "socializing" the use of automated performance measures in a context where they can prove their worth as supports, rather than final arbiters, for assessing exercises.
Weaknesses • Technology supports for SEs and distributed SEs are still sub-optimal **and** evolving [10]), especially for automated behavioral or team performance measures; can quickly become a bottleneck [2] • Planning the data collection for instrumented/SEs is a significant investment in time, effort and skill, not always done well, limiting value of AAR [2], [12] • Effectiveness of AAR is sensitive to (mis)match between techniques, technologies and event characteristics; e.g., can drop to d < 0.6 when individual AARs provided for team training [3] • Requires facilitators to be skilled in both facilitation and AAR technologies – often not the case (insufficient training for facilitators) [2], [6]. • Insufficient standardization of training/exercise data formats, as well as standardized AAR processes, to support standardization of AAR support technologies	**Threats** • The complexity, required effort, (poor) usability, immaturity and lack of evidence for the effectiveness of technologies in AAR can limit their acceptance and use • The collection and transmission of event data from live instrumented or networked SE events poses security/privacy risks, and may be vulnerable to cyberattacks • Development of automated measures for team/collective performance is likely to remain a challenge for many years in the future • Development of automated measures for team/collective performance is likely to remain a challenge for many years in the future • Integration of multiple data sources is challenging; integrating across multiple types of environments (e.g., LVC) may become a deterrent. • AAR in coalition exercise conditions raises the issue of interoperability for AAR processes and products. • Acceptance of automated measures to determine individual and team exercise performance might lag behind the maturity levels of the technology adoptions (regulatory and ethical issues); resistance, especially by instructor staff, may delay adoption of supportive AAR technologies.

7.4.1 Strengths

The AAR, in its many incarnations, is a well-established and accepted process. Many AAR technologies are available and well-understood. A number of meta-analyses, including the recent set in Refs. [3] and [4] attest to its effectiveness and track-record, in particular in training contexts. While these meta-analyses have not considered the range of augmentation technologies discussed in this chapter, they provide evidence for the effectiveness of objective measures in AARs in general. Also, given the widespread use of some kind of augmentation technology (either via instrumented ranges or measures built into SEs) in AARs in general, these meta-analyses provide some indirect indication of the value of technologies in AARs.

7.4.2 Weaknesses

It is generally accepted (though not systematically investigated) that technology supports for AARs in synthetic, and especially distributed synthetic, environments are still lacking and require more development [2], [7], [10]. While basic data collection is fairly well understood, especially for well-defined tasks and small teams, automated performance metrics and analysis of large sets of performance data for complex team tasks are still very much areas of development, as attested to in the chapters on adaptive instructional systems and performance monitoring. Also, the planning, technological resources and expertise required in supporting AARs with such measures can constitute a barrier to organizations making effective use of AAR technologies [2], which may lead to falling back on simpler methods driven mostly by unaided observations and notes. What little evidence about augmentation technologies can be gleaned from meta-analyses indicates that mismatches between facilitation formats (and my implication, the supporting technologies) and the task context (individual vs collective, type of task) can greatly reduce the effectiveness of AARs, underlining the potential for misuse of augmentation technologies in AARs.

7.4.3 Opportunities

Given the central role of data collection and analysis technologies for supporting AARs, the primary opportunities for technical advancements in AARs likely lies with progress being made in monitoring technologies and data analytics. The prospect of ML and DM techniques to uncover patterns in large sets of unlabeled data may reduce the burden on AAR and training event planners for planning data collection and interpreting the results. Developments in practical and affordable VR and AR display technologies will open up new opportunities for facilitating distributed AARs or augmenting AARs conducted in person, even directly the live training environment. A growing acceptance of, and need for, virtual and distributed means for conducting training (incorporating the adaptive training and performance monitoring systems discussed elsewhere in this report) will also help advance opportunities and developments for data collection, analysis and visualization for AARs.

7.4.4 Threats

As with the opportunities just discussed, the threats to AAR technology are likely related to challenges with data collection for training and performance, which have been previously discussed: data privacy, security, and the ethics of using machines (automation) to assess personnel with potentially career-altering consequences. Technical threats for advancing AARs technologies include the sheer burden of managing advanced, specialized technologies, especially when it comes to processing large data sets from training events. It is not uncommon for organizations to vastly underestimate the resources (technical and human) required to employ technically advanced, data-heavy methods; a problem that has been described as a "hidden technical debt" in Ref. [13]. So long as these "technical debts" remain hidden rather than acknowledged and properly addressed, they remain serious threats to the advancement of data-reliant augmentation technologies in AAR.

7.5 RECOMMENDED NEAR-TERM APPLICATIONS

Given the breadth of domains and task types for which AARs are performed, and the lack of systematic investigation of the effectiveness of technology use in AARs, it is difficult to provide recommendations on the best use of specific augmentation technologies. The clearest recommendation that can be made, based on Keiser and Arthur's meta-analyses [3], [4] is for the continued use of objective measures (over subjective recall) in AARs. However, even the range of objective measures alone defies simple description. In line with the discussions on adaptive instructional systems and performance monitoring in this report, it is possible to say with that performance (outcome) metrics for simple, well-defined tasks can be reliably employed in support of AARs, whereas metrics trying to capture user states (vs. performance) or performance in complex, open-ended tasks are less mature and are harder to use effectively. At the extreme, the use of ML or data mining approaches to deriving performance feedback are still very much the object of active investigation and are not recommended for career- or mission-critical training.

Keiser and Arthur's meta-analyses also indicate that mismatches between AAR approach, training task and training audience can lead to significant decreases in the effectiveness of AARs. For instance, they note that highly-structured ("top-down") AAR processes work well with trainees performing highly-structured tasks, in domains with high expectations of organizational structure (e.g., typical military). However, trainees performing more dynamic, open-ended tasks may benefit from less structured, more bottom-up AAR approaches. While this in itself does not argue for a specific technology application, it does lead us to recommend that training and AAR planers ensure that the augmentation technologies they use, in particular for visualizing feedback and facilitating discussions, be well aligned with the nature of the task and the organizational culture of the trainees. This may also argue somewhat in favor of trying to align the AAR environment with the training environment to the extent possible (i.e., performing the AAR in the training range of the SE where the event was conducted, conditions permitting), however this contention requires further investigation.

Finally, Keiser and Arthur [4] note that some AAR approaches use a type of "canned feedback" where trainees view a pre-recorded video or animation of other personnel (presumably SMEs) performing the task they were supposed to perform, rather than their own performance. The rationale behind such "canned" feedback is to create a "psychologically safe" space by singling out the performance of particular individuals in the AAR. In technology terms, this can be aligned with the feedback visualization technologies discussed above. Keiser and Arthur note that such "canned" feedback technologies consistently reduce the effectiveness of AARs and should be avoided. More generally, Keiser and Arthur note that the effectiveness of AAR methods aimed at fostering "psychological safety" (e.g., starting an AAR by asking participants how they feel) is inconclusive at best and are not recommended at this time. From a technology perspective, it is recommended avoiding the use of technologies for the sole purpose of promoting psychological safety (e.g., using audience response systems) without further investigation.

7.6 SCIENCE AND TECHNOLOGY INVESTMENTS

7.6.1 Current Research

The basics of AARs in live instrumented training ranges have been well-established for decades, and commercially-available weapons engagement systems typically include AAR functionalities as a matter of course. Accordingly, most efforts in this area are being dedicated to standardizing AAR methods (in particular data formats), often in the context of standardizing the larger live training system. A notable example is the NATO UCATT standard for live urban operations training, which includes an AAR component specification [9].

Similarly, as SE-based training systems have become more established, a need for including AAR capabilities within them was identified [7]; since then, most vendors of SE systems also include AAR modules or functions as a basic feature, as can be gleaned from a survey of commercially available training systems [2], [6]. Thus, research on advancing SE as training tools advances, advances in simulation technologies that can be used in AARs is also advanced. Prime examples are this are the ongoing development of automated performance metrics and visualization methods for complex data sets in virtual environments. The reader is referred to those sections for more discussion of these topics (in particular with regard to performance data).

One notable area of ongoing investigating in SE which merits further discussion is the development of tools for supporting distributed SE, and distributed AARs within them. The development of technologies and methods in support of the US Air Force Distributed Mission Training and Distributed Mission Operations were the focus of intense research activity in the early and mid-2000s, and involved coalition partners, including some NATO nations (Chapter 7 in Ref. [14]). The concept of distributed SE for training has also been adopted by other services (see other chapters in Ref. [14] for Army and Navy applications). While the research activity in developing distributed SEs seems to have subsided as the technology has become more accepted and implemented by training organizations, there is ongoing work in developing, and especially validating AAR methods for such distributed SEs. A notable recent example is described in Ref. [10].

7.6.2 Recommended Research Investments

A notable gap in the set of meta-analyses by Keiser and Arthur [3], [4] is the lack analysis specifically with respect to augmentation technologies in AARs. As discussed above, the only explicit mentions of AAR technologies in these two meta-analyses are to the broad classes of objective performance measures and "canned" feedback visualizations. The recent study by Jarrett et al. [10] comparing the effectiveness of co-located and distributed AARs stands out for its relative uniqueness: the authors themselves acknowledge the scarcity of studies examining the effectiveness of distributed AAR methods. Similarly, a recent meta-analysis of the training effectiveness of VR and AR technologies [15] did not include a single explicit example of these technologies used in AARs. Thus, an obvious first recommendation on investing in AAR research is to increase effort in studying the effectiveness of AAR augmentation technologies. Data on the effectiveness of specific technologies should be a key driver in directing future AAR research investments, and this information appears to be lacking, or at least not easily accessible to most. Given the large size of the AAR research literature (Keiser and Arthur reviewed 5,639 papers in [4]), it is likely that a meta-analytic approach with the explicit goal of identifying AAR technology types would yield fruitful insights. The effectiveness of distributed AARs might be one area in particular that could benefit from further empirical effort, as per Jarrett et al.

As discussed in the previous section on research efforts and the SWOT analysis above, an obvious next frontier for research is in further investigating automated performance metrics, in particular for complex team and open-ended tasks. ML and DM methods might be particularly relevant here. As these are already the focus of considerable R&D investment for a variety of reasons, the main recommendation with respect to AARs is to ensure that effort is invested in integrating the findings from these fields in AAR methods and systems.

Leveraging developments in visualization technologies, notably VR and AR, remain important for advancing AAR applications. The ongoing investigation of virtual visualization methods in distributed AARs should be sustained; as with AAR capabilities within instrumented live ranges, the time is propitious to invest effort in standardization efforts, especially in service of coalition inter-operability. These could leverage emerging standardization efforts for virtual technologies and content [16]. One "frontier" or emerging area worthy of further R&D effort is the application of AR technologies to AARs in live training environments, which may help to realize the long-aspired-to Live-Virtual-Constructive (LVC) construct, help up by some as a "holy grail" of training modernization [8].

7.7 REFERENCES

[1] Department of the Army (1993). Training Circular 25-20: A Leader's Guide To After-Action Reviews. Published by the Headquarters of the Department of the Army, Washington DC. https://www.hsdl.org/?view&did=775082 (Accessed 12 Jun 2022)

[2] Fournier, H. (2019). Land Vehicle Crew Training System (LVCTS) National Research Council (NRC) "Work Packages After Action Review." DRDC Contract Report, DRDC-RDDC-2019-C178, Defence Research and Development Canada.

[3] Keiser, N.L., and Arthur, W., Jr. (2021). "A Meta-Analysis of the Effectiveness of the After-Action Review (or Debrief) and Factors that Influence its Effectiveness." Journal of Applied Psychology, 106(7), pp. 1007-1032. DOI: /10.1037/apl0000821

[4] Keiser, N.L., and Arthur, W., Jr. (2022). "A Meta-Analysis of Task and Training Characteristics that Contribute to or Attenuate the Effectiveness of the After-Action Review (or Debrief)." Journal of Business and Psychology, 37. DOI: 10.1007/s10869-021-09784-x

[5] Villado, A.J., and Arthur, W., Jr. (2013). "The Comparative Effect of Subjective and Objective After-Action Reviews on Team Performance on a Complex Task." Journal of Applied Psychology, 98, pp. 514-528. DOI: /10.1037/a0031510

[6] Boivin, E., and Mokhtari, M. (2011). "A Canadian After-Action Review Process Improvement Roadmap." Defence Research and Development Canada Report, DRDC-VALCARTIER-SL-2011-397.

