NORTH ATLANTIC TREATY ORGANIZATION

SCIENCE AND TECHNOLOGY ORGANIZATION

AC/323(HFM-300)TP/896

www.sto.nato.int

STO MEETING PROCEEDINGS **MP-HFM-300**

Human Autonomy Teaming

(L'association Humains-Machines autonomes)

STO Human Factors and Medicine (HFM) Panel Specialists' Meeting held on 15 – 17 October 2018 in Seasouth, United Kingdom.

Published April 2019

Distribution and Availability on Back Cover

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 April 2019

Copyright © STO/NATO 2019
All Rights Reserved

ISBN 978-92-837-2221-2

Single copies of this publication or of a part of it may be made for individual use only by those organisations or individuals in NATO Nations defined by the limitation notice printed on the front cover. The approval of the STO Information Management Systems Branch is required for more than one copy to be made or an extract included in another publication. Requests to do so should be sent to the address on the back cover.

Human Autonomy Teaming

(STO-MP-HFM-300)

Executive Summary

A number of applications have been identified by the military that require IR laser sources. These include – platform protection, remote sensing, improved target identification and improved targeting. Several laser technologies are being developed by teams across NATO to address these requirements. These laser technologies include

- using already established lasers at other wavelengths then frequency shifting to the required wavelength,
- direct generation of the required wavelengths either in fiber lasers or Quantum Cascade/Interband Cascade lasers (QCL/ICL).

In order to be deployable by the military any laser and system based around the laser must be affordable (both to buy and through life), have low size and weight, require low power and be robust. The SET-210 meeting brought together experts on IR laser technologies and systems experts that use those lasers. The current state of development of both lasers and systems was presented and ideas exchanged about how to take the technologies forward.

Good progress has been made across all the laser technologies presented with CW powers of 10's W and pulse energies in the 1J region being reported. Broadband and tuneable outputs have also been achieved. Progress was also reported on beam delivery through fibers and manufacturing approaches to provide reliable supply of lasers and other optical components. Results from systems based on currently available laser sources showed the potential of the technologies being explored.

All the laser technologies reported showed potential as a source for at least one of the applications identified, therefore, development should continue on them all. There should be opportunity for engagement between system and laser developers to ensure laser development is directed at meeting system needs. Size, weight and robustness can be addressed by developing systems with few or no free-space optics – i.e. using fiber or QCL/ICL type lasers and making use of waveguides and fibers to transmit the laser beam. Broadband/super-continuum sources covering the IR waveband are needed for remote chemical detection. Multi-functionality of the laser and system should be a development aim as this will be desirable to the end user. Reliable, affordable supply of QCLs is essential as these will be important either as direct laser sources or as pump sources for fiber lasers.

L’association Humains-Machines autonomes

(STO-MP-HFM-300)

Synthèse

Les militaires ont identifié un certain nombre d’applications qui ont besoin de sources laser infrarouges. Il s’agit de la protection des plates-formes, de la détection à distance, de l’identification améliorée des objectifs et de la désignation améliorée des objectifs. Plusieurs technologies laser sont en cours de développement au sein de l’OTAN pour répondre à ces besoins. Ces technologies laser incluent :

- l’utilisation de lasers déjà établis à d’autres longueurs d’onde, puis le déplacement de la fréquence jusqu’à la longueur d’onde requise ;
- la production directe des longueurs d’onde requises, soit dans des lasers à fibre, soit dans des lasers à cascade quantique (LCQ) ou interbandes (LCI).

Pour pouvoir être déployés par les armées, les lasers et systèmes reposant sur ceux-ci doivent être économiques (à l’achat et pendant leur durée de vie), de petite taille, légers et robustes et nécessiter peu d’énergie. La réunion du SET-210 a rassemblé des experts des technologies laser infrarouges et des experts des systèmes qui utilisent ces lasers. L’état actuel de développement des lasers et des systèmes a été présenté et des idées ont été échangées sur la manière de faire progresser les technologies.

Toutes les technologies laser présentées ont fait de beaux progrès : des puissances d’onde porteuse de plusieurs dizaines de watts et des énergies pulsées proches de 1 J ont été signalées. Des signaux accordables et à large bande ont également été obtenues. Des progrès ont été signalés au sujet de la transmission du faisceau par les fibres et des démarches de fabrication visant à assurer un approvisionnement fiable de lasers et autres composants optiques. Les résultats des systèmes basés sur les sources laser actuellement disponibles ont montré le potentiel des technologies étudiées.

Toutes les technologies laser signalées ont montré du potentiel en tant que source pour au moins l’une des applications identifiées. Par conséquent, il faudrait continuer à les développer toutes. Des rencontres entre développeurs de systèmes et développeurs de lasers devraient être organisées pour veiller à ce que le développement des lasers soit orienté de manière à répondre aux besoins des systèmes. Les questions de dimensions, de poids et de robustesse peuvent être réglées en développant des systèmes comportant peu ou pas de propagation optique en espace libre, autrement dit, utilisant des lasers à fibre ou de types LCQ / LCI et des guides d’ondes et des fibres pour transmettre le faisceau laser. Les sources à large bande ou supercontinuum couvrant la gamme d’ondes infrarouges sont nécessaires à la détection chimique à distance. La fonctionnalité multiple du laser et du système devrait être un objectif de développement, car elle sera souhaitable pour l’utilisateur final. Un approvisionnement fiable et abordable de LCQ est essentiel, car les LCQ seront importants soit comme sources laser directes, soit comme sources de pompage pour les lasers à fibre.

Programme Committee

Chair

Prof. Dr. Frank Flemisch

Fraunhofer FKIE/RWTH Aachen

GERMANY

email: mark.draper.2@us.af.mil

Co-Chair

Dr. Mark DRAPER

US Air Force Research Laboratory

UNITED STATES

email: mark.draper.2@us.af.mil

Group Mentor

Prof. Dr. Adelbert BRONKHORST

TNO Human Factors

NETHERLANDS

email: adelbert.bronkhorst@tno.nl

Towards Intelligent Adaptive Human Autonomy Teaming: on 2018 NATO SCO HFM-300 Symposium

Technical Evaluation Report

Robert M. Taylor
Defence Science and Technology Laboratory
Air Platforms Division, Air Systems Autonomy
UNITED KINGDOM

rmtaylor@dstl.gov.uk

ABSTRACT

This report provides a review of the proceedings, aims and objectives, and on the structure, content and delivery of the NATO SCO HFM-300 Symposium on Human Autonomy Teaming (HAT), Southsea UK, 15-17 October 2018. The background context is described, including the development of the themes, evolving from the introduction of automation to manned platforms and mission information management systems. Themes covered include operational requirements, human-autonomy teaming structure, autonomous capabilities that support teaming, HAT interaction and design, and HAT institutional integration. Critical evaluation of technical quality is provided using a combination of HAT metrics and analytical tools, calibrating operational relevance, exploitability, scientific quality, technical capability and innovation, with integration, analysis and interpretation of the evidence. The report summarises the state of art, considers the risks and opportunities, and identifies potential future directions, in particular the need for Intelligent Adaptive Human Autonomy Teaming.

1.0 EXECUTIVE SUMMARY

Over 160 human factors researchers, scientists and technologists gathered in October 2108 in Southsea UK, located on the south coast near Portsmouth to discuss human factors of automation, Artificial Intelligence (AI) and systems with increasingly autonomous capabilities. With advances in computation, artificial intelligence, machine learning, sensors and connectivity, machines gain increasing capabilities to act more autonomously. Applications include the development of automated and autonomous driving, autonomous sea and underwater vessels, autonomous air vehicles for imagery collection, and also cyber defence, image analysis and autonomous logistics. As systems become more intelligent and autonomous, automation can transition from being used as a simple tool to becoming regarded as more of an intelligent teammate. In recognition of this transitioning in concept of automation utility and use, the meeting was convened to discuss human cooperation and collaboration with automation, specifically Human-Autonomy Teaming (HAT). HAT is a design principle for systems that can allow users to more fully engage, interact and cooperate with intelligent automation to perform task more efficiently and effectively, deriving benefits for performance from teaming synergy, with maximum agility and adaptability to unexpected events. Like automation, HAT raises critical challenges related to tractability, veracity, authority, responsibility, accountability, directability, observability, shared comprehension and trust. For legal and ethical reasons, and for assured effective command and control, humans must retain user sovereignty, enabling a certain level of control with the ability to influence and direct system behaviour as required. If designed correctly, effective HAT technologies will enable autonomous systems to operate robustly in hazardous environments, support the tempo of rapid operations, and operate around and interact with human, manned platforms and other autonomous systems in highly constrained and chaotic environments. The report provides a review of the proceedings, aims and objectives, and considers the structure, content and delivery of the Symposium deliberations. The background context is described,

including the development of the themes, evolving from the introduction of automation to manned platforms and mission information management systems. Themes covered include operational requirements, human-autonomy teaming structure, autonomous capabilities that support teaming, HAT interaction and design, and HAT institutional integration. Critical evaluation of technical quality is provided using a combination of HAT metrics and analytical tools, calibrating operational relevance, exploitability, scientific quality, technical capability and innovation, with integration, analysis and interpretation of the evidence. The report summarises the state of art, considers the risks and opportunities, and identifies potential future directions in particular the need for cognitive technologies and systems designed to support Intelligent Adaptive Human Autonomy Teaming. Detailed summarising information on the individual technical contributions is provided as an Annex.

KEY WORDS: Human factors; Automation; Cognitive technologies; Artificial intelligence; Autonomy; Teaming; Adaptive systems; Command and control; Test and evaluation; Human effectiveness; Technical quality; Technology readiness; Human readiness; Capability maturity; Operational relevance.

2.0 INTRODUCTION

"Optimising human and machine capabilities in teams that maximise strengths and mitigate weaknesses is essential"
Human Machine Teaming, UK MoD Joint Concept Note 1/18.
UK Ministry of Defence, Development, Doctrine and Concepts Centre. May 2018. [1]

In 1942, the short story "Runaround" written by Isaac Asimov, inspired perhaps by Mary Shelley's 1818 Frankenstein, first described three basic laws aiming to assure safe behaviour of Robots, subsequently included in "I, Robot" published in 1950. Asimov's now famous Laws of Robotics were designed to assure human authority and safety. Now, developments in technology have made automation systems in all its forms - networked, integrated and embedded - a ubiquitous fact of modern life. This technology revolution is particularly challenging for the military needing to exploit the benefits of increasingly effective autonomous systems. The predominant question has now become how to assure efficient and effective Command and Control of autonomy enabled capabilities in the complex, dynamic military environment.

Automation has evolved and been transformed into autonomy and autonomous systems through Artificial Intelligence (AI) and agent-based systems. Automation replacing human involvement has become the dominant default strategy for many "industrial" dull, dirty and dangerous work domains. Replacement automation is considered by many philosophers and thought leaders, such as the Symposium Keynote Dr Peter Hancock, to be ultimately a "zero sum", no-win game strategy for humanity [2]. To counter this, the scientific and technical community, concerned with human-machine interaction and human-computer systems design, seek a more balanced, efficient and effective Teaming "win-win" approach.

Contemporary narratives invoke Human-Autonomy Teaming (HAT) as a counter to concerns about automation "brittleness", use and abuse, reliability and trust, mitigating the risks of cognition and automation bias, and enabling capability through a human-computer partnership in a synergistic teaming relationship. Conversely, teaming risks creating unintended over-dependence and bias, if the relationship is incorrectly balanced, reducing human ethical agency and sovereignty. A positive teaming strategy seeks to harness and exploit for betterment of human use the evolving developments in automation and autonomy. Enabling technology developments include big data base analytics, information fusion, information management and information planning tools and algorithms, and more recently advances in machine learning, learning re-use, and learning transfer. These technologies are coupled with human capabilities through interaction, dialogue, information exchange, decision aiding and learning transfer. Delivered effectively, augmenting human cognition for problem solving and decision making, they add significantly to the power of human understanding and reasoning, knowledge, experience, learning, imagination, creativity and innovation. Collectively, this teaming synergy aims to provide optimisation in human systems cognition effectiveness,

delivering more intelligent adaptive and agile solutions, and more deliberate, deceptive, decisive and superior military combat power. Exploiting human capability is identified in the US Third Offset strategy as a unique, enduring and powerful resource [3]. The belief is that by prioritising HAT for augmentation of human cognition, working collaboratively in an intelligent adaptive partnership, the benefits of synergy should create a "win-win" situation from both a technology and a human-centred military perspective.