[7] Meliza, L.L., Goldberg, S.L., and Lampton, D.R. (2007). "After Action Review in Simulation-Based Training." Chapter 8 in "Virtual Environments for Intuitive Human-System Interaction – Human Factors Considerations in the Design, Use, and Evaluation of AMVE-Technology." NATO STO Technical Report RTO-TR-HFM-121-Part-II. NATO Science & Technology Organization, Neuilly-sur-Seine, France.

[8] Allen, G.W., Lutz, R., and Richbourg, R. (2010). "Live, Virtual, Constructive, Architecture Roadmap Implementation and Net-Centric Environment Implications." ITEA Journal, 31, pp. 355-364.

[9] NATO STO (2018). "Urban Combat Advanced Training Technology Architecture." NATO STO Technical Report STO-TR-MSG-098, NATO Science & Technology Organization, Neuilly-sur-Seine, France.

[10] Jarrett, S.M., Glaze, R.M., Schurig, I., Muñoz, G.J., Naber, A.M., McDonald, J.A., Bennett Jr., W., and Arthur Jr., W. 2016). "The Comparative Effectiveness of Distributed and Colocated Team After-Action Reviews." Human Performance, 29(5), pp. 408-427, DOI: 10.1080/08959285.2016.1208662

[11] Reinig B.A., and Shin, B. (2002). "The Dynamic Effects of Group Support Systems on Group Meetings." Journal of Management Information Systems, 19(2), pp. 303-325, DOI: 10.1080/07421222.2002.11045728

[12] Huddlestone, J.A., and Pike, J. (2012). "Team and Collective Training Needs Analysis (TCTNA): Identifying Training Requirements and Specifying Solutions." Journal of Battlefield Technology, 15(1), pp. 49-56.

[13] Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J., and Dennison, D. (2015). "Hidden Technical Debt in Machine Learning Systems." In Proceedings of Advances in Neural Information Processing Systems 28 (NIPS 2015).

[14] NATO STO (2005). "Virtual Environments for Intuitive Human-System Interaction – National Research Activities in Augmented, Mixed and Virtual Environments." NATO STO Technical Report RTO-TR-HFM-121-Part-I, NATO Science & Technology Organization, Neuilly-sur-Seine, France. DOI: 10.14339/RTO-TR-HFM-121-Part-I

[15] Kaplan, A.D., Cruit, J., Endsley, M., Beers, S.M., Sawyer, B.D., and Hancock, P.A. (2021). "The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis." Human Factors 63(4), pp. 706-726. DOI:10.1177/0018720820904229

[16] Wild, F., Perey, C., Hensen, B., and Klamma, R. (2020). IEEE Standard for Augmented Reality Learning Experience Models. In Proceedings of 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 1-3. DOI: 10.1109/TALE48869.2020.9368405.

Chapter 8 – INSIGHTS, LESSONS LEARNED, AND CONCLUSION

Jerzy Jarmasz and Benjamin Goldberg

8.1 SUMMARY OF AUGMENTATION DOMAIN CHAPTERS

8.1.1 General Overview

Our group's review of five performance dependent augmentation domains (Adaptive Instruction and Accelerated Readiness, Mission Preparation and Rehearsal, Real-time Support and Remote Control, Cognitive Monitoring and Optimization and After-Action Review) identified successful applications of augmentation tools and methods. The review also identified technical challenges and barriers to adoption across each domain, with attention to required research to drive the maturation of an augmentation tool or method. Looking at the similarities across each task and human performance context also reveals a number of themes and issues that are common across all application spaces.

With respect to the technologies themselves, it is apparent that the same technologies reoccur across many domains; for instance, Augmented Reality (AR) and performance monitoring technologies can be used in training as well as real-time job support, and Artificial Intelligence (AI) has applications in all augmentation domains. While the context of performance related tasks vary, the management and utilization of data supporting tasks follow similar input-process-output methods [1]. As well, there does not seem to be a technology that is limited to a single domain. That said, different technologies are more or less prominent in different domains as related to the performance timeline established in Chapter 2 of this report. For instance, virtual (synthetic) environments are most prominent in training (including AAR) and mission rehearsal applications, whereas AR seem to be more suited to real-time support applications based on its ability to interface with an operational space.

The technologies discussed in this report are also at varying levels of maturity. For instance, haptic and tactile interface technologies lag behind visual and auditory ones in fidelity, reliability and usability, however newer visual interfacing technologies still present significant challenges (e.g., cybersickness with VR displays [2], AR field of view, occlusion, rendering in bright conditions, etc.). AI-based methods (especially for adaptive training and performance metrics) work well for well-defined tasks, but much less so for open-ended tasks and interactions reliant on communication and natural language understanding; distributed and collective contexts still pose challenges for all applications, whether in live or synthetic environments. In more general terms, most augmentation technologies are still more suited to controlled environments (e.g., training centers, home station, physically constrained work environments) and their usability and reliability in non-permissive, rugged, less-secure operational environments remains work in progress.

Despite the variability in tasks and human performance requirements, the various augmentation domains share a key feature, they are all heavily data-driven and data-dependent. At a minimum, all of these applications require underlying data models (e.g., of task domains, operational environments, terrains, physiology, etc.) to track key variables of operator performance, and require in-the-moment context to drive augmentation functions aligned with task outcomes and human performance constructs. While these assumptions are recognized, the application of these data-driven methods is not trivial. Augmentation applications that provide (near) real-time feedback or support require data analytics methods and computational capacity across multi-modal sources of information. In addition, applications in distributed task environments increasingly require reliable, secure, high-speed and low latency data networks that can be deployed in austere or non-permissive environments. Advancements in these data-related technologies (e.g., 5G wireless networks, data mining methods, hybrid computing) promise to evolve augmentation domains to new levels; at the same time, current limitations with these technologies create significant bottlenecks for augmentation applications at a ubiquitous level across the enterprise. Ambitions to implement advanced data analytics and machine learning methods often incur significant but unanticipated "technical debts" in terms of the supplementary infrastructure required to support them [3].

In addition to the technical challenges, there are important factors and issues of a non-technical nature facing their broad adoption. Referencing the data requirements discussion above, some of the most important challenges are the security and privacy issues involved in collecting, storing and transmitting operator performance and state data, whether in training or operational settings. Significant ethical issues also exist. Specifically, this involves the use of data to support decisions that may affect a human-centered objective, regardless of whether its aligned to operational or personal goals. This includes data sourcing that impacts a user's career progression (e.g., training performance) or employment status (e.g., operational performance), or that may impact a user's experienced state (i.e., performance monitoring systems that may affect activity or sleep cycles). In addition, ubiquitous access to reliable augmentation tools and methods will create a user-centered dependency to meet desired performance objectives. Studying how technology becomes a cognitive performance crutch (e.g., navigation dependency using GPS) is important, as technology should not replace knowledge, skill and competency elements required to perform a task when a technology is no longer available. Regardless, differential access to augmentation technologies, whether due to organizational policy, logistics (e.g., insufficient systems for everyone), or usability (e.g., eye accommodation issues or cybersickness with head-mounted displays) also raise issues of equity and fairness in how augmentation technologies support operational requirements across organizational structures.

Many of the technical and non-technical issues still facing augmentation technologies would be alleviated by further definition of standards for system specifications, data requirements, employment and interoperability of these technologies. This is critical for future acquisition programs, as there is a need to establish implementable and reproducible engineering requirements that provide the level of detail and specification to optimize the utility of augmentation technology outside a controlled lab setting. Some efforts at standards developments are presented in the chapters above; while these efforts are important and promising, they face the challenge of requiring a broad set of end-user communities to adopt and use these technologies in the absence of validated standards, in order to generate the data and use cases required to mature and verify standards in the form of heuristics and best practices.

The issue of end-user adoption of augmentation technology can be seen as a bias between the enthusiasm and hype generated by the promises these technologies hold, and the actual track record of these tools and methods, based on the evidence available to assess their effectiveness. Sometimes, users rush to adopt an augmentation technology in the absence of clear benefits for its use (e.g., see the discussion of some technologies for AAR in Chapter 7). Using technologies before they are mature or have been proven to be effective can result in initial enthusiasm turning to discontent and even abandonment, a phenomenon that is vividly captured in the "hype cycle" concept developed by Gartner Inc. [4]. An example of such hype (and the accompanying concern it generates) that is very much in the public eye at the time of this writing is the so-called "Metaverse" construct, a ubiquitous virtual social environment being promoted by big commercial IT companies such as Facebook Inc. and others [5]. As depicted in Figure 8-1, the Metaverse is currently in its conception phase, with a likely build of inflated expectations of the technology industry does not balance with state of the art with the state of the possible. Of particular interest is the shift in emerging technologies between 2017 and 2022. In 2017, both Virtual Reality (VR) and Augmented Reality (AR) were defined, with VR in the Slope of Enlightenment phase where the boundaries of the technology space were well understood and being utilized accordingly; however, AR was nestled in the Trough of Disillusionment phase, where the capabilities available to the public did not meet expectations based on several overlapping factors [6]. In 2022, neither of those capabilities are referenced, but rather there is an emphasis on digitizing human and engineering processes, while building up the concept of a decentralized, ubiquitous Metaverse.

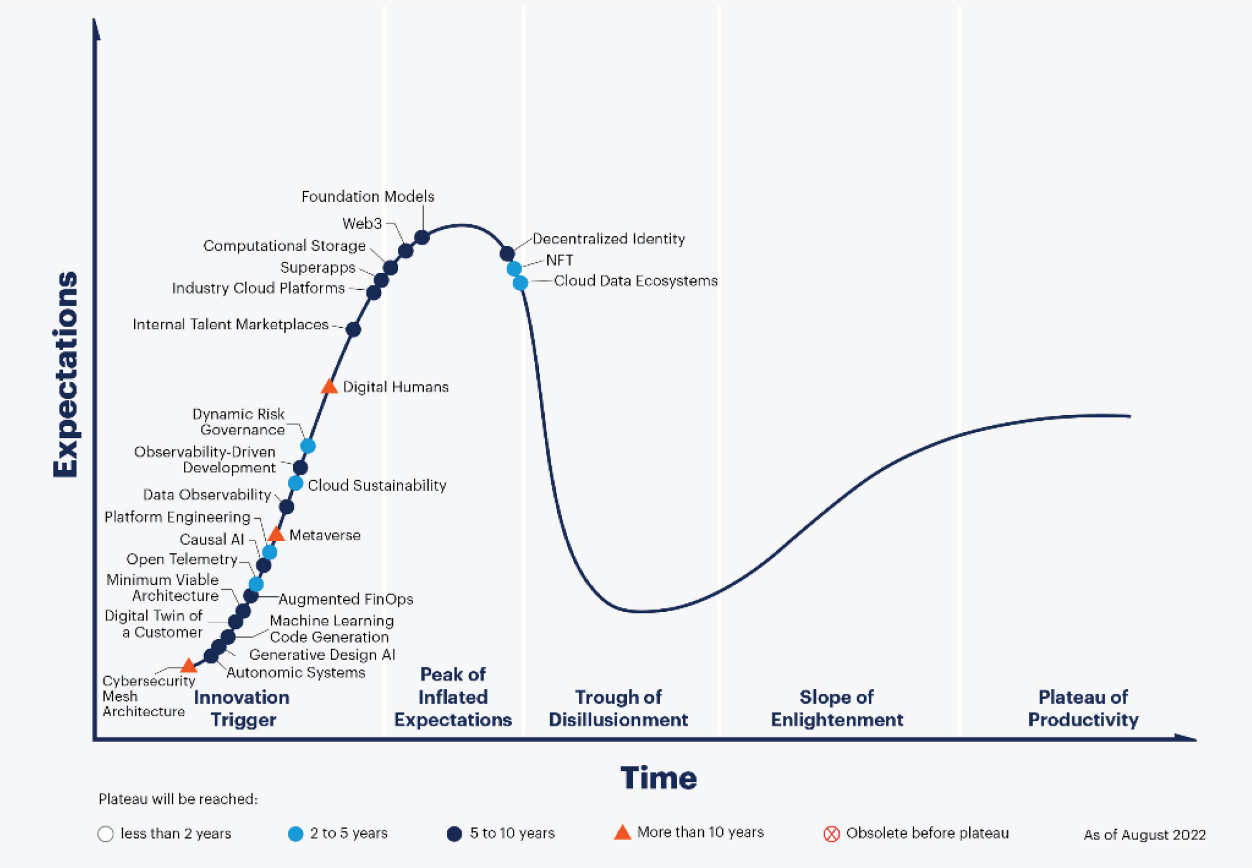


Figure 8-1: Gartner Hype Cycle for Emerging Technologies, 2022 [6].