Helpfully, models for the required HAT relationship have been visualised and immortalised in a highly anthropomorphised, entertaining and notionally idealised form, by the crew supportive behaviours of the R2D2 and C-3PO "Droids" in 1977-1983 Star Wars film trilogy. Now, the challenge for the scientific and technical community is to identify and define the essential characteristics of effective teaming relationships, including behaviours, communication, interactions and interdependencies, between humans and intelligent autonomous systems. The required relationship needs to be based on veracity and rationality, truth and reasoning, without resort to empathic anthropomorphism risking biased, uncalibrated reliance and trust.

Meanwhile, automation technology continues to evolve. Recent developments in the application of AI through machine learning (ML), distributed networked communication in computational systems enabling large scale collaboration, and big data analytics, together provide a technical convergence with powerful synergistic, transformative potential. Evolving technical progress has increased the possibility of realising soon concepts such as Artificial General Intelligence (AGI), ML, social intelligence, machine creativity or "Elastic Thinking", ethical robotics, intelligent adaptive systems, critiquing systems, autonomous agent systems, and reliably safe autonomous vehicles.

Verification and validation of the veracity of AI data, information and decisions that is uncertain and non-deterministic in nature, present significant challenges for AI exploitation in the military domain [4]. Smart advanced, anticipatory intelligence and planning, and force multiplication always matters. Critical military applications involve high levels of complex dynamic risk, where the consequences of inaccuracy can be severe. In combat operations, and many dynamic missions, tactical flexibility and the ability of humans to comprehend, conceptualise and perform the unexpected effectively, often creates the battle-winning edge.

Notwithstanding these concerns, the technological developments raise distinctly the near term prospect of at least co-equal, and most probably, asymmetrically high machine team member computational cognitive competencies. Technology trends, scientific evidence, popular wisdom, and official policy (e.g. U.S. Third Offset Strategy [3], US DoD DSB Role of Autonomy Report 2012 [4], UK MoD JCN 1/18 Human Machine Teaming [1]) increasingly indicate that the correct efficient, effective and safe relationship for future assured command and control should be confidently based on Human-Autonomy Teaming principles, promoting human capital and retaining human sovereignty. Human-Autonomy Teaming should be modelled on a balanced, co-equal, peer-to-peer, human and autonomy, mixed initiative teaming with the benefits of symbiotic synergistic interdependencies.

Joint human-machine cognitive systems, or AI augmented cognition, are increasingly becoming recognised as the future gold standard for intelligent military command and control. The teaming relationship needed is one that is designed to fully exploit both the human and AI autonomy capabilities for achieving decision superiority, agility and adaptability, in the complex, dynamic, uncertain military battle-space. Critically, both as an enabler and design constraint, it is necessary, axiomatic and legally imperative, that the required teaming relationship is to be achieved whilst maintaining human ethical agency and authority for assured effective command and control of weapon systems lethal force. This means retaining human authority, responsibility and accountability for meaningful, safe and effective human monitoring and supervision over system control and performance effects, including enabling processes and functions, at effective command and control "cognitive echelon" levels. This means maintaining human agency, competency, skills and knowledge at critical system decision making levels, enabling the exercising of critical human ethical judgment over the veracity of Human-Autonomy Team decision making processes and their outcomes.

Teaming principles and processes are needed to ensure the correct human engagement and decision supervision veracity in collaborative human-machine decision making processes. This can be achieved by team members working collectively to mitigate errors from cognition and automation bias. Automation bias is a widely recognised risk whereby humans exhibit a propensity to accept information from automated decision making processes, and to ignore contradictory information made without automation.

Variability inherent in human performance impacts adversely on delivering the human supervisory control requirement. Limitations on human cognition and information processing, such as the restricted bandwidth of sensory processes and the restricted span of attention, cause unawareness, inattention and distraction errors, which coupled with fatigue, lead to vigilance decrement and impaired error-prone performance. Additionally, limitations on human learning, memory and recall, and on human comprehension and reasoning, contribute to bias and error in decision making, effecting course of action planning and action execution.

Passive supervisory control solutions risk poor human engagement. The human condition needs active engagement for optimum performance. Interdependent, interactive, dynamic and adaptive teaming processes demand human engagement and facilitate supervisory control. This includes common ground, shared situation awareness, dialogue and essential information exchanges, mutual risks recognition and mitigation, and collaborative problem solving. All these are required features and defining characteristics of HAT needed for human assured effective command and control.

In theory of cybernetics [5], the science of communications and automatic regulatory control systems, in both machines and living things, working in a “coequal-by-design” teaming partnership can be considered to create the potential to benefit from a symbiotic and synergistic relationship. In a synergistic relationship, the combined effect is greater than the sum of their separate effects. Thus, in theory, through synergy, HAT can potentially create force multiplication, with effects increasing decision superiority and extending the battle-space winning edge.

This Technical Evaluation Report seeks to add value to the Symposium Proceedings Report, providing additional information for readers’ guidance. Guidance is provided so as to better appreciate of the Proceedings Report, to assist navigation through its structure and contents, and to facilitate analysis and interpretation of the Symposium outcomes.

This report includes some introductory information on the background and context of the Symposium, including the historical development towards the current focus and themes. This is followed by details of the Symposium’s specific aims and objectives, together with a high level summary of the resultant Symposium Programme structure and content. Next, the detailed Technical Evaluation of the Symposium Programme content and outcomes is provided, delivered using a structured evaluation framework with HAT analytical tools and assessment metrics for analysis, integration and interpretation. The evaluation seeks to summarise the state of art, the risks and opportunities, and potential future directions, with emphasis on the structure, content and delivery. Detailed summarising information on the individual technical contributions is provided in the Annex to the report.

2.1 Background

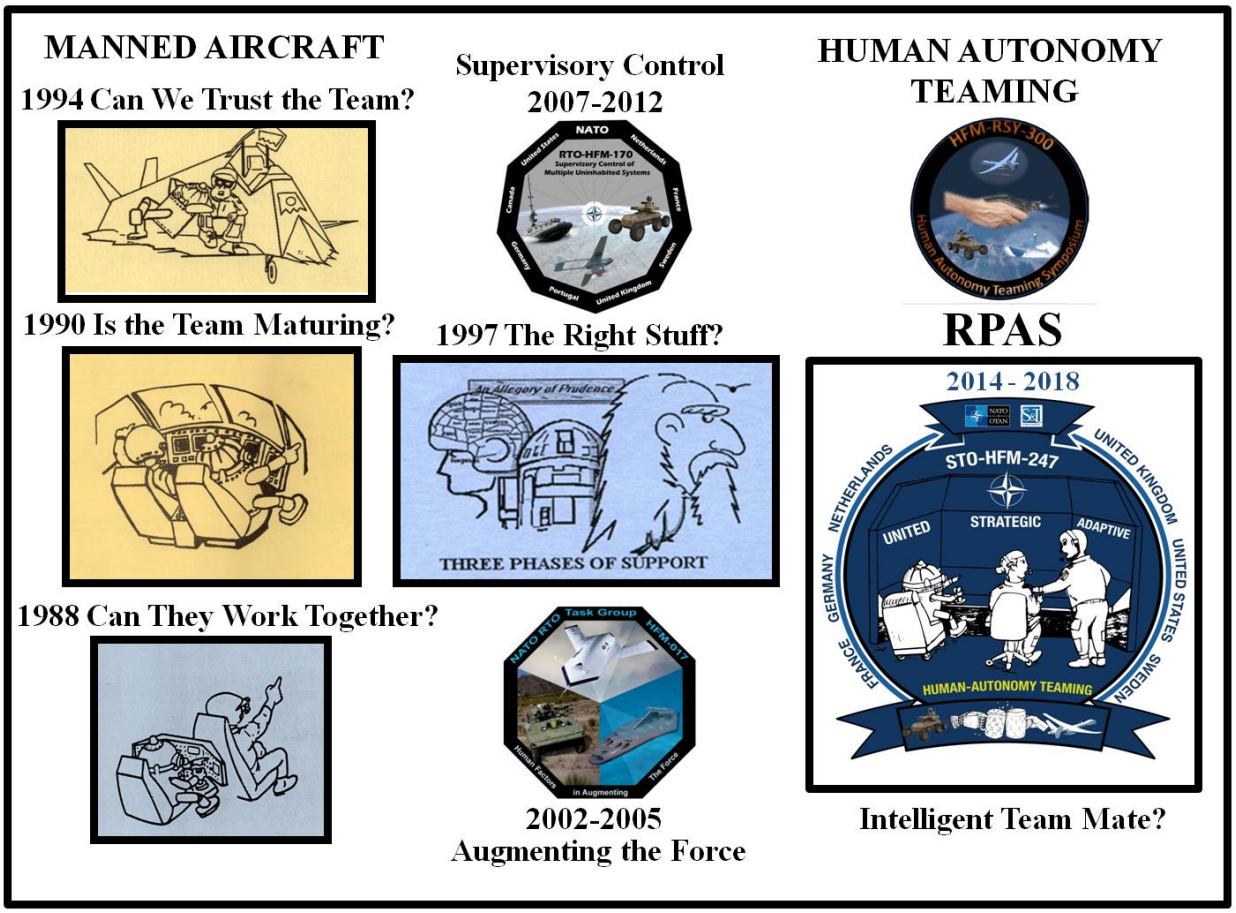


Figure 1: Precursor Human Autonomy Teaming Collaboration Activities.

The origins of human factors research on HAT can be traced to the 1980's era with the introduction of automation into manned aircraft operations. This introduced concepts for adaptable or adaptive automation, or pilot aiding technologies and intelligent decision support systems, with Assistant, Associate and Coach Levels of capability proposed for mixed initiative collaborative problem solving. The era is exemplified best by the US Pilot's Associate programmes (1985-1992), with multiple Knowledge-based Systems (KBS) planners and managers, as illustrated below [6].

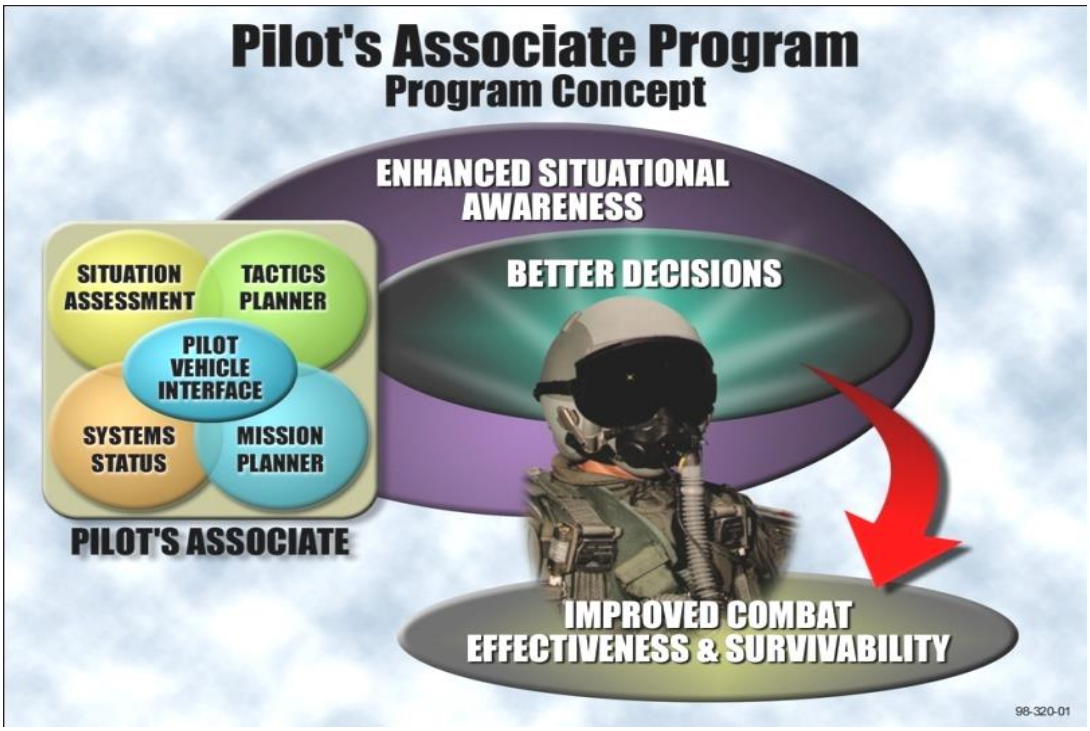


Figure 2: USAF Pilot's Associate Programme Concept [6]

Contemporary FR (Co-Pilote Electronique), GE (Cooperative Automation, Cockpit Assistant System CASSY, Crew Assistant Military Aircraft CAMA), and UK (Mission Management Aid MMA, Cognitive Cockpit KBS) applied research programmes built further understanding of the challenges and issues.

The concepts of intelligent pilot aiding coincided with the growth of interest in Cognitive Science and Cybernetics. Developments in the field of Cognitive Science and Cybernetics, from 1983 led to the introduction of ideas of Cognitive Systems Engineering for the design of socio-technical systems. This included Jens Rasmussen's influential 1986 Skills, Rule and Knowledge-based decision making framework [7], and Eric Hollnagel's ideas for Joint Cognitive Systems layered control, which were applied to UAS C2 through NATO RTO HFM TG-017 in 2007 [8]. It also coincided with interest in the introduction of AI computational techniques to mission planning and mission management systems engineering, such as probabilistic Bayesian reasoning engines, neural networks, agent-based systems, as well as Knowledge Based Systems.

At this time, between 1988 and 1997, a series of highly productive invitation workshop meetings took place on Human-Electronic Crew Teamwork, involving military, scientific and engineering specialists. Teamwork and trust provided the linking theme and thread [9, 10, 11, 12]. The workshops were organised jointly by representatives of the USAF, RAF and Germany Air Force human factors research institutions, funded by USAF EOARD. The focus then was on monitoring evidence and collaborating on H-E Crew Teaming maturity. The idea was to provide support and encouragement for an evolving concept, rather than seeking to define understanding of design requirements. Subsequently, the mission initiative was passed over to NATO through the RTO HFM MP-004 meeting in Edinburgh, 20-22 April 1998, on Collaborative Crew Performance in Complex Operational Systems [13]. This was followed by the NATO HFM Task Group 017 on UMV Augmenting the Force, 2002-2005 [14, 15], and Task Group 170 on Supervisory Control, 2007-2012 [16, 17], then the recent HFM Task Group 247 on Human Autonomy Teaming, 2014-2018, assuming peer-to-peer intelligent teammate capabilities, Final Report in press.

Since 2000, with the introduction of Remotely Piloted Air Systems (RPAS), focus has switched to multiple uninhabited air/land/sea vehicles (MUxV) autonomous systems supervisory control. Here, human supervisory control is executed from manned Ground Control Stations (GCS), exemplified by the USAF MUSCIT Vigilant Spirit GCS research facility [18], identifying multiple levels of human interaction, as illustrated in Table 1 below [19], and the evolving UK MoD QinetiQ UxV autonomy GCS research capability, currently QUARC [20]. Additionally, there is increasing interest in support for Manned-Unmanned Teaming (MUM-T) in air systems, in particular for rotorcraft operations, with airborne operator control of multiple UAVs. In 2001 US DARPA initiated their Advanced Research Programme on Augmented Cognition, focusing on cognitive state monitoring and manipulation, enhanced cognition and associated task adaptation, including manned cockpit and UAS GCS applications [21].

Table 1: Levels of Human Interaction (Eggers & Kelchner, 2012 [19])

Level	Name	Description
1	**Human Operated**	A human operator makes all decisions. The system has no autonomous control of its environment although it may have information-only responses to sensed data.
2	**Human Delegated** *(Management by consent)*	The vehicle has the capability to perform many functions independent of human control when delegated to do so. This level encompasses automatic controls, engine controls, and other low-level automation that must be activated or deactivated by a human input and must act in mutual exclusion of human operation.
3	**Human Supervised** *(Management by exception)*	The system can perform a wide variety of activities when given top-level permissions or direction by a human. Both the human and the system can initiate behaviors based on sensed data (e.g., conflict avoidance maneuver), but the system can do so only if within the scope of its currently directed tasks.
4	**Fully Autonomous**	System receives goals from humans and translates them into tasks performed without human interaction. A human would still be capable of entering the loop in an emergency or capable of changing the goals.

Moving forward towards the present day, it is possible to observe the influence of AI on manned cockpit applications as evidenced in 2015 5th Generation aircraft, F35 / Lightning II. US Air Force Chief Scientist, Gregory Zacharias, believes that dynamic human cognition remains more highly capable than computers [22]. AI impacts on F-35 sensor fusion capability, mission data fusion and mission data base, for decision aiding, using AI algorithms data and information processing management techniques. The F-35s sensor fusion uses computer algorithms to acquire, distil, organize and present otherwise disparate pieces of intelligence into a single picture for the pilot. Additionally, the F35 Autonomous Logistics Information System (ALIS) provides an information infrastructure, transmitting aircraft health and maintenance action information to the appropriate users on a globally-distributed network. ALIS uses early applications of AI to make assessments, process checklists, organize information and make some decisions without human intervention. Much higher degrees of autonomy and manned-unmanned teaming are likely to emerge in the near future including aiding airborne control of multiple UAV wingmen using AI and autonomy to perform sensing, reconnaissance and targeting functions.

Currently, HAT remains an evolving philosophy, increasingly informed by evidence to support formulation & definition of system design requirements. Examples include MUM-T Levels of Interoperability (LoI) and GCS Essential Information Exchange Requirements (EIERs).

In summary, the desired teaming capability, relationship and system should be characterised as follows:

- Assured effective Command and Control with balanced synergistic force mix of human and automation autonomy resources, optimising human and machine capabilities in teams that maximise strengths and mitigate weaknesses.
- Commanded and governed by user sovereignty, with control regulated by human ethical and moral agency, values and standards.
- Constrained by, and adaptive to the limitations of human cognition, i.e. span of attention, sustainment of vigilance, learning and memory, decision making cognition bias (automation bias, confirmation bias).
- Empowered by complex computational data and information processing capability, coupled with human cognitive capability for reasoning in uncertainty, with imagination, innovation and creativity in decision making in complex novel situations with unexpected events.

The principal research requirement remains to identify the defining characteristics of effective teaming relationships between this human and autonomy capability force mix.

3.0 PROGRAMME

Table 2: NATO SCO HFM-300 HAT Symposium Aims and Objectives

VISION GOALS	MEANS HOW			ENDS
	BARRIERS	PERSPECTIVES		
• Transition from a simple tool to more of an intelligent teammate. • Allow users to more fully engage, interact & cooperate with intelligent automation to perform tasks more efficiently, with maximum agility in response to unexpected events. • Increase effectiveness through flexible and robust Teaming. • Humans must retain a certain level of control with the ability to influence system behaviour as required. • Operate robustly in hazardous environments, support the tempo of rapid operations, or operate around & interact with humans, manned platforms, and other autonomous systems in highly constrained and chaotic environments.	• Overcome challenges: tractability, authority, directability, observability, adaptability, shared comprehension, trust, responsibility. • Connect domains: ground, air, sea, space and cyber • Bridge science with applications and operations	Overall system	• The big picture: Human autonomy teaming and cooperation in defence from a strategic perspective • Function division and relationships /roles (who does what functions, and in what kind of relationship?)	State of the art
		Technological factors	•Teaming and cooperation requires sufficient autonomy: Autonomous capabilities that support teaming	
		Human factors	• Human Machine interaction, communication & cooperation • Comprehension (how do humans & autonomy understand one another?) • Joint Learning/Training (collaborative learning methods, crew resource management, etc.)	Balance risks and opportunities
		Operational issues	• The operational picture: Requirements, hopes and concerns from military • Command and Control aspects of HAT • Assessing risks and benefits of HAT: Performance and Effectiveness Metrics	Map future directions
		Legal and ethical questions.	• Ethical, Legal and Social aspects of HAT: Values, Authority, Trust, Responsibility, Accountability, morality	

The aims and objectives of the NATO SCO HFM-300 Symposium, as described above in Table 2, set out in the meeting Call for Papers, and summarised in the Symposium Programme notes, were reiterated at the meeting by the NATO HFM Panel representative and Symposium Co-Chair, Dr Flemisch, DEU in his keynote. They include a description of the vision and goals, identification of the means for achieving the objectives, described in terms of barriers and perspectives, and a summary of the desired outcomes. It would have been appropriate and valuable to hear directly also the thoughts and aspirations of the other Symposium Co-chair Dr Mark Draper, USA. Dr Draper was Chair of the recent 2014-2018 NATO HFM 247 HAT Task Group, and has prior relevant experience spanning multiple NATO HFM Task Groups and Symposia on automation and autonomy. Dr Draper willingly shares his inspiring and motivating vision of HAT research informed from a military operations background perspective, and as a leading Human Effectiveness researcher, specialising in Multi UxV Control at the US Air Force Research Laboratory [23, 24]. Regrettably, in my humble opinion, an opportunity lost to envision future military requirements, risks, benefits and directions for Human Autonomy Teaming.

Table 3: Programme Perspectives Contributions

PERSPECTIVES CONTRIBUTIONS		
HFM RSY 300 PERSPECTIVES		**CONTRIBUTION PRIMARY PERSPECTIVE**
Overall system	• The big picture: Human autonomy teaming and cooperation in defence from a strategic perspective	HFM. Flemisch: Issues, drivers, risks.
	• Function division and relationships /roles (who does what functions, and in what kind of relationship?)	KN2. Hancock: Automation, human roles, balance.
Technological factors	• Teaming and cooperation requires sufficient autonomy: Autonomous capabilities that support teaming	7. Baltzer: Collision avoidance. 8. Neerincx: Patterns, work agreements. 9. Kelley: Search planning. 11. Luchero: Information provenance. 14. Diggelen: Social intelligence. 15. Thomas: Policy management. P1. Coronado: Autonomics framework. P2. Coronado: Task management.
Human factors	• Human Machine interaction, communication & cooperation	4. Johnson: Interdependency analysis. 10. Murphy: Imagery analysis. 18. Earthy: Ergonomics standards. 20. Kersholt: Social relationships. P3. Zelnio: Imagery analysis.
	• Comprehension (how do humans & autonomy understand one another?)	6. Shively: Communication dialogue. 12. Nirenburg: Natural language. 21. Chen: Transparency, trust. 23. Miller: Displaced transparency.
	• Joint Learning/Training (collaborative learning, crew resource management etc)	2. Roberts: HFI Organisational risks.
Operational issues	• The operational picture: Requirements, hopes and concerns from military	KN1. Amison: Military requirements.
	• Command and Control aspects of HAT	1. Rohling: Command management. 5. Frost: Functional division. 13. Schmitt: Manned-unmanned teaming.
	• Assessing risks and benefits of HAT: Performance and Effectiveness Metrics	3. Keirl: Effectiveness metrics.
Legal and ethical questions.	• Ethical, Legal and Social aspects of HAT: Values, Authority, Trust, Responsibility, Accountability, Morality	17. Koch: Ethical, legal arguments. 22. Butcher: Trust attitudes.