In terms of broad application, successful training and on-the-job technology adoption depends on a multitude of factors [7] and requires a whole-life-cycle view to define these factors across all application elements [8]. Nevertheless, from the point of view of our RTG's activity, one key factor is the need to maintain ground truth awareness on the research to date focused on the effectiveness of these tools in applied settings. This requires a strategy to gather and update data on effectiveness studies with careful attention to the domains they're being studied within. This is critical, especially with respect to enabling evidence-based decision making around augmentation technology investments when the market is active with rapid enhancements to hardware, innovative data techniques to build insights from interaction, and research studies being executed in the dozen. Making definitive recommendations in a volatile technology sector requires a strategy to maintain Situational Awareness across high profile research communities and industry partners investing in applied research. As discussed in previous chapters, the evidence for the effectiveness of various augmentation technologies is variable and is very much an evolving field. A notable initiative in this regard is initiation of a new NATO research task group seeking to develop and expand databases of evidence and tools to support decisions on the use of eXtended Reality (XR) technologies based on evidence from the research literature and end-user input (NMSG ET-052, "Common Framework for the assessment of XR Technologies for Use in Training and Education" [9] and a proposed NMSG RTG "XR4T Portfolio of Evidence"). This is an important activity that can support reporting strategies for future NATO task groups. It'll be important to consider how an information technology infrastructure designed to track bodies of evidence across a field of research can support research task group market research and literature reviews focused on concurrency and relevancy. This will enable a decision support tool for research and acquisition communities, with careful attention to monitoring trends and outputs from study designs and human-subjects experiments.

Other opportunities for advancing R&D on augmentation technologies were identified earlier in this report. Two worth mentioning here are the increasing digitization of the work/operational space, and synergies to be gained from combining insights across augmentation domains. With respect to the ongoing digitization of work, this will create enabling conditions, namely with respect to data-related technologies, to support augmentation applications, as well as possibly engender more social (user) acceptance for data-driven tools such as augmentation technologies. A gap recognized by the group was a lack of common approaches for operationalizing the competency (i.e., knowledge, skill, and other abilities) requirements for an individual or team completing defined tasks or missions. Adopting standards for defining machine readable competency frameworks will be necessary to enable augmentation technology injects across an operational timeline from training to execution. While the domains may present varying challenges from a data capture, translation and system action standpoint, they will adhere to similar performance and task outcome constructs that disparate technologies must interoperate with. There are multiple alliance nations and services investing in this space, and some relevant standards being developed (e.g., the S6000T International specification for training analysis and design [10]), but there needs to be a concerted effort to mature a standard that can be applied across the training and operational contexts the augmentation technologies reported upon were designed to support.

Another interesting development in this regard, which was not within scope for our RTG, is the increasing use of digital twins, that is, real-time digital (virtual) representations of physical entities (platforms or humans) are intended to respond in nearly identical ways to the systems they represent for specific applications [11]. Digital twins are already used in maintenance and safety applications and may allow for the further development of augmentation technologies for real-time job support as well as just-in-time training or mission rehearsal. The development of digital twins for humans in particular is fraught with both promise and challenge and entails significant R&D effort over longer horizons than digital twins for platforms. While there are successes in modeling the physical properties of a digital twin in support of human factors and ergonomic design studies, accurately mimicking the cognitive and affective attributes still requires significant investigation at the basic and applied research level.

In terms of synergies across augmentation domain, one that stands out in particular to our RTG is advancements in performance optimization aligned to physiological state mediation. The earlier chapters above have already highlighted the potential usefulness of leveraging physiological performance monitoring to adapt an instructional system to optimize learning, or user interfaces to optimize performance. Chapter 6 of this report discusses non-invasive monitoring technologies focused on leveraging wakefulness and fatigue states to predict impact on performance. A further bound in such applications is suggested by the progression of basic research in neurostimulation in augmenting cognitive processes, for instance mental schema generation and encoding during early skill acquisition phases. This basic research could eventually be applied to augment and enhance human performance, alongside and in combination with other augmentation technologies. Currently, NATO HFM RTGs HFM-334 (Applying Neuroscience to Performance: From Rehabilitation to Human Cognitive Augmentation) and HFM-311 (Cognitive Neuroenhancement: Techniques and Technology) are examining the applicability of such basic research to the defence domain. Such research may eventually yield practical Brain-Computer Interfaces (BCI) for use in contexts where the usual manual or voice interfaces may not be practical due to task constraints [12]. BCI stands as a research domain at the "bleeding edge" of human performance augmentation whose application to defence and security remain to be investigated.

8.1.2 Synthesis of SWOT Analyses Across Domains

In previous chapters, each augmentation domain considered by the RTG was subjected to a Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis. Points recurring across the different SWOT analyses, echoing the discussion of common themes and challenges above, are summarized in Table 8-1 presenting a SWOT analysis of augmentation technologies in general.

Table 8-1: Commonalities Across Augmentation Domains SWOT Analysis.

SWOT Analysis of Commonalities Across Augmentation Domains	
Strengths • When used appropriately and within limitations, augmentation technologies have been shown to improve learning/preparedness/performance • Can overcome physical limitations/challenges in operational settings (e.g., distance, exposure to hazards during training) • Improved (and improving) computation capacity, IT infrastructure and interface technologies support the bulk of the augmentation technologies discussed here • Most of the relevant technologies are much more portable, unobtrusive and robust than before • Most of the augmentation applications discussed here have a proven track record going back decades, and some degree of empirical validation, at least for well-defined tasks in controlled or “permissive;” established “wins” over the years • Increased acceptance of augmentation technologies through increased digitization of work and everyday life • “Digitization” of data required for augmentation technologies can allow for more rigorous performance monitoring (learning and training) and the application of performance standards (at least for well-defined tasks)	**Opportunities** • Growing acceptable of digital/virtual technologies in the consumer sphere enhances acceptability in military applications • Improvement in collaboration technologies and distributed work/training spaces create opportunities for augmentation in collaborative contexts • Ongoing technological advances (e.g., 5G, ML, quantum computing, sensors) will continue to create augmentation opportunities • ML/data mining in particular continue to create new opportunities for fusing multiple data sources and extracting patterns from large and otherwise intractable data sets that can now be generated with augmentation technologies • Evolving real-world mission sets (e.g., Global Powers Competition) will require more flexibility and streamlining in training and missions, also creating opportunities • The increasing digitization of the work environment creates opportunities for both on-the-job supports and embedded training (data sources) • “Digital twins” as another opportunity to bridge real-life and synthetic/simulated data for a variety of augmentation applications (not truly addressed in chapters) • COVID as an example of a driver that makes various augmentation technologies more attractive
Weaknesses • Many augmentation applications still have low maturity, especially outside of well-defined tasks and controlled/permissive environments • Even when component technologies are advanced, integrated solutions still require improvements	**Threats** • Data/cyber security threats exist across all applications. Constrained task environments on local networks present less of a threat • Organizational culture and user-buy in regarding regulatory and privacy issues with user data (e.g., General Data Protection Regulation in European Union countries)

SWOT Analysis of Commonalities Across Augmentation Domains	
Weaknesses (cont'd)	**Threats (cont'd)**
• Many more advanced interface technologies still create human factors problems (e.g., cybersickness, usability of complex technologies, ergonomics of monitoring systems) • Standards still lacking/weak in for many applications • Systematic effectiveness/validation data still lacking outside of controlled or "laboratory" settings • Require organizational investments in increased, specialized technical expertise • Investment in institutional resources still considerable due to ecosystem required • Data generation and management (whether for content or collected from users) presents a significant burden (so-called "technical burden") • Relevant interface technologies are still predominantly visual and auditory (tactile/haptic weak) and do not support high degree of unstructured interactivity (e.g., NLP still does not support free-form dialogue well) • Requires investment in instrumented training and operational environments • High-speed / capacity / low latency IT wireless networks still difficult to deploy in many training and operational settings; current infrastructure not up to task for real-life or live settings • Integration of technologies, both within a domain and across domains (e.g., seamless transition from training to operations) still lacking	• Ethical challenges with applying augmentation (esp. automation) technologies to performance or training evaluations that have consequences on career progression • "Fragility" and usability challenges of technology in non-permissive environments (or live environments for training); maintenance burden requiring proactive full life-cycle management • Pace of evolution of technology and operational concepts outstrip speed of development of instructional or mission support processes and systems • High upfront (content development, data management, infrastructure) costs can weigh heavily in a 'return on investment' analysis, which in turn can threaten acceptance • General lag between operational and training applications and institutional policies • Inadequate attention to interoperability issues in joint or coalition settings can undermine use in collective settings • Inappropriate trust in automation (both over- and under-reliance) can severely undermine the use of augmentation technologies • Potential to erosion of critical knowledge and skills through (over)use of on-the-job support technologies (so-called "cognitive crutch" phenomenon)

8.2 RECOMMENDATIONS AND CONSIDERATIONS ON THE APPLICATION OF AUGMENTATION TECHNOLOGIES

On the basis of the analyses and considerations above, we offer the following considerations on best practices and recommendations for applying the current state of the art in augmentation technologies. These include areas where augmentation technologies are proving to be effective, and areas that require further improvement.

8.2.1 Best Practices in Applying Augmentation Technologies

- Application of augmentation technologies must continue to be guided by human factors considerations and empirical evidence for their effectiveness; technologies supporting operational or job requirements should be applied based on mission-function-task analyses and user-centered design requirements, and technologies supporting training should additionally be guided by instructional design principles.
- Adaptive training technologies are well-suited (i.e., have a proven track record) for individual and small team training, especially for well-defined task contexts.
- Applications leveraging mainly visual information and cognitive/procedural tasks with limited time constraints on task execution are the best applications for current tech base, especially virtual for immersive adaptive training practice.
- Virtual/remote presence (including distributed simulation) technologies are mature enough for “office” type collaborative work (as many of us have experienced during the COVID-19 pandemic), and for distributed training for procedural skills and mission sets. Careful consideration of fidelity requirements is important in applying these to meet training and/or on-the-job objectives.
- AR and MR are proving to be effective for certain types of training (e.g., maintenance and medical skills) and real-time support at home base or in generally “permissive” settings.
- Synthetic and instrumented live environments have seen widespread adoption for collective training and mission rehearsal; they present a ready source of training and performance data that should be leveraged to accelerate the maturation of augmentation applications for performance monitoring, assessment and AARs. In turn, the integration of such augmentation technologies increases the value proposition for synthetic and instrumented live environments. The evolution and increased adoption of synthetic and instrumented live environments should be advanced in combination with performance data-based augmentation technologies directly aligned with human performance constructs.
- To the extent possible, augmentation technology applications should make use of existing or emerging standards for performance and training data (e.g., AR Learning Experience Model aka ARLEM [13]; eXperience Application Programming Interface aka xAPI [14]). The continued development and refinement of these standards is crucial to the success of augmentation technologies and should continue to be advanced.