The response to the meeting call, summarised in Table 3, whilst not overwhelming, provided sufficient offerings for the Programme Committee to form a relevant, broad and interesting programme. In total, there were 21 presented contributions, 2 Keynotes and 3 Poster papers. The programme was able to address the challenges by bridging science and applications across domains, with strong contributions in key areas, offering a range of perspectives. Key perspectives attracted multiple contributions, including Technological factors (8), Human factors (11) and Operational issues (5). Understandably, many contributions covered

multiple perspectives, whilst some more narrowly focussed on specific concerns. Thankfully, there were only two cancellations to the published programme. This included one unfulfilled invited contribution to a wider perspective Panel covering Legal and Ethical issues, proposed by the HFM Panel Representative. Three deep technical offerings, with relatively narrower focus, were judged more suitable for closer individual scrutiny best delivered directly as poster papers. Incomplete reserve offerings – marginally relevant research objectives, plans only, without results - were included late to fill programme gaps, not necessarily enhancing the programme effectiveness. The meeting was privileged to receive two excellent Keynote presentations providing diverse, but complementary, Military and Academic expert perspectives, both highly valued and much appreciated by the audience.

Table 4: Programme Summary

HAT PROGRAMME SESSIONS TECHNICAL QUALITY ASSESSMENT MEANS, RANKS & PERFORMATIVE EFFECTS									
Session 1		**Session 2 & 3**		**Session 4**		**Session 5 & 6**		**Panel**	
Operation Requirements & Measuring HAT Effectiveness		**Human Autonomy Teaming Structure (Parts 1 &2)**		**Autonomous Capabilities That Support Teaming**		**HAT Interaction and Design**		**HAT Institutional Integration**	
1 Rohling TG247 Command management	TQ 5.00 P10 Performed effects	5 Frost TG247 Functional division	TQ 5.84 P3 Performed effects	10 Murphy Imagery analysis	TQ 5.67 P6 Performed effects	13 Schmitt TG247 Manned – unmanned teaming	TQ 6.00 P1 Performed effects	17 Koch Ethical, legal arguments	TQ 3.68 P19
2 Roberts HFI organisation risks	TQ 3.76 P18	6 Shively TG247 Communication dialogue	TQ 4.45 P16 Recognised effects	11 Luchero TG247 Information provenance	TQ 5.97 P2 Reasoned effects	14 Diggelen Social intelligence	TQ 5.48 P8 Reasoned effects	18 Earthy Ergonomics standards	TQ 4.67 P13 Reasoned effects
3 Keirl TG247 Effectiveness metrics	TQ 5.12 P9 Performed effects	7 Baltzer TG247 Collision avoidance	TQ 4.53 P15 Performed effects	12Nirenburg Natural language	3.63 P20	15 Thomas TG247 Policy management	TQ 5.80 P4 Performed effects	20 Kersholt Social relationships	TQ 4.27 P17 Reasoned effects
4 Johnson Inter-dependency analysis	TQ 4.80 P12 Recognised effects	8 Neerincx TG247 Patterns, work agreements	TQ 4.97 P11 Recognised effects			21 Chen TG247 Transparency, trust	TQ 4.60 P14 Performed effects		
		9 Kelley Search planning	TQ 5.63 P7 Reasoned effects			22 Butcher Trust attitudes	TQ 3.33 P21		
						23 Miller TG247 Displaced transparency	TQ 5.70 P5 Reasoned effects		
TQ Aggregate	**4.67**		**5.08**		**5.09**		**5.15**		**4.20**

The resultant programme of 21 submitted presentations, shown in Table 4 above, was performed over two days, divided into four themes, delivered in six Technical Sessions, plus one Panel, the later intended to include invited Legal and Ethical contributions, and entitled Institutional Integration. The structure was a balance between the HFM Panel vision and aspirations for a Symposium covering multiple perspectives, and the reality of the submissions received and the quality of scientific evidence provided. The presentations relied extensively and appropriately on 11 contributions provided from the HFM-TG-247 HAT Task Group, drawing upon Nations' leading relevant programme activities, and 10 additional offerings from activities either closely connected to the work of HFM-TG-247, or from the wider scientific community.

The programme balance was mixed, judged in terms of the full spectrum of targeted perspectives. Fundamentally, as required, the balance is a mix bridging science, applications and operations. It was relatively strong on technological and human factors, with a fairly representative, but not complete or deep coverage, of operational issues. However, there was only a weak response on legal and ethical questions. This is not entirely surprising, given that the event targeted a mostly scientific and technological, rather than legal, community. The Programme breakdown shows estimates of the assessed Technical Quality (TQ), and rated rank position order (P1-21), and the nature of the reported effects, Performed, Reasoned, or Recognised.

The aggregated estimates of Technical Quality for the 7+1 programme sessions showed that the HAT Integration and Design Sessions, with 6 papers, obtained the highest aggregate rating with TQ 5.15. In contrast, the HAT Institutional Integration Panel session with only 3 papers, and one cancelled, scored the lowest aggregate rating of TQ 4.20, close to mid scale. In between, the Operation Requirements and Measurement Session with 4 papers was rated TQ 4.67, the Human Autonomy Teaming Structure sessions, with 5 papers, was rated TQ 5.08, and the Autonomous Capability that Support Teaming Session was rated TQ 5.09. This represents a reasonable balance of above average, good quality across the Technical Sessions groupings, with all sessions rated above the rating scale nominal mid-point (4.00). The weakest rated session, the Panel on HAT Institutional Integration, a somewhat artificial thematic collective, intended to cover legal and ethical issues, had the weakest technology capability development content.

The timing schedule with gaps allowed good time for welcome questions, dialogue and discussion of members of the audience with presenters, and provided opportunity for the Co-Chair Panel Representative to respond and share at length his opinions and perspective. Disappointingly, the series of three optional Workshops, scheduled for the morning of the third and final day, were not performed due to weak participant support. The proposed Workshop themes – Operational Assessment RTAG, Surprise and Startle, Future AWACS - were with hindsight perhaps too esoteric for the wider target audience, and evidentially not sufficiently appealing to counter the maritime attractions of not-so-sunny Southsea and Portsmouth. This indicates possibly, that for future reference, a more integrated approach to Workshop scheduling, embedded within the schedule of presentations, might be more popular, accessible and effectively delivered.

4.0 TECHNICAL EVALUATION

4.1 Evaluation Approach

The primary function of the TER is to represent and critically review the state of the art, focusing on the scientific evidence for HAT, and to highlight directions for future travel. The scientific evidence offered by the research community in this diverse body of work is essentially proof of concepts, models and design requirements, including evidence on performative effects, and argumentation on costs and benefits, risks and opportunities of HAT. The general aim of the evaluation was to offer a basis for reflecting critically on the strength of scientific evidence provided, be it argumentative or factual proof, conflicting or corroborating, converging or diverging, or confirmatory or disproving scientific evidence. From this analysis, it is expected to emerge evidence of barriers and gaps, indications of challenges and opportunities, and lines for development with likely directions for future new work.

The technical evaluation process utilised an assessment approach combining expert opinion value judgements on the quality of Symposium presentations, inevitably subject to bias and selection sampling error, coupled with post-hoc, deep technical analysis of the documentation prepared for publication, providing corroborating evidence based on factual accuracy. The Symposium was characteristically rich on positive academic rhetoric and enthusiastic supportive argument concerning the value and necessity of HAT solutions. Disappointingly, the proceedings were generally less strong, and some relatively poor, on providing hard, objective, factual evidence to validate and prove the risks and benefits of teaming concepts advocated for improved performance and effects. This imbalance between rhetoric and evidence is perhaps characteristic of popularism, but equally

indication of the relative immaturity of the research in the domain. This twin-track critical evaluation approach was provided to some extent so as to temper, and if necessary to counter balance, the prevailing positivistic enthusiasm inflation. It offers a broadly based, comparative appraisal of the symposium technical quality and state of the art, based on a variety of relevant assessment dimensions, with verification and validation of veracity from diverse perspectives. The intention was to identify rationale for consensus in expert opinions rather than to illustrate and understand divergence. In attempting to quantify the qualitative value of the work, the focus was on representing comparative rather than absolute value, supported where possible by relevant theory, analysis, rationale and factual evidence.

TECHNICAL ANALYSIS	TECHNICAL QUALITY					
	Operational Relevance	Scientific Quality	Exploitability	Technical Capability	Innovation	AGGREGATE TECHNICAL QUALITY
Wider System RSGDEC	★					★
Performed Effects		★				★
Technology Readiness Level			★		★	★
Human Readiness Level			★			★
Human Autonomy Teaming				★		★
Decision Models REMEDE				★		★
Capability Maturity Level				★		★

Figure 3. Technical Evaluation Assessment and Analysis Matrix

4.2 Technical Evaluation Assessment Team

At the Symposium Programme Planning Meeting in Sophia, the HFM Panel Representative indicated a preference for an independent and impartial Technical Evaluation Report. The proposed TER volunteer was a serving member of the Symposium Programme Committee and a long-time active advocate of HAT. Partly in response to these concerns, and as a demonstration of good practice, a small multi-national team of subject matter expert volunteers was assembled by invitation of the agreed TER lead, to share responsibility and support the TER task.

The TER Team primary role was to provide a small but representative subset sample of qualitative assessments of the NATO HFM 300 HAT Symposium papers' presentations. The HAT TER Team individual assessors were drawn from members of the NATO SCO HFM Task Group 247 on HAT, 2014-2018. All served on the

HFM SY 300 HAT Symposium Programme Committee. Additionally, most had participated relatively continuously in the relevant precursor HFM Automation/Autonomy Task Groups, HFM-TG-017 Augmenting the Force (2002-2005), and HFM-TG-170, Supervisory Control, (2007-2012). The HAT TER Team can be considered to be recognised all as respected Nations' technical leaders in the field, bringing a wealth of relevant knowledge and experience to the task of HAT technical evaluation. Thus, for the present purposes, they are considered to be most credible and reliable judges of technical quality of the Symposium contributions, albeit with a positive bias towards supporting HAT research. The HAT TER team members were as follows:

- Professor Gilles Coppin - Head of Laboratory, STICC/CNRS UMR 6285, Telecom Bretagne, Breast, France.
- Antony Grabham – Senior Scientist, Air Systems Autonomy, Platform Systems Division, Defence Science and Technology Laboratory, DSTL Portsdown West, Fareham, UK.
- Dr Chris Miller - Senior Scientist, Smart Information Flow Technologies, Minneapolis, USA.
- Univ.-Prof. Dr. Ing. Axel Schulte, Dean, Aerospace Engineering Department, Bundeswehr University, Munich, Germany.
- Dr. Rojier Woltjer – Deputy Research Director, Swedish Defence Research Agency, FOI, Stockholm, Sweden.

Additional assistance, filling in some assessment gaps, was provided by Dr. Timo Rohling, Fraunhofer FKIE, Germany. Invaluable assistance in distribution and management of the assessment protocols, and in the resultant data compilation, was provided by William R. Ellis, Engineer, Air Systems Autonomy, Platform Systems Division, Defence Science and Technology Laboratory, DSTL UK.

The assessments were performed using a standard Dstl structured Technical Quality assessment protocol. The protocol captured the individual assessor's appraisals of generic operational, technical and scientific attributes of the work, using a simple 7-point Likert low-high rating scale. The assessed attributes and definitions were as follows:

- Operational relevance – Significance of impact on military operations
- Exploitability – Likelihood of effective exploitation
- Technical capability – Builds critical S&T capability to meet future needs
- Scientific quality – Use of sound approach, methods and techniques
- Innovation – Pushing boundaries and production of something new and novel

NATO HFM 300 HUMAN AUTONOMY TEAMING – TECHNICAL EVALUATION REPORT NAME OF TECHNICAL ASSESSOR:							
Paper/Poster Number: Authors: Title:							
TECHNICAL QUALITY	Low ASSESSMENT RATING High						
	1	2	3	4	5	6	7
Operational relevance – Significance of impact on military operations							
Comments:							
Exploitability – Likelihood of effective exploitation							
Comments:							
Technical capability – Builds critical S&T capability to meet future needs							
Comments:							
Scientific quality – Use of sound approach, methods & techniques							
Comments:							
Innovation – Pushing boundaries & production of something new and novel							
Comments:							
Summary :							

Figure 4. Technical Evaluation Assessment Rating Protocol

4.3 Technical Assessment Results

The Technical Evaluation assessment ratings provided by the subject matter expert SME team of reviewers, judged on the Low 1-7 High rating scale, were aggregated across the five technical components. This provided a Technical Quality (TQ) estimate for the collective endeavour, averaged across the assessment categories, of TQ 4.95, with a range of TQ 3.33-6.00 varying across the 21 assessed contributions.

Table 5: Results of Technical Evaluation Team Assessment Ratings

No.	Author
1.	Rohling
2.	Roberts
3.	Keirl
4.	Johnson
5.	Frost
6.	Shively
7.	Baltzer
8.	Neerincx
9.	Kelley
10.	Murphy
11.	Luchero
12.	Nirenburg
13.	Schmitt
14.	Diggelen
15.	Thomas
16.	Cancelled
17.	Koch
18.	Earthy
19.	Cancelled
20.	Kersholt
21.	Chen
22.	Butcher
23.	Miller

Technical Evaluation Mean Ratings (No. – Mean)

Rank Order	Operational Relevance	Exploitability	Technical Capability	Scientific Quality	Innovation	AGGREGATE TECH QUALITY
1	10 - 6.33	18 – 6.33	5 – 6.40	11 - 6.33	23 – 6.50	13 – 6.00
2	15 – 6.33	15 – 6.33	9 – 6.33	13 – 6.20	11 – 6.33	11 – 5.97
3	13 – 6.20	13 – 6.20	13 – 6.20	5 – 5.80	9 – 5.67	5 – 5.84
4	5 – 6.00	23 – 6.00	11 – 6.00	10 – 5.67	14 – 5.60	15 – 5.80
5	11- 6.00	10 – 5.83	15 – 6.00	23 - 5.50	5 – 5.20	23 – 5.70
6	3 – 5.80	5 – 5.80	8 – 5.50	14 – 5.40	13 – 5.20	10 – 5.67
7	1 – 5.80	14 – 5.60	14 - 5.40	15 – 5.33	10 – 5.17	9 – 5.62
8	17 - 5.50	3 – 5.60	3 – 5.40	7 – 5.25	20 – 5.00	14 – 5.48
9	23 - 5.50	1 – 5.60	1 - 5.40	8 – 5.17	15 – 5.00	3 – 5.12
10	14 - 5.40	9 – 5.33	10 – 5.33	21 - 5.00	12 – 4.83	1 – 5.00
11	21 - 5.33	11 – 5.17	21 – 5.00	9 - 5.00	8 – 4.50	8 – 4.97
12	9 – 5.33	4 – 5.00	4 – 5.00	12 – 5.00	4 – 4.40	4 – 4.80
13	4 – 4.80	7 – 4.75	23 – 5.00	4 – 4.80	3 – 4.40	18 – 4.67
14	2 – 4.80	8 – 4.67	6 – 5.00	18 - 4.67	18 – 4.33	21 – 4.60
15	6 – 4.75	2 – 4.40	7 – 4.08	20 - 4.67	7 – 4.08	7 – 4.53
16	18 – 4.67	21 – 4.33	20 – 3.67	3 – 4.40	17 – 4.00	6 – 4.45
17	8 – 4.67	6 – 4.25	18 – 3.33	6 – 4.25	1 – 4.00	20 – 4.27
18	7 – 4.50	20 – 4.00	2 – 3.20	2 – 4.20	6 – 4.00	2 – 3.76
19	20 – 4.00	22 – 3.67	12 – 3.00	22 – 4.00	21 – 3.33	17 – 3.68
20	22 – 3.67	17 – 3.50	22 -2.67	1 – 4.00	22 – 2.67	12 – 3.63
21	12 – 2.67	12 – 2.67	17 – 2.50	17 – 2.00	2 – 2.20	22 – 3.33

The aggregated ratings of Technical Quality were mostly above the scale nominal mid-point value of TQ 4.00 (17/22). This indicates a good standard of Technical Quality ranging across a wide range of differing technical contributions, with relatively few exceptions judged to be of marginally value and relevance.

Not all contributions received the same maximum level of scrutiny from 5 TER Team Assessors. This was due to changes in the availability of the volunteer assessors which reduced as the meeting progressed. Individual assessors' ratings of items showed variability over 2-3 scale point intervals. The individual assessors' scores showed some differences in the levels of sensitivity and discrimination exercised, typical for subjective judgements on complex, knowledge-based tasks. Assessors tended to rate the set of contributions consistently at relatively lower or higher levels on the scale, indicating baselining differences. The variability reflects the required independence in scoring, and includes subjectivity in scale baselining, calibration and interpretation. Consequently, the ratings results are aggregated across individual assessors and across assessment categories, averaged for simplification and presentation, but not to imply normative distributions. They should be interpreted with due caution, respecting sampling error associated with such relatively small sample sizes. The plotted data show that the averages match the central tendencies in the data presented, and none appear to be overly influenced by extreme outlying exceptional values. The data can be treated as indicative of the comparative value of contributions, and not as factual accuracy, but used for judging quality in relative terms rather than absolute values. For the purposes of evaluation, the aim was to look for consistency within and between assessors, and to find general trends and patterns, seeking to inform and guide a comparative qualitative assessment, rather than focusing on differences in absolute assessment values.

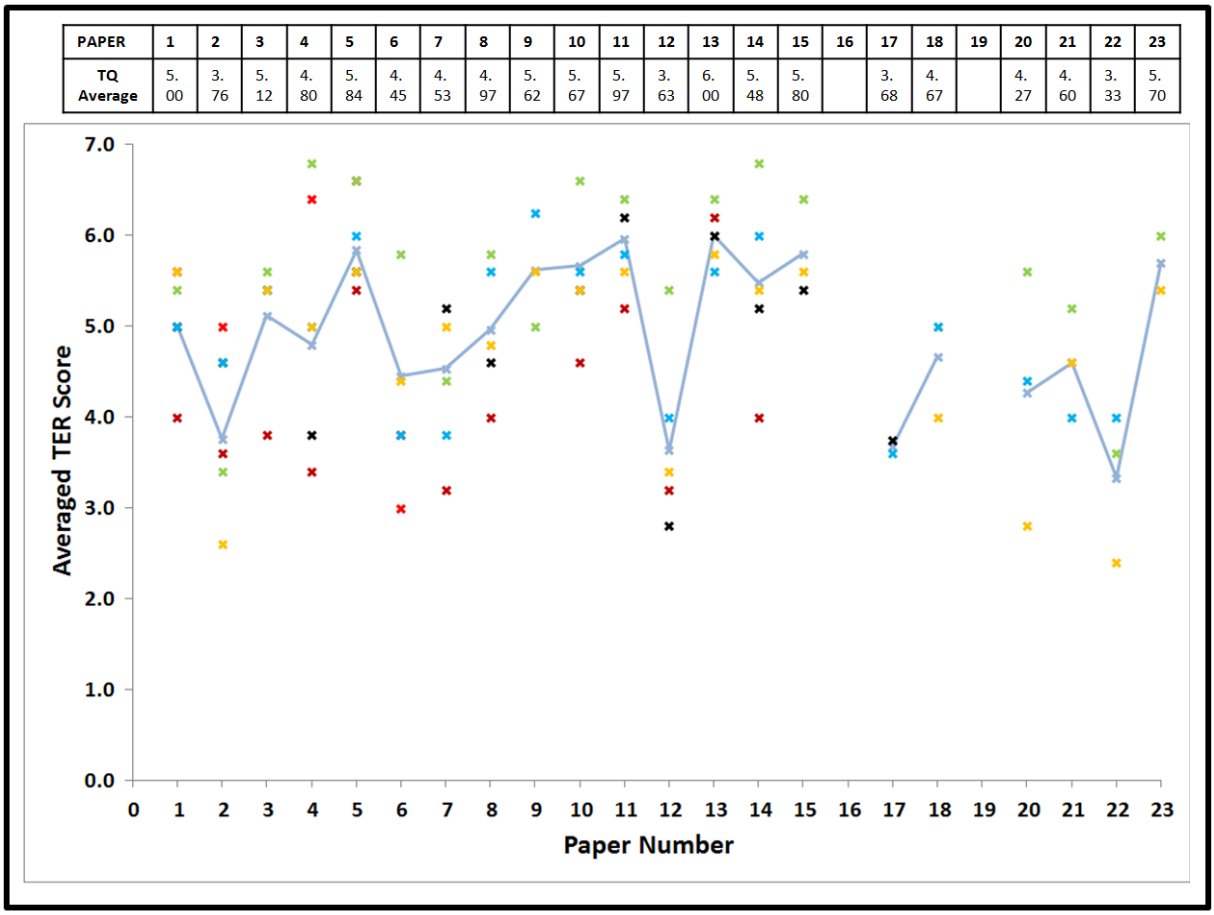


Figure 5: Averaged Ratings for Symposium Contributions

The underlying pattern evidenced a reasonable degree of inter-rater consistency and agreement. The level of agreement was considered sufficient to draw general conclusions with reasonable confidence over the relative strengths and weaknesses, value and impact of the assessed contributions.

There was close inter-rater agreement, spanning 1 scale interval point, on the strongest assessed contributions, specifically Manned-Unmanned Teaming (13. Schmitt, TQ 6.00), Information Provenance (11. Luchero, TQ 5.97), MUxV Functional Division (5. Frost, TQ 5.84), Policy Management (15. Thomas, TQ 5.80) and Displaced Transparency (23.Miller, TQ 5.70).

Close agreement was found for the least strong assessed contributions, indicating floor and ceiling clustering effects. The widest range of assessments, and least inter-rater agreement, spanning 3 scale interval points, was for intermediate strength contributions. These included notably Interdependency Analysis (4. Johnson, TQ 4.80), Communication Dialogue (6. Shively, TQ 4.45), and Social Relationships (20. Kersholt, TQ 4.27). Most intermediate strength contributions exhibited in common soft analytical approaches, without strong technical and operational attributes. While these contributions were all theoretically consistent and logically sound, they all reported without performed effects, with effects recognised or reasoned only. They lacked factual scientific evidential proof of efficacy and effectiveness.