8.2.2 Caveats and Areas that Require Further Research

- The application of augmentation technologies in general is still hampered by the lack of a formalized data strategy to capture, contextualize, and operationalize data in support of augmentation methods. Continued efforts on developing data standards and processes supporting both system and human performance will help address this.
- Reliable evidence on the effectiveness of augmentation technologies applied to specific contexts is still lacking, hampering evidence-based decision making about their use. Efforts on collecting effectiveness data and conducting meta-analyses (such as the ones reported in specific performance domain chapters) should be continued.

- Collective and distributed task contexts still pose challenges, especially in synthetic and mixed (LVC) task environments. Integration of various environments across networks remains incomplete and still requires considerable effort in addressing classification and security issues. Team performance monitoring, either in real-time or for collective AARs, does not yet scale up well to large collectives. We note, however, that this challenge is being addressed by initiatives such as the US Army's Synthetic Training Environment (STE) initiative, whose goal is to establish an enterprise level Training Management Tool that connects to and consumes data from the "edge" devices that soldiers use during a distributed collective exercise.

- Haptic and tactile modalities are still not mature enough to reliably support human performance or training that require interaction with or feedback from physical objects. Virtual mediation of interaction with physical objects remains work in progress, though the relevant technologies continue to advance rapidly. Augmented or mixed reality implementations that leverage physical objects in the real world are still typically more feasible and usable than fully virtual implementations in such cases.

- Complex, open-ended tasks, especially those requiring advanced tactics, or interpersonal interactions with open-ended dialogue and outcomes, also not yet well-suited for applications involving automated performance metrics and automated training support (e.g., adaptive instruction systems, synthetic teammates), despite significant recent advances in text-based natural language processing. Complex behaviors (whether system or human) still typically require the intervention of role-players or exercise controllers.

- Content generation and performance analysis remains labor-intensive and computationally costly. While data analytics and machine learning methods hold great promise for automating these processes, the current state of the art is such that users of synthetic environments or augmentation technologies that capture large amounts of user performance must still be prepared for significant investments in expert human labor and computational power.

- Accordingly, the increasing use of augmentation technologies requires increasing technical expertise for military organizations; if the development of this expertise is not prioritized or invested in, the application of many/most of the augmentation possibilities discussed here will remain challenging and limited. The AI and analytics technologies being advanced to address some of the reliance on human expertise will themselves require new competencies and organizational culture for effectively working with automated systems and AI as "teammates."

- Wireless/cellular data limitations (bandwidth, latency, reliability) still limit many real-world uses of augmentation technologies, especially for real-time support and leveraging of performance data in live instrumented ranges. While advances in cellular data standards (e.g., 5G) are starting to address many of the technical issues, the security challenges and vulnerability of wireless systems (especially commercial) in non-permissive environments will require significant continued effort before fully connected systems can be reliably deployed in operations.

8.3 LONGER TERM R&D INVESTMENT RECOMMENDATIONS

Looking past the current state of the art, our RTG sees the need to invest in the following R&D domains:

- Continue to support Human Factors studies on the effectiveness and usability of augmentation technologies. A particularly important topic in this respect is a more systematic examination of organizational (including individual personnel) barriers to the effective implementation, use and management of augmentation technology capabilities. Some steps in this regard have been taken in Canada [15] and the Five Eyes [16]. Extending this approach to the NATO sphere for specific application domains, generally being more deliberate about full life-cycle human-systems integration practices, may be of benefit.

- Haptics technologies need to be further advanced in order to better support physical (tactile, force) feedback cues and provide more physical affordances in virtual environments. Application domains that would particularly benefit from further investment in where haptics interfaces (especially when coupled with AR or VR systems) include maintenance and medical (both training and on-the-job).
- Visual AR technologies need to be further advanced and especially in-theatre applications in relation to optical see-through display brightness and field-of-view properties along with reduced optical stray light leakage causing higher signature of soldiers wearing the devices to be more detectable from e.g., the naked eye, night vision googles and ISTAR sensors.
- The “hardening” of augmentation techs for military contexts is a critical aspect of deploying these technologies; both technical and HF work in this domain needs to continue. In addition to the improvements to visibility in operational lighting conditions for AR mentioned above, other general “ruggedization” improvements would include protection from environmental conditions (water, dust, extreme temperatures, deep water), protection from rough handling, improved ease of repairability. Such ruggedization should be accomplished in the context of human-systems integration efforts of augmentation technologies mentioned above.
- The criticality of data generation and analytics across the augmentation domains implies a need to continue to invest in R&D on automated content generation (aka generative data analytics) and data analytic methods for operational and training data. ML and data mining as techniques to enable performance monitoring and fusing of performance data streams in complex and team/collective tasks environments (for adaptive training, mission rehearsal, performance optimization and AAR) should be a particular focus. As stressed above, such efforts should be guided by human and performance and mission requirements.
- Regarding AI, research on an interoperable data strategy is imperative. Context rich data sources are required to drive model development, ML and data mining processes. There is no clear understanding on what data needs to be captured and stored to truly drive human performance ML processes. Storing raw data is a start, but it is not of much use without overlayed context to understand the relationship of that data towards a task or individual contributor.
- The ethical and security aspects of augmentation technologies in the military occupations need to continue to be examined; in particular, countermeasures to vulnerabilities of augmentation, and especially data-driven systems need to be considered to avoid disruption of Allied training and operational capabilities. Specific areas of concern include protecting critical information on personnel from being inappropriately accessed and preventing bad data or mis-information being injected into systems by bad actors.
- Continue to advance human performance-centered data strategies, standards and models.
- Investigate the organizational and individual competencies that will be required for the “future of work” with more automated, AI-driven job aids and support systems (possibly taking the role currently assigned to human teammates).
- The specific issue of how augmentation and automation change workflow and affect/moderate operator capacities needs to be examined more closely as they may pose long-term threats to operational effectiveness. Two topics of particular note here are preparing personnel for human-machine teaming and the “cognitive crutch” issue, which can be expressed as the potential for skill and knowledge decay through using augmentation.
- Continue to advance human modeling (performance, cognition, physical); human (and team) digital twins.

8.4 THE EVOLVING AUGMENTATION TECH SPACE AND THE CHALLENGES OF REPORTING STRATEGIES FOR FUTURE RTGS (MOSTLY COPIED FROM EARLY REPORT OUTLINE)

One clear theme that emerges from the above discussions is the continuously increasing pace of technology development, and the challenge of reliably aligning that development to human performance requirements in an evidence-based way. Whereas many technological advances in the augmentation space used to be driven by military requirements, they are increasingly impelled by consumer (entertainment, gaming) and commercial (business intelligence, commercial logistics and aviation) requirements. In many cases, military technology applications are in a "catch-up" posture relative to these commercial drivers. Furthermore, the operational and security environment is itself rapidly evolving at the time of writing, as contemporary events are showing; the resulting changes in mission and performance requirements also complicate the study of technology applications for supporting training and human performance. As a result, the investigation of military technology applications seems to be increasingly concerned with adapting commercial or consumer technologies to military context rather than "staying ahead of the curve." Consequently, our RTG believes that this points to a fundamental challenge for research groups such as NATO RTGs tasked with documenting and assessing the state of the art in many technological spheres. During the lifespan of this RTG alone, the augmentation technology scene changed considerably, making some initial assumptions obsolete, and creating challenges for fixing a definitive scope and reporting plan for the activity. The typical (traditional?) NATO RTG report development and sharing model cannot keep up with the dynamic technology scene, and recommendations generated that way tend to become rapidly outpaced by the community at large.

The RTG explored and discussed many options for making its findings relevant and current during its various meetings. All of the group's in-person meetings (following the kick-off) included plans for engagement and attempts at data collection with scientific, military and industry SMEs, which were executed with varying degrees of success (see Annex B for list of extramural activities). Attempts to systematically elicit and structure SME assessments of augmentation technologies were made using an "instantiation card" format based on structured foresight methodologies [17], as well as sponsoring a Special Event session at the 2019 Interservice / Industry Training Simulation and Education Conference (I/ITSEC) with spotlight demonstrations on state-of-the-art augmentation tools and methods. Later, an initiative on "crowdsourcing" experiences and assessments of augmentation technologies from end-users on a continuous basis using web-based technologies at the UK MOD was discussed and considered for adaptation by the RTG [18] Finally, the concept of soliciting complementary input on technology advances from Industry using a web-based survey was considered and actively discussed with NATO STO and activities sponsored through the NATO Industrial Advisory Group [19]. In the end, while the RTG was successful in engaging various user and SME groups, feasibility issues and the COVID-19 pandemic prevented these various plans for developing a more dynamic and "living" record of the RTG's activities from coming to fruition.

Nevertheless, our RTG proposes that future NATO task groups or studies examining dynamic technology areas such as augmentation technologies need to consider more dynamic and responsive formats for capturing information on their domain and presenting their findings. An example of such an approach would be to develop a web-based (rather than report-based) reporting framework. Leveraging the data input and content management capabilities possible with web-based technologies, a web-based framework could also be used to source data from a broader sampling of subject matter experts and requirements holders (both military and technical). It would also enable frequent (if not real-time) updating and curating of findings, ensuring the product retains relevance and currency over time. Such an endeavor would require certain constraints to ensure quality control, appropriate access by relevant parties, and appropriate focus of effort. For instance, study topics or themes requiring iterative, dynamic updates would have to be identified at the time of definition as suitable for a dynamic reporting format. Timelines and criteria for updating or re-contextualizing a study's findings would have to be defined. NATO STO infrastructure and support would be critical for implementing and managing a suitable web-based product. A critical consideration is the appropriate definition of criteria by which to frame technology applications investigations and assessments, namely ensuring a balance between operational end-user requirements and aspirations vs evidence-based evaluation of effectiveness and best practices.

8.5 CONCLUDING THOUGHTS

It is important to succinctly wrap up this final report with some final take-away perspectives. For the sake of brevity, these items are presented in a bulleted format, with a goal of summarizing the RTG objectives, the approach defined, the activities executed, the challenges endured, and the outcomes generated from this RTG. We are in an exciting time for technological advancement. Understanding the challenges and lessons learned from this activity will be critical to establishing a more formal and data-driven approach to tracking the maturation of human-centered and performance-driven technological tools and methods.

- In order to meet its stated objective of assessing augmentation technologies, the RTG needed to define a scope for the technologies considered, and a framework (five performance domains contextualized along an operational timeline) against which to assess these technologies. Defining the scope and framework proved challenging, reflecting both the multiplicity of perspectives on the concept of augmentation technologies, and the extremely dynamic nature of the technological space.
- Technologies in each domain were analyzed with respect to human performance requirements in military contexts, using a Strength-Weaknesses-Opportunities-Threats (SWOT) analysis. These analyses enabled the RTG to apply findings from recent research literature to relevant performance requirements.
- Synthesizing the analyses across the five domains, we performed a global SWOT for the augmentation technologies within the RTG's scope and extracted cross-cutting themes. Chief among these is the requirement for developing comprehensive data strategies aligned with human performance requirement to support the development, application, maintenance and evolution of augmentation technologies.
- Developing such data strategies requires work on standards (technical and human performance), addressing security (incl. classification), privacy and ethical issues, and overcoming the technical challenges of providing practical connectivity in non-permissive environments. Many efforts, within NATO and elsewhere, are underway to address these issues, but it will be critical for research and operational communities across our nations to sustain and coordinate these efforts and avoid their fragmentation.
- We further provided an overview of current "best practices" and recommended areas for future investment, based on the performance domain analyses and the cross-cutting themes extracted therefrom. Unsurprisingly, many of these recommendations involved the limitations of current data-driven capabilities and the need to coordinate their rapidly evolving advancement with standards aligned to operational human performance requirements.
- The challenges with defining a scope and keeping pace with rapid evolution in technology and operational mission sets (given the evolving security environment) led the RTG to re-consider and challenge the very process and reporting construct of the typical RTG (an example of which is this very report).
- Looking to the future, we recommend that NATO STO consider (and support research tasks or studies on) more dynamic (e.g., updateable) and accessible formats for conducting studies and reporting their findings, such as web-based formats, especially for domains such as augmentation technologies with a high degree of dynamicity and change.