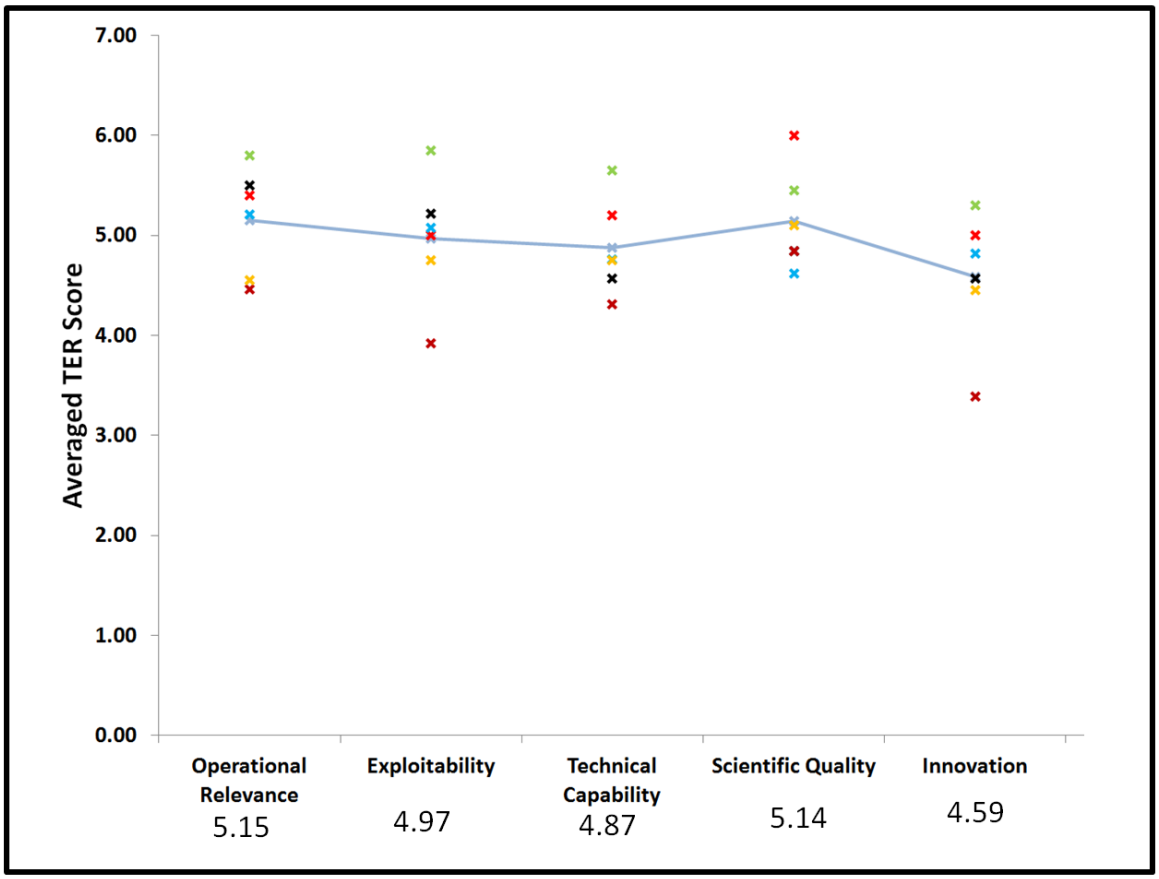


Figure 6: Averaged Ratings for Assessment Categories

The averaged ratings for the five technical assessment categories exhibited further evidence of broad inter-rater consistency and agreement across the individual SME evaluators, ranging across a span of 1-2 scale point intervals. Average ratings for Operational Relevance (OR 5.15), Technical Capability (TC 4.87) and Scientific Quality (SQ 5.14) were consistently above mid scale rating. Similarly, average ratings for Exploitability (Exp 4.97) and Innovation (Inn 4.59), with two exceptions, tended towards the high end of the rating scale.

4.4 Technical Assessment Analysis

Following the meeting, the available documentation on the contributions was analysed in depth by the TER author. This analysis was conducted using a set of HAT specific, Dstl technical assessment frameworks and technical evaluation criteria. The aim of the analysis was to provide additional detailed factual, corroborating evidence, as back up to assessors' subjective appraisals. They provided focussing on the status of the contributions, expressed in more precise technical terms. This includes common distinguishing features, strengths and weaknesses, underlying patterns and trends, priorities, risks and opportunities, ultimately seeking means to illustrate the direction of travel.

4.4.1 Operational Relevance

Operational relevance refers to the significance of impact on military operations. The military relevance of Uninhabited Military Vehicles (UMVs) was considered in 2007 by the NATO HFM Task Group 017 on Human Factors on Augmenting the Force, with the rationale for a wide the range of applications extensively documented in the TG-017 Final Report [25]. Ideally, the Military operational relevance and impact of HAT

concepts needs to be judged in terms of detailed applications performed in context of specific military missions and tasks. A clinical analysis of the applicability of HFM TG 247 related HFM SY-300 Symposium contributions is reported in the Military Relevance Chapter of the of NATO SCO HFM HAT Task Group 247 2018 Final Report [26]. Three mission use cases were investigated, namely Intelligence, Surveillance and Reconnaissance (ISR), Base Protection and Troops in Contact (TIC). Five HAT capabilities were available for examination by HFM TG 247, all of which were reported in contributions to the SY 300 HAT Symposium:

- UK1 Dstl/QinetiQ Adaptable Autonomy and Policy Management (15. Thomas);
- US1 SPAWAR Task Management and Plan Monitoring (11 Luchero; P2 Coronado);
- US2 AFRL IMPACT MUxV Planning and Control (5. Frost);
- US3 NASA Bi-directional Communications (6. Shively);
- US4 USARL Transparency (20. Chen).

The analysis identified utility levels judged for HAT capabilities in phases of the mission use cases, their detailed application and benefit, and principal mission risks probability and severity. Task management, plan monitoring, policy management and transparency were judged most frequently as high utility capabilities.

Table 6: HFM TG 247 Military Relevance Mission Analysis

MILTARY USE CASE MISSION	UTILITY	HFM TG 247 HAT TECHNICAL ACTIVITY				
		UK1 Adaptable Autonomy & Policy Management	US1 Task Management & Plan Monitoring	US2 IMPACT MUxV Planning & Control	US3 Bidirectional Communication	US4 Transparency
ISR Intelligence, Surveillance, Reconnaissance	HIGH	8	8	1	4	6
	MED	0	2	4	0	0
	LOW	3	1	2	1	0
Base Penetration	HIGH	8	9	3	1	6
	MED	0	2	3	4	1
	LOW	2	0	1	0	0
TIC Troops in Contact	HIGH	9	9	1	0	0
	MED	0	1	2	0	3
	LOW	0	0	0	2	0
TOTAL		30	32	17	12	16

A variety of Military operations applications were covered in the HFM SY 300 HAT Symposium contributions. These include:

- Unmanned Vehicle Air/Land/Maritime Command and Control C2 (1, 2, 3, 5, 8, 11, 13, 15)
- Aircraft Manned-Unmanned Teaming (3, 13, 15)

- Multi UxV Coordination (1, 3, 5, 6, 8, 9, 11, 13, 15, 21, 23)
- Base Protection (1, 3, 5, 11, 14, 15, 21)
- UAS Tactical Operations C2 Airspace Management (3, 15, P2)
- Airborne Surveillance and Reconnaissance (1, 3, 13, 14, 15, 21, 23, P2)
- Land vehicle escort and overwatch (1, 5)
- Maritime Underwater Search (9)
- MSTAR (P3) and ISTAR PED FMV Imagery Analysis (10)
- Emergency and Disaster Response (8, 11)
- Casevac, Medivac, Troop Transportation (13)
- Land Manoeuvre Vehicle Operation transportation collision avoidance (4, 7)

Applications also covered included single-pilot Commercial Air Operations (6), Air Traffic Collision Avoidance (4, 6), Clinical Medicine (12), Recruitment (22) and RAIS Robotic Autonomous Intelligent Systems (18).

Considered in assessed order of merit, the averaged ratings for Operational Relevance (OR) of the 2018 NATO SCO HFM SY-300 Symposium contributions were the highest of the five assessment categories, at OR 5.15, with a range of averages between OR 2.67-6.33. 12 of the 21 total assessed contributions rated greater than 5.00 for Operational Relevance. Thus, the Operational Relevance of this diverse body of work was judged in general to be of a relatively high standard.

Operational Relevance – Top Ten
10. Murphy - Imagery analysis
15. Thomas – Policy management
13. Schmitt – Manned-Unmanned teaming
5. Frost – Functional division
11. Luchero – Information provenance
3. Keirl – Effectiveness metrics
1. Rohling – Command management
17. Koch – Ethical, legal arguments
23. Miller – Displaced transparency
14. Diggelen – Social intelligence

The most notable examples of strong operational relevance include the work on Imagery analysis (10. Murphy, OR 6.33); Policy management, including Negotiation (15. Thomas, OR 6.33); Manned-unmanned teaming (13. Schmitt, OR 6.20); MUxV Functional division (5. Frost, OR 6.00); and Information provenance (11. Luchero, OR 6.00). Manned-unmanned teaming also features strongly in the work on Effective metrics (3. Keirl, OR 5.80)

This relatively high valuation of the Operational Relevance of the work reflects the current high levels investment in autonomy technology, and the anticipated significant impact expected on military operations. It both confirms and justifies the strong military support for research on human effectiveness with autonomy technology.

A wider background analysis performed by the TER author post meeting, considered the mapping of the relationships of the contributing activities with respect to a generic military research and development system

framework for technical evaluation of capability requirements. This wider system research requirements framework - RSGDEC - comprises 6 components, namely Requirements Drivers, Cognitive Systems Engineering (CSE) Solutions, Functional Governance, Development Enablers, Performative Effects, and Enabled Capabilities. The resultant framework profile mappings for contributions should be regarded as nominal rather than absolute. They are intended to broadly differentiate rather than accurately define and measure relevant wider differences in requirements context, based on the evidence reported, indications and assumptions when information is lacking.

The RSGDEC background analysis indicated that 8 of the 11 HAT symposium contributions with highest rated aggregate Technical Quality (Paper Numbers 1, 3, 5, 8, 10, 11, 13, 15) had either a complete (6/6), or near complete (5/6) RSGDEC Mapping Profile. This finding can be considered as evidence of the benefits for research technical quality and delivery of a research procurement environment driven by a comprehensive understanding of military capability requirements.

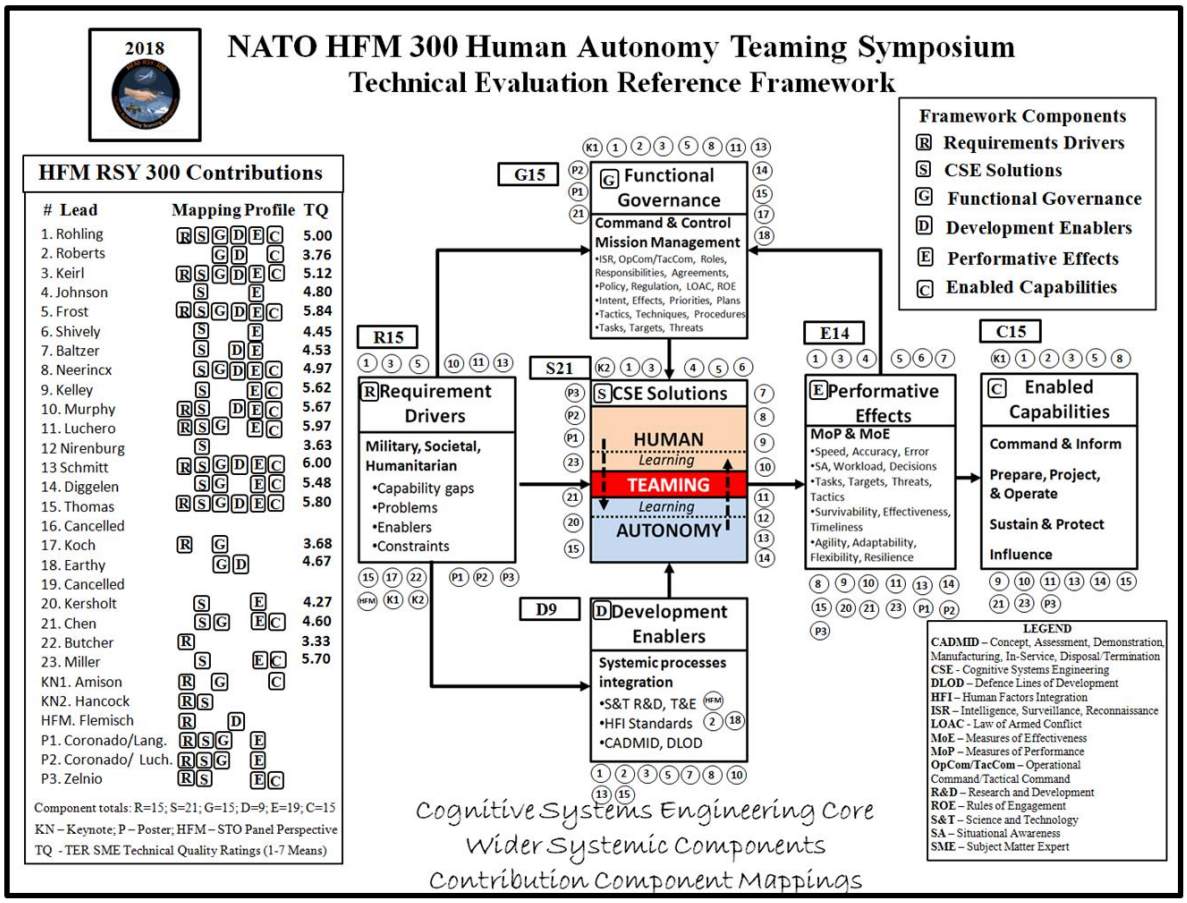


Figure 7: Results of Wider System Research Requirements Analysis

Wider Context Top Ten

15. Thomas – Policy management **RSGDEC** (OR 6.33)
13. Schmitt – Manned-Unmanned teaming **RSGDEC** (OR 6.20)
5. Frost – Functional division **RSGDEC** (6.00)
1. Rohling – Command management **RSGDEC** (OR 5.80)
3. Keirl – Effectiveness metrics **RSGDEC** (OR 5.80)
10. Murphy - Imagery analysis **RS_DEC** (OR 6.33)
11. Luchero – Information provenance **RSG_EC** (OR 6.00)

8. Neerincx – Patterns, work agreements **_SGDEC** (OR 4.67)
14. Diggelen – Social intelligence **_SG_EC** (OR 5.40)
21. Chen – Transparency, trust **_SG_EC** (OR 5.33)

Comparative analysis indicates that relatively strong wider system research requirements RSGDEC profiles are strongly associated with high operational relevance. The results showed that 8 of the Top Ten Operational Relevance (OR) contributions had strong RSGDEC profiles (10, 15, 13, 5, 11, 3, 1, 14).

The value of strong military support foundations for HAT research is also evidenced by the level international participation in the NATO HFM RSY 300 HAT symposium meeting, and sustained during the pre-cursor 3-year NATO HFM HAT Task Group 247. Representatives from the NATO Nations' leading military technology research laboratories and research programmes on autonomy systems technology and human effectiveness have taken a full and active part in both of these NATO SCO HAT related activities. It is noteworthy that 11 of the 21 symposium papers, plus 2 poster contributions, were provided by authors and co-authors who were active participants in HFM TG247 on HAT. This indicates the value of effective NATO SCO HFM Task Group activities in forming, evolving and delivering strong productive collaborative research through communities of specialist interest.

4.4.2 Scientific Quality

Scientific quality refers to the use of a sound approach, methods and techniques. A variety of research methods techniques were applied in the contributions. These included the following:

- Computer based modelling and simulation testing,
- Laboratory-based controlled variable prototype experimental test methods,
- Single operator/user/subject participant test trials,
- Military operator and University student test participants,
- Multiple operator system testing,
- Small and large scale simulations,
- Virtual Synthetic Environment (SE) test and evaluation,
- Live/Virtual/Constructive (LVC) mixed,
- Field test environments.

When evaluating scientific quality, the focus is on the choice of research approach, methods and techniques. This is judged typically in terms of theoretical consistency, levels of proof, methodological rigour, robustness and reliability. Testing rigour includes control of variables, sensitivity and discrimination of measurements and metrics, sampling variance and error, and diagnostic, prognostic and predictive power.

Appropriately for an S&T community, Scientific Quality (SQ) averaged second highest assessment ratings at 5.14, marginally behind Operational Relevance, with a range of 2.00-6.33. 12 of the 21 contributions were assessed with Scientific Quality equal or greater than 5.00. Thus, the Scientific Quality of this diverse body of work was judged generally to be relatively high.

Scientific Quality Top Ten
11. Luchero – Information provenance
13. Schmitt – Manned-Unmanned teaming
5. Frost – Functional division

10. Murphy - Imagery analysis
23. Miller – Displaced transparency
14. Diggelen – Social intelligence
15. Thomas – Policy management
7. Baltzer – Collision avoidance
8. Neerincx – Patterns, work agreements
21. Chen – Transparency, trust

The most notable examples of good SQ include the work on the following: Information provenance (11. Luchero, SQ 6.33); Manned-unmanned teaming (13. Schmitt, SQ 6.20); Functional division (5. Frost, SQ 5.80); Imagery analysis (10. Murphy, SQ 5.67); Displaced transparency (23. Miller, SQ 5.50).

Of both interest and concern, examined from a technical evaluation perspective with particular regard to scientific quality, are the variations in the levels of proof, or scientific factual evidence, provided by the contributions. Levels of proof, obtained through testing, are needed in support of the validity, efficacy and/or effectiveness of the particular HAT concept, hypothesis, contention, design proposition or problem solution. Consideration of the performative effects of HAT proposals and solutions is essential for HAT proof of concept. It is not a foregone conclusion that all forms of HAT are inherently and necessarily beneficial. Proof of HAT performance is needed to weigh and balance correctly the advantages and disadvantages, pro's and con's, biases and performance variable trades that are needed, and indeed essential, viewed from a control system perspective. We need to know if, why, when, where, and how HAT is superior to HHT, AAT, or indeed HA, HH, AA without teaming. The work on Functional division (5. Frost, SQ 5.80), describes use of a novel human Confederate simulation approach to an experimental comparison of performance of Human-Autonomy Team versus Human-Human Team for MUxV C2. The work contrasts an organisation based on an operator hierarchical structure compared with one based on assigned roles, contrasting operator driven versus role driven structures. This relates closely to earlier UK Dstl/QinetiQ work on UAS C2 organisation structure (ATLAS UAS SE Trial, Farnborough, Jan 2012) contrasting the traditional person-to-platform organisation with a person-to-purpose, flexible "cloud" services architecture [20]. Control of system performance judged in terms of goals, missions, functions and tasks, and ultimately effects, is essential for successful evolution of the management, optimisation and ultimately adaptiveness in HAT capability maturity. For the purposes of discrimination and analysis, performative effects reported in the authors' contributions were classified nominally as either Recognised, Reasoned or Performed, in order indicative of increasing level of proof provided, or alternatively, by default, Not Considered.

Performed Effects Reported
1. Rohling – Command management SQ 4.00
3. Keirl – Effectiveness metrics SQ 4.40
5. Frost – Functional division SQ 5.80
7. Baltzer – Collision avoidance SQ 5.25
10. Murphy – Imagery analysis SQ 5.67
13. Schmitt – Manned-Unmanned teaming SQ 6.20
15. Thomas – Policy management SQ 5.33
21. Chen – Transparency, trust SQ 5.00
P1. Zelnio – Imagery analysis

Performed Effects Reasoned
9. Kelley – Search planning SQ 5.00
11. Luchero – Information provenance SQ 6.33
14. Diggelen – Social intelligence SQ 5.40
20. Kersholt – Social relationships SQ 4.67
23. Miller – Displaced transparency SQ 5.50
P2. Coronado/Luchero – Task management

Performed Effects Recognised

4. Johnson – Interdependency analysis SQ 4.80
6. Shively – Communication dialogue SQ 4.25
8. Neerincx – Patterns, work agreements SQ 5.17
P1. Coronado/Lange - Autonomics framework

Performed Effects Not Considered
2. Roberts – HFI Organisational risks SQ 4.20
12. Nirenburg – Natural language SQ 5.00
17. Koch – Legal, ethical arguments SQ 2.00
18. Earthy – Ergonomics standards SQ 4.67
22. Butcher – Trust attitudes SQ 4.00

This nominal analysis indicated that effects were reported as Performed in 9 of the total 21+3 poster contributions (1, 3, 5, 7, 10, 13, 15, 21, P3) with an average SQ 5.21, Reasoned with, but not performed, in 6 (9, 11, 14, 20, 23, P2) with an average SQ 5.38, and Recognised only in 4 contributions (4, 6, 8, P1) with an average SQ 4.74. Performative effects were not considered explicitly in 5 of the 21 contributions (2, 12, 17, 18, 22) with an average SQ 3.97. It is significant that 5 of the contributions reporting Performed Effects are included in the SQ Top Ten. Furthermore, nine of the authors reporting either Performed or Reasoned Effects are included in the aggregated Technical Quality Top Ten. Those authors reporting proof of performative effects should be congratulated. Others should be encouraged to do think seriously about doing so, as a scientific imperative, in future work on HAT.

4.4.3 Exploitability

Exploitability refers to the likelihood of effective exploitation. Averaged ratings Exploitability (Exp), provided the third assessment highest ratings at Exp 4.97, with a range of Exp 2.67-6.33, and with 12 contributions rated greater than Exp 5.00. Thus, likelihood of exploitation arising from this diverse body of work was judged to be generally high.

Exploitability Top Ten
18. Earthy – Ergonomics standards.
15. Thomas – Policy management.
13. Schmitt – Manned-Unmanned teaming.
23. Miller – Displaced transparency.
10. Murphy - Imagery analysis.
5. Frost – Functional division.
14. Diggelen – Social intelligence.
3. Keirl – Effectiveness metrics.
1. Rohling – Command management.
9. Kelley – Search planning.

The most notable examples of Exploitability include the work on the following: Ergonomics standards (18. Earthy, Exp 6.33); Policy management (15. Thomas, Exp 6.33); Manned-unmanned teaming (13. Schmitt, Exp 6.20); Displaced transparency (23. Miller, Exp 6.20); Imagery analysis (10. Murphy, Exp 5.83).

4.4.3.1 Technology Readiness

A range of factors can be considered to influence the likelihood of exploitability, including relevance, maturity and readiness. One relatively simple method for judging exploitability of HAT research Symposium contributions is to consider the individual contributions in terms of Technology Readiness Levels. Technology

Readiness Levels (TRL), originally developed by NASA, now widely used in the Defence Equipment Procurement community, provide a type of measurement system to assess the maturity of a particular technology. Using TRLs, each technology project can be evaluated against the parameters for each technology level, and assigned a TRL rating based on the projects progress. There are nine technology readiness levels, ranging from TRL 1 as the lowest level, to TRL 9 as the highest level. When a technology is at TRL 1, scientific research is beginning and the results are being translated into future research and development. At the opposite end of the scale, TRL 9 applies when a technology has been fully developed and tested, or "flight proven" during a successful mission. The 9 TRL levels progress through distinct stages of development and test environments: commencing with basic technology research, rising through research to prove feasibility, then technology development and technology demonstration, then rising through technology system/sub-system development, ultimately rising to technology system test, qualification and operation. Detailed definitions are provided for levels hardware and software, with descriptions of supporting evidence.

The Symposium contributions were assessed using the UK MoD TRL definitions. This provided a range of assignments, ranging in from TRL 1 to TRL 7. None were judged to be at TRL 5, TRL 8 and TRL 9. Near equal numbers of contributions (3/4) were associated with the other TRL levels.