8.6 REFERENCES

[1] Gao, Y., and Janssen, M. (2020). "Generating Value from Government Data Using AI: An Exploratory Study." In G. Viale Pereira, M. Janssen, H. Lee, I. Lindgren, M. P. Rodríguez Bolívar, H. J. Scholl, and A. Zuiderwijk (Eds.), Electronic Government – 19th IFIP WG 8.5 International Conference, EGOV 2020, Proceedings, pp. 319-331. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Vol. 12219 LNCS). Springer. DOI: /10.1007/978-3-030-57599-1_24

[2] NATO STO (2021). "Guidelines for Mitigating Cybersickness in Virtual Reality Systems." NATO STO Technical Report STO-TR-HFM-MSG-323. NATO Science & Technology Organization, Neuilly-sur-Seine, France.

[3] Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J., and Dennison, D. (2015). "Hidden Technical Debt in Machine Learning Systems." In Proceedings of Advances in Neural Information Processing Systems 28 (NIPS 2015).

[4] Kondo, Y., Asatani, K., and Sakata, I. (August 2022). "Evaluating Emerging Technologies on the Gartner Hype Cycle by Network Analysis: A Display Technology Case Study." In 2022 Portland International Conference on Management of Engineering and Technology (PICMET), pp. 1-11. IEEE.

[5] Kraus, S., Kanbach, D.K., Krysta, P.M., Steinhoff, M.M., and Tomini, N. (2022). "Facebook and the Creation of the Metaverse: Radical Business Model Innovation or Incremental Transformation?" International Journal of Entrepreneurial Behavior & Research 28(9).

[6] Gartner. (2022). "Gartner Identifies Key Emerging Technologies Expanding Immersive Experiences, Accelerating AI Automation and Optimizing Technologist Delivery." Gartner. https://www.gartner.com/en/newsroom/press-releases/2022-08-10-gartner-identifies-key-emerging-technologies-expanding-immersive-experiences-accelerating-ai-automation-and-optimizing-technologist-delivery

[7] Emond, B., and Durand, G. (2022). "Support to Training and Education Modernization: Training Technologies Adoption." DRDC Contract Report, DRDC-RDDC-2022-C325, Defence Research and Development Canada

[8] Parkinson, G. (2022). "Training Technology Support Tool Development: Final Report." DRDC Contract Report, DRDC-RDDC-2022-C295, Defence Research and Development Canada.

[9] NATO STO (2022). "Common Framework for the Assessment of XR Technologies for Use in Training and Education." NATO STO Technical Report STO-TR-MSG-ET-052. NATO Science & Technology Organization, Neuilly-sur-Seine, France.

[10] AeroSpace and Defense Industries Association of Europe (2021). International Specification for Training Analysis and Design S6000T-B6865-06000-00 Issue No. 2.0. http://www.s6000t.org/docs/S6000T%20Issue%202.0.pdf (Accessed 29 Nov 2022).

[11] Mihai, S., Yaqoob, M., Hung, D.V., Davis, W., Towakel, P., Raza, M. et al. (2022). "Digital Twins: A Survey on Enabling Technologies, Challenges, Trends and Future Prospects." IEEE Communications Surveys & Tutorials.

[12] Brunyé, T.T., Brou, R., Doty, T.J., Gregory, F.D., Hussey, E. K., Lieberman, H. R. et al. (2020). "A Review of US Army Research Contributing to Cognitive Enhancement in Military Contexts." Journal of Cognitive Enhancement, 4(4), pp. 453-468.

[13] Wild, F., Perey, C., Hensen, B., and Klamma, R. (December 2020). "IEEE Standard for Augmented Reality Learning Experience Models." In 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), pp. 1-3. IEEE.

[14] Walcutt, J.J., and Schatz, S. (2019). Modernizing Learning: Building the Future Learning Ecosystem. Advanced Distributed Learning Initiative.

[15] Jarmasz, J., and Martin, B. (2019). "Distributed Simulation for Training: Promises, Barriers and Pathways." In Proceedings of STO NATO Modelling and Simulation Group (NMSG) Symposium 159 on "Multinational Interoperability: Agility for Military Training and Operational Applications, Innovation in Enterprise Level Consortiums and M&S Technology Development," 11 – 12 October 2018, Ottawa, Canada, AC/323(MSG-159)TP/873.

[16] Schatz, S., Rankin, P.D.C.F.K., Huff, P.D.K., O'Brien, C.P., and Andrejevic, B. (2019). "Twenty-Five Emerging Trends in Learning and their Implications for Military Partners: An International Study." In Proceedings of the 2019 I/ITSEC Conference, Orlando, FL USA.

[17] Adlakha-Hutcheon, G., and Jarmasz, J. (2019). "Fostering Innovation in Training and Education through a Novel Technology Assessment Approach." DRDC Scientific Letter DRDC-RDDC-2019-L308, Defence Research and Development Canada.

[18] Olonilua, M., Black, S., Forrest, E., and Bayerl, P.S. (2020). "The Assessment and Generation of Evidence for XR Approaches in Training." NATO STO Meetings Proceedings Paper, STO-MP-MSG-177-12. NATO Science & Technology Organization, Neuilly-sur-Seine, France.

[19] Proietti, P., Abate, A. and Sab-Möbus, B. (2021). "Assessment of Human Augmentation Technologies for Exploitation in Battlefield." NATO Industrial Advisory Group (NIAG) Study Group 253.

Annex A – INSTANTIATION CARD (ICARD) TEMPLATE AND EXAMPLES

During an HFM-297 bi-annual meeting hosted by Defence Research and Development Canada (DRDC) in Toronto, we conducted a team exercise to develop a framework for reporting metrics and insights linked to technologies of interest. We leveraged a Technology Instantiation Assessment method [1], which focuses on the development and validation of “Instantiation Cards” (i.e., iCards). An iCard is used to summarize available information on specific use cases with alignment to empirical evidence and those parties and players engaged in the field. The goal is to establish a reporting format that is easy to consume and provides best practice guidelines on what tasks a technology is well-suited for, and what research is required to mature its utility to other domains of interest.

For a full overview on how this methodology was applied against the technology reporting objectives of HFM-297, please see Adlakha-Hutcheon and Jarmasz [1]. For this annex, we are providing the iCard templates that were used to guide this workshop activity, followed by example iCards that were generated by the task group (reproduced with permission from Ref. [1]).

Please note the iCards shown in Figure A-1 to Figure A-4 are provided as a record of the RTGs through process on characterizing and assessing augmentation technologies and are not to be construed as definitive statements on these, nor as endorsements of the products or organizations mentioned therein.

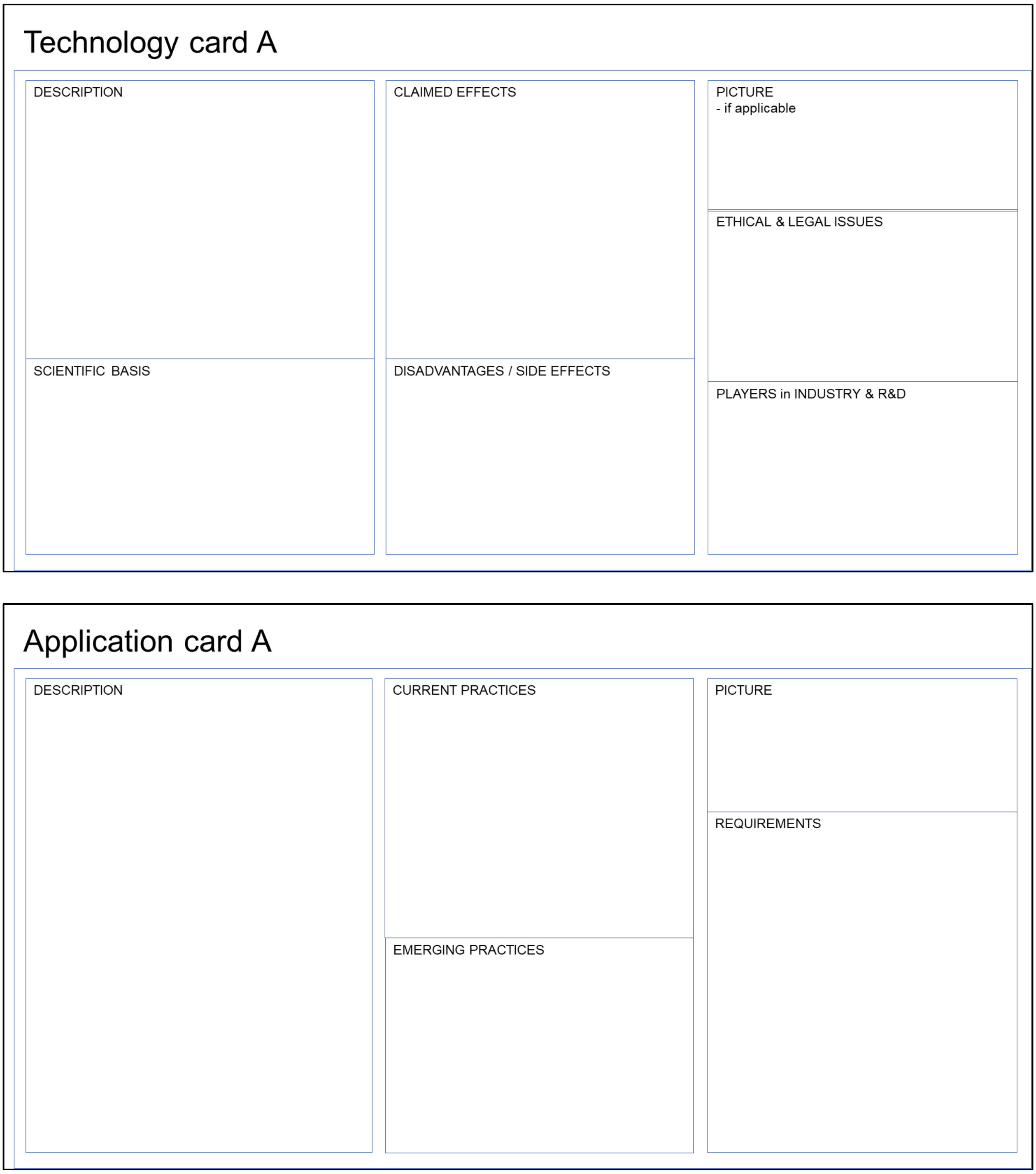


Figure A-1: (Top) Technology iCard Template – Used to Establish an Aggregated Representation of a Technology and the Underlying Evidence Showing its Effectiveness Across all Applications; (Bottom) Application iCard Template – Used to Represent a Focused Perspective on a Technology Applied Within the Boundaries of Defined Tasks, Conditions and Standards.

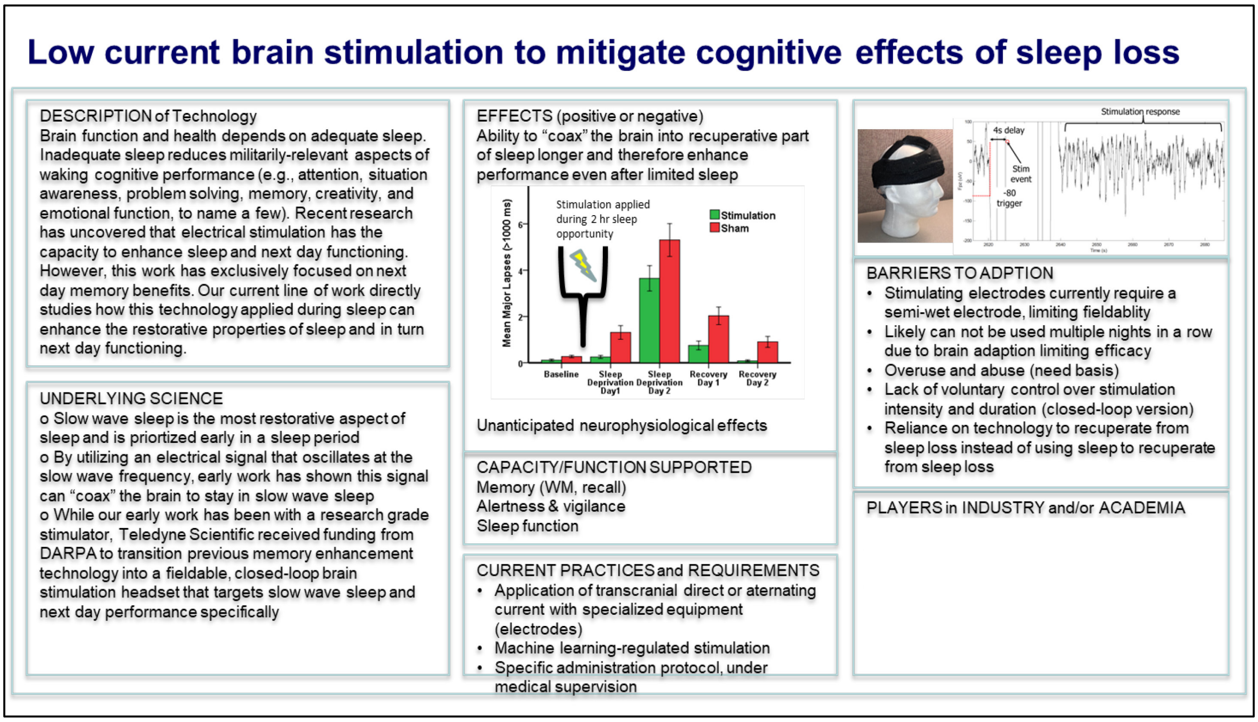


Figure A-2: iCard Example on Low Current Brain Stimulation Developed During the HFM-297 Meeting at DRDC Toronto.

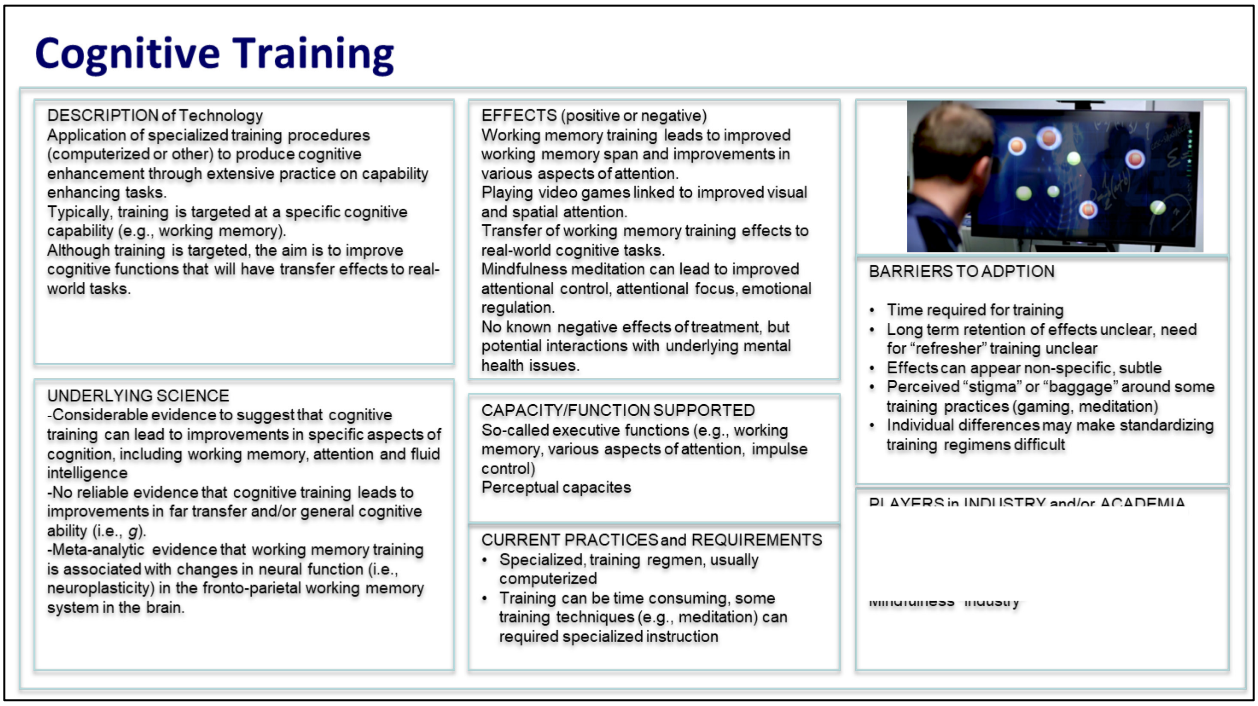


Figure A-3: iCard Example on Cognitive Training Developed During the HFM-297 Meeting at DRDC Toronto.

Augmented Training through Adaptive Instructional Systems

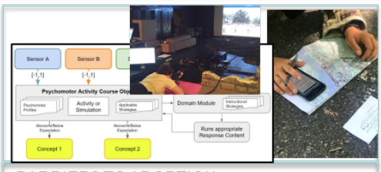
DESCRIPTION of Technology
Adaptive Instructional Systems (AIS) are derived applications that accelerate knowledge and skill acquisition through functions that target personalization and balanced feedback and coaching. AISs augment interaction to the needs of each individual learner or team where actions by the tutor are determined by policies designed to optimize learning, performance, retention, and transfer of skills to the operational setting.

UNDERLYING SCIENCE (cite references and/or patents)
o AIS are products of interdiscplenary research across learning sciences, AI, computer science, psychology, and instructional design
o Original AISs are based on foundational cognitve frameworks used to drive computational representations of of how knowledge and skill is formalized in memory
o Based on constructive learning theory and learning by doing philosophy
o Defined as critical technology need for future training strategies across all U.S. DoD services

EFFECTS (positive or negative)
- Accelerated & more effective skill acquisition (indiv & team; gains of .7 SD above classroom training – meta-analysis)
- Supports instructor process, individualized trg
- More objective assessment of training
- Logistical challenges/burden for instructional system
- Potential for training systems to be "gamed" by trainees
- Potential for mis-diagnosis of learner performance/needs

CAPACITY/FUNCTION SUPPORTED
- Individual Skill acquisition (acceleration)
- Teamwork skills
- Self-regulated training model
- Objectives-driven assessments
- Diagnosis of learning needs

CURRENT PRACTICES and REQUIREMENTS
- Extensive connectivity (IoT)
- Requirements for standards & certification processes
- Evolving civil standards (xAPI, adaptive instructional systems WG, etc.)
- Competency frameworks (e.g. NATO TG IT&ED)



BARRIERS TO ADOPTION
o Connectivity issues (internet of things)
o Limited applicability in dynamic simulation-based environments used for military training
o AIS in the team-context is not-well studied
o Lack of Fidelity (Natural Language depedencies)
o Time, cost, and personnel requirements to build an AIS

PLAYERS in INDUSTRY and/or ACADEMIA

Figure A-4: iCard Example of Adaptive Instructional Systems Developed During the HFM-297 Meeting at DRDC Toronto.

A.1 REFERENCES

[1] Adlakha-Hutcheon, G. and Jarmasz, J. (2019). "Fostering Innovation in Training and Education through a Novel Technology Assessment Approach." DRDC Scientific Letter DRDC-RDDC-2019-L308, Defence Research and Development Canada.

Annex B – HFM-297 ENGAGEMENTS AND EXTRAMURAL ACTIVITIES

During execution of the Research Task Group (RTG) there was careful consideration in using the semi-annual meetings and other outreach opportunities to directly engage with relevant stakeholders both at the research and end-user level. This allowed the group to get better perspective on the tasks augmentation technologies were being applied against, while also serving the goal of better defining the performance implications for this technology space at-large. This resulted in several workshops and interactive events that incorporated demonstrations and briefings linked to current research and development efforts (see Table B-1 for stakeholder engagements). In this Annex we will briefly describe each event and any relevant outcomes and products generated from those interactions.

Table B-1: HFM-297 Workshop and Demonstration Events.

Event	Location (Date)	Description
End-User Research and Development Workshop (Described Below)	United States Military Academy, West Point, NY USA (Spring 2018)	Interactive discussions and hands-on demonstrations of augmentation technologies across several use cases. Emphasized current domains of investment, with direct attention to barriers of adoption and performance criteria for viable use.
International Ergonomics Association (IEA) World Congress HFM-297 Paper Session (Described Below)	Florence, ITALY (Fall 2018)	Focused paper session with three contributions to the proceedings from RTG members. Served as first outreach activity to engage with experts across government, industry and academia.
End-User Research and Development Workshop	Bundeswehr Command and Staff College – BwCSC, Hamburg, GERMANY (Fall 2018)	Engagement with schoolhouse leadership and practitioners. Provided opportunity to discuss RTG ideas and preliminary standing with military personnel at the BwCSC. The BwCSC provides training, advanced training and further education for German staff officers and upcoming Generals. The BwCSC is also the main pillar of Germany’s brand new Think Tank, The German Institute for Defence and Strategic Studies, exploring trends, innovation and necessary capabilities for the government and its ministry of defence.
HFM-297 Panel at Canadian Armed Force’s “Ready, Set, Innovate” Symposium (Described Below)	Canadian Forces Base (CFB) Borden, CANADA (Spring 2019)	Coordinated and executed a Panel Discussion on HFM-297 activities and objectives. Facilitated discussion with symposium participants and leadership within Canadian Armed Forces’s Military Personnel Generation (MILPERSGEN) Command.

Event	Location (Date)	Description
Industry Engagement and Interactive Matrix Workshop (Described Below)	Stockholm, SWEDEN (Fall 2019)	Hands-on with latest tech and research investments across Swedish industry base. This included tours and demonstrations from Intel, Ericsson and Microsoft. Briefings were also delivered across internationally recognized researchers, with focus on HW specifications linked to human perception and communication requirements. At the conclusion, we executed a workshop examining an interactive framework for building technology recommendations and insights. Goal is to establish new reporting method for RTGs working in a volatile technology sector. Draft framework for consideration is provided below.
Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) HFM-297 Special Demonstration Event (Described Below)	Orlando, FL USA (Fall 2019)	HFM-297 hosted an interactive forum to demonstrate current trends in augmentation technology in the context of human performance. This involved hands-on demonstrations of eleven current and emerging technologies that interact with and stimulate user(s) perceptual systems resulting in higher learning, performance, retention, and/or transfer.
COVID-19 Webinars (Described Below)	Virtual (2020)	Across the 2020 and 2021, COVID-19 impacted any face to face interactions. HFM-297 conducted two virtual webinars to continue engagement with recognized experts in the field: 1) Augmented Reality Enterprise Alliance (AREA) Mark Sage, AREA Exec Dir. 2) Human-Centered Augmented Reality Dr. Aaron Gardony, US Army DEVCOM Soldier Center.

B.1 END-USER RESEARCH AND DEVELOPMENT WORKSHOP; WEST POINT, NY USA

In spring 2018, HFM-297 conducted a workshop at the United States Military Academy (USMA) in West Point, NY. During this engagement, several demonstrations were provided that exhibited cutting-edge methods for utilizing adaptation and augmentation to address human performance requirements across a number of use cases. This involved considerations both at the knowledge and skill acquisition level through training methods and at the operational level with on-the-job supports. Below are descriptions of three demonstrations that were provided during this activity. They are further expanded upon in [1], with attention to the performance constructs that need to be measured and assessed to establish viability of augmentation techniques reliably supporting performance needs. This paper was presented at the International Ergonomics Association's World Congress, which is described in the next section of this Annex. Below are brief descriptions across the types of demonstrations that were provided, and the types of domains these technologies are being matured within. Outside of these three use cases, other domains of deep interest to the end user is visualization of cyber effects in an operational setting, augmentation techniques to manage human-agent teaming, and tactical augmentation capabilities linked to real-time operations.

Figure B-1 highlights research and development of a virtual pilot crafted to support experiential training through augmented reality interfacing. The resulting capability is designed to support crew and flight management activities, with characteristics in place to invoke emotion and confrontation commonly occurring on the flight deck. Initial objectives of the project are to optimize training time on full flight simulators and building scenarios that target communication procedures through verbal speech and body language, while also learning coping mechanisms for varying personality types [2].

Figure B-1: Virtual Pilot Seen through Augmented Reality Headset.

Figure B-2 represents the U.S. Air Force's Pilot Training Next program. They are examining the use of low-cost virtual reality technologies to support initial exposure training on in-flight routines, maneuvers and communication protocols [1]. This supports a new data driven training approach that is centered on adult learning methods, with careful consideration during front-end analysis on what knowledge and skill components immersive training interactions can reliably influence and support. This involves establishing a new training culture with technology insertions where virtual reality and augmented reality can be applied for efficient and cost-effective skill development, while facilitating meaningful sets and reps on the tasks and procedures required to assist in building competency and proficiency.

Figure B-2: Pilot Training Virtual Reality Configuration with Low-Cost Interfacing.

Figure B-3 associates with technology applied in developing a virtual command center for first-responder coordination and logistics operations [3]. The concept is based on a 'Connected City' with a network of resources and assets that can be used to plan and coordinate across multiple organizations (e.g., police, SWAT Team, medics, National Guard, firemen, etc.). The idea is that an individual can tap into a network of resources to gather data, intelligence, and visual confirmations associated with specified areas of interest.

Figure B-3: Emergency Operations Center Virtual Command Map.

B.2 INTERNATIONAL ERGONOMICS ASSOCIATION (IEA) WORLD CONGRESS HFM-297 PAPER SESSION; FLORENCE, ITALY

In the fall of 2018, HFM-297 hosted a paper session at the World Congress for the International Ergonomics Association, which is conducted once every three years. The goal was to establish a dedicated paper track that focused directly on implications between augmentation technology and learning science. This was an early outreach activity within the execution of the RTG, with an aim to engage with academic and industry leaders to better socialize activities within NATO's Science & Technology Office.

Three papers were authored by RTG members, including:

- Robert Sottilare's "Applying adaptive instruction to enhance learning in non-adaptive virtual training environments" [4].
- Thomas Alexander's "Virtual and Augmented Reality: Innovation or Old Wine in New Bottles?" [5]
- Benjamin Goldberg's "The Connection Between Constructs and Augmentation Technologies: Measurement Principles Linked to Training and Performance" [1]. This paper incorporated insights and outcomes from the end-user workshop facilitated at West Point, described above.

B.3 HFM-297 PANEL AT CANADIAN ARMED FORCE’S “READY, SET, INNOVATE” SYMPOSIUM; CANADIAN FORCES BASE BORDEN, CANADA

At the Spring Meeting of HFM-297 (30 April – 2 May 2019, Toronto, Canada), the group was invited to participate in the annual Canadian Armed Forces (CAF) Individual Training and Education (IT&E) Symposium, organized by the CAF’s Military Personnel Generation (MILPERSGEN) Command. This annual meeting brings together representatives from CAF Training Establishments, Learning Support Centres and the CAF Training Development Centre with external subject matter experts on education and training, to exchange ideas about the latest trends in training and education. The theme for the 2019 IT&E Symposium was “Ready, Set, Innovate” and accordingly, RTG HFM-297 was invited to provide an international perspective on innovation for training and performance improvement technologies. The briefing given by HFM-297 at the meeting (see slides below in this Annex) was also an opportunity for the RTG to solicit input from the IT&E community gathered at the MILPSERGEN Symposium on the group’s initial structured assessment of 4 augmentation technology areas (Head Mounted See Through Display (AR) and Navigation Technologies, Augmented Training through Adaptive Instructional Systems, Cognitive Training and Low current brain stimulation to mitigate cognitive effects of sleep loss), using the “iCard” methodology, developed in the earlier at the same RTG HFM-297 meeting (see more detailed description in Annex A as well as Ref. [6]). While the symposium format did not allow for a detailed discussion of each iCard separately, general themes about augmentation technologies were elicited from the audience. The main points elicited from the symposium were:

- The training SMEs expressed clear enthusiasm for change in the training “philosophy” of CAF, and a sense that this change is necessary;
- However, there does not seem to be a clear way forward to effect the desired changes and to manage their knock-on effects. Techniques like the one discussed here (technology instantiation assessment or iCard method) could provide a means to start addressing aspects of this gap; and
- Along these lines, concerns were expressed that the insertion of new training technologies might introduce new “burdens” on training staff, in terms of new requirements to operate and manage unfamiliar technologies ([6], p. 4).

Despite the interruption of the in-person IT&E Symposia due to the COVID-19 pandemic, training stakeholders in the CAF remain committed to leveraging “emerging technologies” for modernizing training, and to updating training and education doctrine, policy and processes based on findings from research. Thus, the efforts of NATO RTG HFM-297 continue to be of relevance to organizations such as the CAF Training and Education authorities.

B.4 INTERACTIVE MATRIX WORKSHOP; STOCKHOLM, SWEDEN

In the fall of 2019, HFM-297 conducted a workshop activity to explore interactive methods for building final reports within an RTG. This was motivated by the challenge of reporting against a technology sector that is rapidly maturing, with a goal of providing focused recommendations on acquisition best practices and identifying areas that require focused research investment.

Figure B-4 represents a draft framework of the concept that was developed. In this instance, each column represents a difference performance domain augmentation technology is being developed to support, and each row represents a different perspective on time. Near-term identifies domains and tasks that are well-suited for augmentation technology application, mid-term represents areas that are progressing within a laboratory environment, but are not ready for widescale adoption, and far-term define ideal end-states that can be used to help roadmap research and development needs to meet those overarching objectives. This framework serves as a starting point to guide and organize the creation instantiation Cards (iCards, see Annex A) that align to the performance domain of interest.

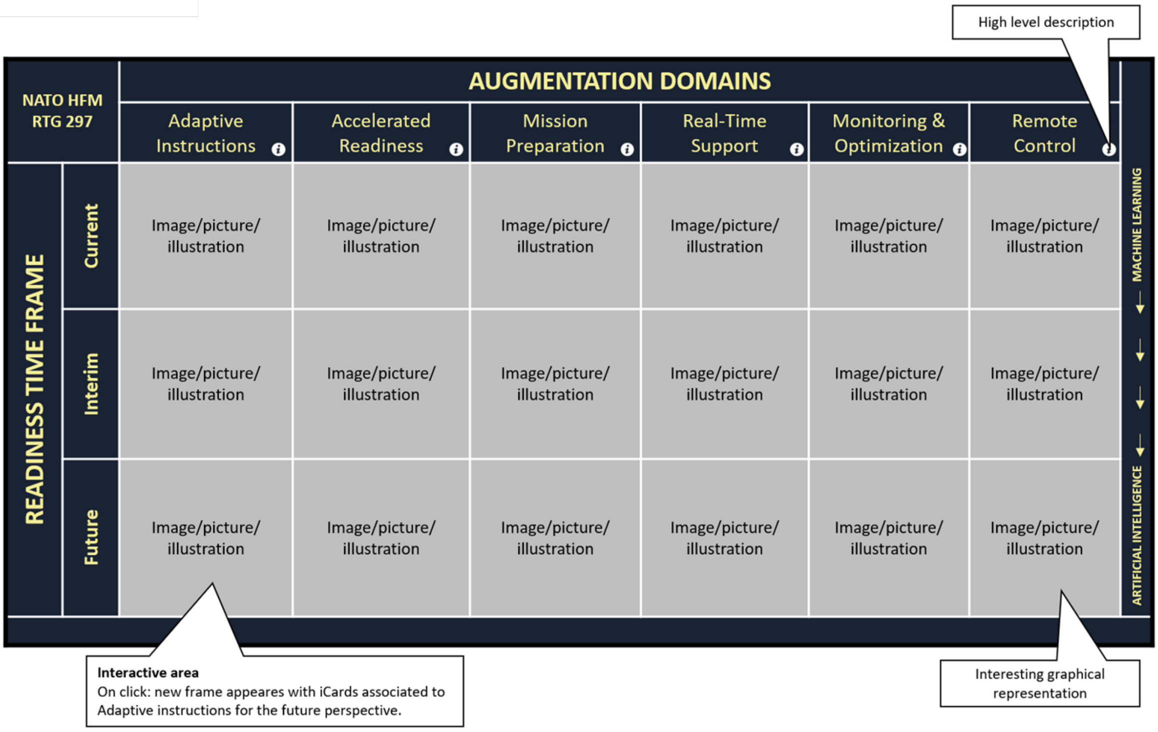


Figure B-4: Draft Interactive Matrix for Organizing Insights and Recommendations Across the Augmentation Domain of Interest.

B.5 INTERSERVICE / INDUSTRY TRAINING, SIMULATION AND EDUCATION CONFERENCE (I/ITSEC) HFM-297 SPECIAL DEMONSTRATION EVENT; ORLANDO, FL USA

In fall 2019, HFM-297 executed a hands-on Special Event at the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC). I/ITSEC is known as the largest defence focused simulation and training conference in the world and provides a unique opportunity to directly interact with researchers, practitioners, end-users, and senior leaders. The RTG proposed a demonstration focused event that would spotlight specific augmentation technologies being researched and developed in support of human performance requirements (see Figure B-5 for event description in I/ITSEC conference program).

B.6 COVID-19 WEBINARS

During the COVID-19 pandemic we had two invited expert briefings that led into our final report writing phase. The first briefing introduced the RTG to the Augmented Reality Enterprise Alliance (AREA) and was delivered by Mr. Mark Sage (AREA Executive Director). AREA is a non-profit alliance aimed at assisting in the acceleration of AR capabilities with an emphasis on ecosystem growth. As described on their website [7] "The AREA is working to help enterprises maximize the impact of AR by providing up-to-date resources and neutral, reliable guidance that make the path to AR adoption surer, shorter, and smoother. By identifying opportunities and challenges, disseminating information, spearheading research, promoting dialogue, and providing a forum for AR providers and enterprises, the AREA is clearing a path to interoperable AR enabled enterprise systems that fully deliver on their promises." It is important to identify appropriate engagement at the government and defence level, as the solutions these organizations are maturing align to future capability needs. Influencing the use cases and requirements the AREA targets will be critical to putting these technologies in the hands of those who need them faster.

FOCUS EVENT

New and Emerging Augmentation Technologies for Training and Operations within the NATO Alliance Nations

SHOWCASING THE STATE OF THE ART IN HUMAN PERFORMANCE AUGMENTATION

THURSDAY, 5 DECEMBER
1330 – 1500 • ROOM S310C
FE13

Moderators

Beth Biddle, Ph.D.
Technical Fellow/RTG Member, Boeing Research & Technology, The Boeing Company (United States)

Benjamin Goldberg, Ph.D.
Senior Scientist/RTG Member and Co-Chair, U.S. Army CCDC-Soldier Center, STTC (United States)

Panelists

Thomas Alexander, Ph.D.
RTG Member/Co-Chair, Federal Institute for Occupational Safety and Health (BAUA; Germany)

Jerzy Jamais, Ph.D.
RTG Member, Toronto Research Centre – Defence Research and Development Canada (Canada)

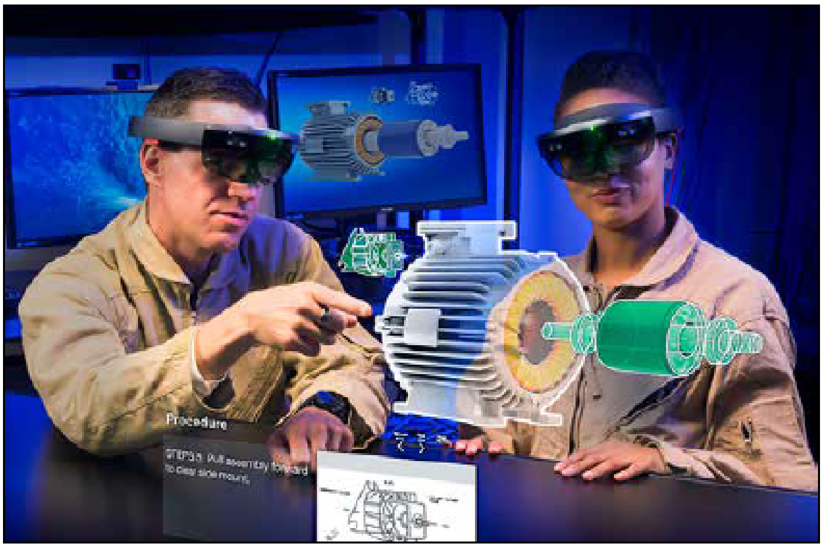
Glenn Gunzelmann, Ph.D.
RTG Member U.S. Air Force Research Laboratory (United States)

Peder Sjolund, Ph.D.
RTG Member, Skydome (Sweden)

Ian Greig, Ph.D.
RTG Member, Defence Science and Technology Laboratory (United Kingdom)

Dexter Fletcher, Ph.D.
RTG Member, Institute for Defense Analyses

LTC Vincent Capaldi, M.D.
RTG Member, Walter Reed Army Institute of Research



In this Special Event, we showcase the North Atlantic Treaty Organization (NATO) Research Task Group (RTG) focused on the assessment of human performance oriented augmentation technologies. The event will start with a short engagement with the RTG to review task group objectives, followed by an overview of the innovative technologies that will be showcased. We provide an interactive forum to demonstrate current trends in augmentation technology in the context of human performance. This will involve hands-on demonstrations of eleven current and emerging technologies that interact with and stimulate user(s) perceptual systems resulting in higher learning, performance, retention, and/or transfer. As a participant in the Special Event, you will have a chance to engage directly with RTG members and I/ITSEC attendees in this focused context. The event will be documented and included as a chapter in the RTG final recommendation report.

Technology Demonstrations:

- Madigan Army Medical Group (Kyle Couperus)
- Design Interactive (Luke Devore)
- NeuroTracker (Scott Kozak and Lee Sidebottom)
- Modest Tree Media (Sam Sannandeji and Emily Smits)
- Intelligent Automation (Bob Pokorny, Chad Zalkin, Jeff Kish and Lisa Holt)
- Soar Technology (Alyssa Tanaka)
- Charles River Analytics (Caroline Kingsley, Arthur Wollocko and Michael Jenkins)
- HTC Vive (Amy Peck and Frank Black)
- Swedome (Peder Sjolund)
- University of Central Florida METIL Lab (David Metcalf, Tim Welch, Michael Eakins
- Microsoft/Insight Enterprise (Matt Fedorovich, David Eager)

Session Chair: Craig Langhauser, Collins Aerospace

Figure B-5: I/ITSEC Focus Event Description and Demonstration Participants as Highlighted in the Conference Program.

The second briefing delivered over a virtual webinar format was executed by Dr. Aaron Gardony (Scientist, Center for Applied Brain and Cognitive Science; U.S. Army DEVCOM Soldier Center). This discussion centered on design and computation requirements linked to Level of Detail in AR and what is good enough to support the perceptual processes of a human user. Dr. Gardony also presented work on eye-tracking in AR with an emphasis on human-computer interaction and using this data source to optimize interfacing techniques with a dedicated user. See Ref. [8] for a great overview on this topic of interest.

B.7 REFERENCES

[1] Goldberg, B. (2018). "The Connection Between Constructs and Augmentation Technologies: Measurement Principles Linked to Training and Performance." In Congress of the International Ergonomics Association, pp. 370-379. Springer, Cham.

[2] Biddle, B. (2018). "Building a Virtual Reality Pilot: Embedding the End User Throughout the Process." NATO RTG HFM-297 Workshop on Augmentation Technologies. West Point, NY, 2018.

[3] Swedome (n.d.). Command, Collaboration and Communication (C3) Platform. https://www.skydome.se/ (Accessed 15 June 2022)

[4] Sottilare, R.A. (2018). "Applying Adaptive Instruction to Enhance Learning in Non-Adaptive Virtual Training Environments." In Congress of the International Ergonomics Association, pp. 155-162. Springer, Cham.

[5] Alexander, T. (2018). "Virtual and Augmented Reality: Innovation or Old Wine in New Bottles?" In Congress of the International Ergonomics Association, pp. 233-239. Springer, Cham.

[6] Adlakha-Hutcheon, G., and Jarmasz, J. (2019). "Fostering Innovation in Training and Education through a Novel Technology Assessment Approach." DRDC Scientific Letter DRDC-RDDC-2019-L308, Defence Research and Development Canada.

[7] Augmented Reality for Enterprise Alliance (AREA). About the Augmented Reality Enterprise Alliance (n.d.). https://thearea.org/about-us/ (Accessed 01 July 2022)

[8] Gardony, A.L., Lindeman, R.W., and Brunyé, T.T. (February 2020). "Eye-Tracking for Human-Centered Mixed Reality: Promises and Challenges." In Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR), 11310, pp. 230-247. SPIE.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	STO-TR-HFM-297 AC/323(HFM-297)TP/1141	ISBN 978-92-837-2455-1	PUBLIC RELEASE

5. Originator

Science and Technology Organization
North Atlantic Treaty Organization
BP 25, F-92201 Neuilly-sur-Seine Cedex, France

6. Title

Assessment of Augmentation Technologies for Improving Human Performance

7. Presented at/Sponsored by

This report documents the findings of the Human Factors and Medicine Research Task Group 297.

8. Author(s)/Editor(s)	9. Date
Multiple	December 2023

10. Author's/Editor's Address	11. Pages
Multiple	128

12. Distribution Statement

There are no restrictions on the distribution of this document. Information about the availability of this and other STO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Adaptation; Adaptive instructional systems; After action review; Augmentation; Augmented reality; Cognitive monitoring; Extended reality; Fatigue; Human performance; Intelligent tutoring systems; Real-time support; SWOT analysis; Virtual reality

14. Abstract

NATO Human Factors and Medicine (HFM) Research Task Group (RTG) HFM-297 was established to support the "Assessment of augmentation technologies for improving human performance". A framework was developed to analyze these technologies and integrate evidence from the research literature around an operational timeline and varying mission contexts. This resulted in an analysis of augmentation technologies aligned to five performance domains that associate with the preparation, execution and assessment of mission oriented tasks, including: 1) Adaptive instruction and accelerated readiness; 2) Mission preparation and rehearsal; 3) Real-time support and remote control; 4) Cognitive monitoring and optimization; and 5) After-action reviews. A Strengths-Weaknesses-Opportunities-Threats (SWOT) Analysis methodology was applied across each domain to organize and communicate insights on the impact and limitations of augmentation technologies aligned to performance outcomes and variables. This resulted in a set of recommendations centered on best practice application of augmentation technologies and identified areas that require further research investment. Common technology themes across performance domains were identified, with a future focus and emphasis on data and interoperability. There is also a recognition that these technology types are maturing rapidly and maintaining concurrent recommendations proves challenging.

NORTH ATLANTIC TREATY ORGANIZATION

SCIENCE AND TECHNOLOGY ORGANIZATION

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cso.nato.int

DIFFUSION DES PUBLICATIONS
STO NON CLASSIFIEES

Les publications de l'AGARD, de la RTO et de la STO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la STO, 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 STO, de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

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

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la STO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (http://www.sto.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 STO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Bruxelles

BULGARIE
Ministry of Defence
Defence Institute "Prof. Tsvetan Lazarov"
"Tsvetan Lazarov" bul no.2
1592 Sofia

CANADA
DGSIST 2
Recherche et développement pour la défense Canada
60 Moodie Drive (7N-1-F20)
Ottawa, Ontario K1A 0K2

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

ESPAGNE
Área de Cooperación Internacional en I+D
SDGPLATIN (DGAM), C/ Arturo Soria 289
28033 Madrid

ESTONIE
Estonian National Defence College
Centre for Applied Research
Riia str 12
Tartu 51013

ETATS-UNIS
Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

FINLAND
Ministry for Foreign Affairs
Telecommunications Centre (24/7)
P.O BOX 176
FI-00023 Government

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
Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25, H-1885 Budapest

ITALIE
Ten Col Renato NARO
Capo servizio Gestione della Conoscenza
F. Baracca Military Airport "Comparto A"

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket

PAYS-BAS
Royal Netherlands Military Academy Library
P.O. Box 90.002, 4800 PA Breda

POLOGNE
Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

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

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

ROYAUME-UNI
Dstl Records Centre
Rm G02, ISAT F, Building 5
Dstl Porton Down
Salisbury SP4 0JQ

SLOVAQUIE
Akadémia ozbrojených síl gen. M.R. Štefánika, Distribučné a informačné stredisko STO
Demänová 393
031 01 Liptovský Mikuláš 1

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

TCHEQUIE
Vojenský technický ústav s.p.
CZ Distribution Information
Mladoboleslavská 944
PO Box 18, 197 06 Praha 9

TURQUIE
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

AGENCES DE VENTE

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, Ontario K1A 0S2
CANADA

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « 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 STO, RTO et AGARD figurent dans le « NTIS Publications Database » (http://www.ntis.gov).

NORTH ATLANTIC TREATY ORGANIZATION

NATO OTAN

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cso.nato.int

SCIENCE AND TECHNOLOGY ORGANIZATION

DISTRIBUTION OF UNCLASSIFIED STO PUBLICATIONS

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

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

Requests for STO, RTO or AGARD documents should include the word 'STO', '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 STO reports as they are published, please visit our website (http://www.sto.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 STO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Brussels

BULGARIA
Ministry of Defence
Defence Institute "Prof. Tsvetan Lazarov"
"Tsvetan Lazarov" bul no.2, 1592 Sofia

CANADA
DSTKIM 2
Defence Research and Development Canada
60 Moodie Drive (7N-1-F20)
Ottawa, Ontario K1A 0K2

CZECHIA
Vojenský technický ústav s.p.
CZ Distribution Information Centre
Mladoboleslavská 944
PO Box 18, 197 06 Praha 9

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

ESTONIA
Estonian National Defence College
Centre for Applied Research
Riia str 12, Tartu 51013

FINLAND
Ministry for Foreign Affairs
Telecommunications Centre (24/7)
P.O BOX 176, FI-00023 Government

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
Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25
H-1885 Budapest

ITALY
Ten Col Renato NARO
Capo servizio Gestione della Conoscenza
F. Baracca Military Airport "Comparto A"
Via di Centocelle, 301
00175, Rome

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
Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 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 gen. M.R. Štefánika, Distribučné a informačné stredisko STO
Demänová 393
031 01 Liptovský Mikuláš 1

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

SPAIN
Área de Cooperación Internacional en I+D
SDGPLATIN (DGAM)
C/ Arturo Soria 289
28033 Madrid

TÜRKIYE
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

UNITED KINGDOM
Dstl Records Centre
Rm G02, ISAT F, Building 5
Dstl Porton Down, Salisbury SP4 0JQ

UNITED STATES
Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

SALES AGENCIES

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, Ontario K1A 0S2, CANADA

Requests for STO, RTO or AGARD documents should include the word 'STO', '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 STO, RTO and AGARD publications are given in "NTIS Publications Database" (http://www.ntis.gov).

ISBN 978-92-837-2455-1