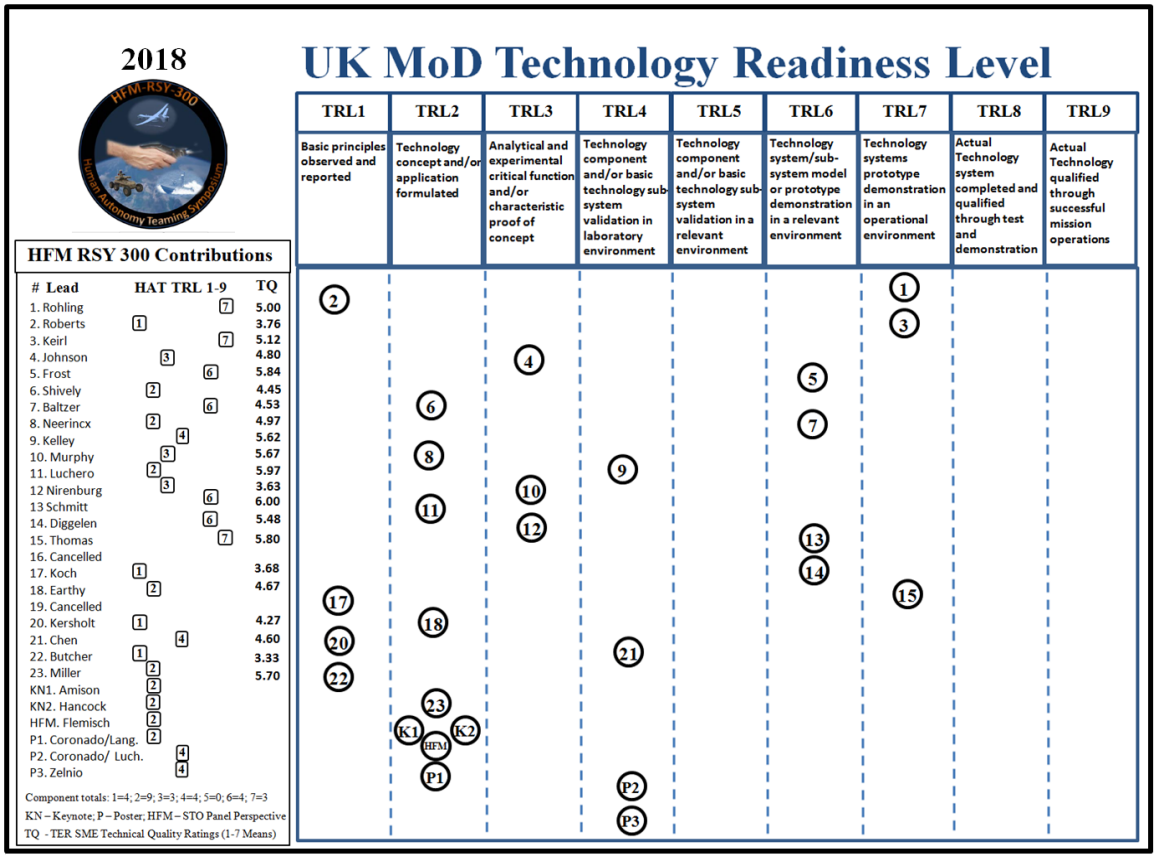


Figure 8: Results of Technology Readiness Levels Analysis

Technology Readiness Top Ten
TRL7 - Technology systems prototype demonstration in an operational environment
1. Rohling – Command management, Exp 5.60
3. Keirl – Effectiveness metrics, Exp 5.60
15. Thomas – Policy management, Exp 6.33

TRL6 - Technology system/sub system model or prototype demonstration in a relevant environment
5. Frost – Functional division, Exp 5.80
7. Baltzer – Collision avoidance, Exp 4.75
13. Schmitt – Manned-Unmanned teaming, Exp 6.20
14. Diggelen – Social intelligence, Exp 5.60
TRL4 - Technology component and/or basis technology sub system prototype validation in a laboratory environment
9. Kelley – Search planning, Exp 5.33
21. Chen – Transparency, trust, Exp 4.33
P2. Coronado/Luchero – Task management
P3. Zelnio – Imagery analysis

Comparative analysis indicates that high TRL work is strongly associated with high exploitability (1, 3, 5, 13, 14, 15) but that there can be exceptions where low TRL work of good TQ can be identified early as having potentially high exploitability (10, 18, 23).

4.4.3.2 Human Readiness

For balance, and good practice, the Symposium contributions were assessed using the Human Readiness Level (HRL) definitions developed relatively recently by US DoD [27, 28].

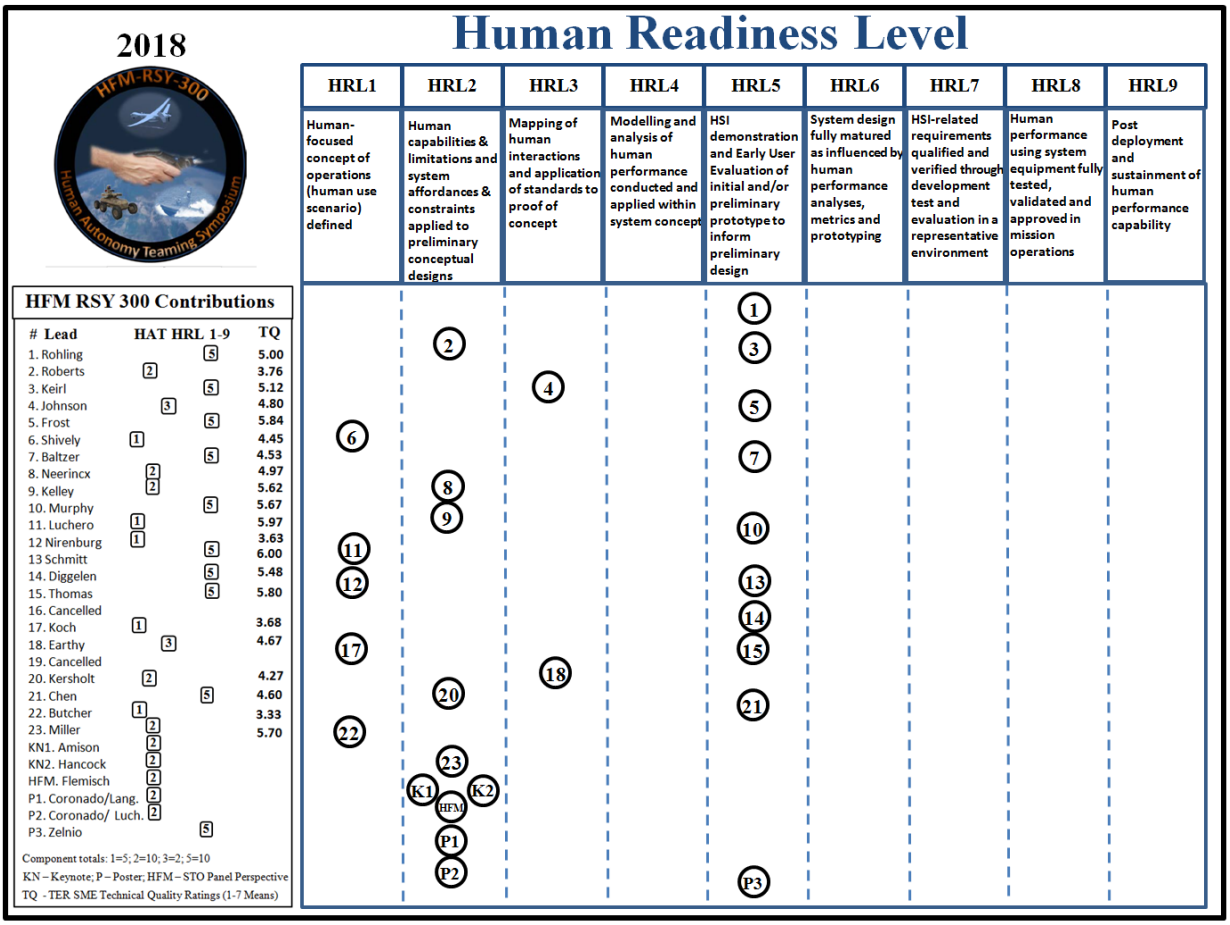


Figure 9: Results of Human Readiness Levels Analysis

Analysis provided a narrower range of mappings, ranging from HRL 1-5, falling considerably short of full readiness maturity development and testing. None of the contributions were judged to be HRL 4 or 6-9.

- 10 contributions (1,3,5,7,10,13,14,15,21) were judged to be at HRL 5: HSI demonstration and Early User Evaluation of initial and/or preliminary prototype to inform preliminary design.
- 2 contributions (4, 18) were judged to be at HRL 3: Mapping of human interactions and application of standards to proof of concept.
- 10 contributions (2, 8, 9, 20, 23, KN1, KN2, HFM, P1, P2) were judged to be at HRL 2: Human capabilities & limitations and system affordances & constraints applied to preliminary conceptual designs.
- 5 contributions (6, 11, 12, 17, 22) were judged to be at HRL 1: Human-focused concept of operations (human use scenario) defined.

The general level of Human Readiness of this body of work, judged by this method, was indicated to be relatively low in comparison with Technology Readiness. This reflects the notion that Human readiness tends to lag behind Technology readiness, unless prioritised. The HRL scale is strongly matched to the application of Human Systems Integration HSI processes for system procurement, and not designed for evaluating S&T R&D. Procurement requires a progressive, logical definition of system design requirements, for development, testing, verification and validation. HSI seeks to manage, control and assure human readiness in systems procurement. This demands more rigorous and thorough analysis and verification of human integration, usability and performance requirements than found most exploratory, investigative research and development projects. Research and development into HAT solutions for systems design, by virtue of its nature, needs to be conducted fully cognisant of the risks and benefits of managing both human and technology readiness requirements.

Human Readiness Top Ten – HRL 5: HSI demonstration and Early User Evaluation of initial and/or preliminary prototype to inform preliminary design.
1. Rohling – Command management, Exp 5.60
3. Keirl – Effectiveness metrics, Exp 5.60
5. Frost – Functional division, Exp 5.80
7. Baltzer – Collision avoidance, Exp 4.75
10. Murphy - Imagery analysis, Exp 5.83
13. Schmitt – Manned-Unmanned teaming, Exp 6.20
14. Diggelen – Social intelligence, Exp 5.60
15. Thomas – Policy management, Exp 6.33
21. Chen – Transparency, trust, Exp 4.33

Comparative analysis indicates that relatively high HRL work is strongly associated with high exploitability (15, 13, 10, 5, 14, 3).

Further analysis showed that HRL is also closely related to proof of performance effects. All 10 contributions reporting performance effects were judged relatively high HRL.

Performed Effects Reported
1. Rohling – Command management, HRL 5
3. Keirl – Effectiveness metrics, HRL 5
5. Frost – Functional division, HRL 5
7. Baltzer – Collision avoidance, HRL 5
10. Murphy – Imagery analysis, HRL 5
13. Schmitt – Manned-Unmanned teaming, HRL 5

15. Thomas – Policy management, HRL 5
21. Chen – Transparency, trust, HRL 5

4.4.4 Technical Capability

Technical Capability (TC) refers to building significant S&T capability to meet future needs. TC averaged TC 4.87 with a range of TC 2.50-6.40, and with 14 contributions judged greater than TC 5.00. On the whole, this provides good evidence of a strong technical capability in this diverse body of work.

Technical Capability Top Ten
5. Frost – Functional division
9. Kelley – Search planning
13. Schmitt – Manned-Unmanned teaming
11. Luchero – Information provenance
15. Thomas – Policy management
8. Neerincx – Patterns, work agreements
14. Diggelen – Social intelligence
3. Keirl – Effectiveness metrics
1. Rohling – Command management
10. Murphy - Imagery analysis

TC can be considered to be measured in terms of a variety of factors, residing in individuals, teams and organisations, including the research team management and technical resources, including team size, skills, competencies, expertise, knowledge and leadership, the equipment and facilities needed for development, test and evaluation, and prior and on-going relevant technical work, reputation and track record. The work on MUxV Functional Division performed at USAF AFRL (5. Frost, TC 6.40), providing a Human Effectiveness led demonstration of the IMPACT Multiple UxV Command and Control capability, is indicative of a singularly impressive TC. The core work focuses on Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies, and flexible delegation interfaces, with a strong emphasis on human factors optimised intuitive interface design for single operator MUxV C2. Autonomous capabilities include cooperative control algorithms for rapid vehicle route planning, intelligent agent reasoning that compares possible courses of actions and makes vehicle allocation recommendation, and autonomic monitoring technologies. In the wider context, IMPACT capability forms the core of a US DoD multi agency programme, led by Dr Mark Draper USAF, including USN SPAWAR, and Army Research Laboratory, under ARPI funding, and a wider international collaboration effort AIM (Allied IMPACT), integrating concepts and technology from international research partners including UK Dstl/QinetiQ (COMPACT Policy Management, Negotiation), Canada DRDC (Authority Pathway) and Australia DSTO (Narrative). A total of eight HFM SY 300 contributions were associated directly or indirectly with this significant collective IMPACT/AIM MUxV C2 endeavour:

IMPACT/AIM MUxV Components
3. Keirl. Effectiveness metrics
5. Frost. Functional division
9. Kelley, Search planning
11. Luchero, Information provenance
15. Thomas, Policy management
21. Chen – Transparency, trust
P1 Coronado/Lange – Autonomics framework
P2 Coronado/Lucero – Task management

For the present purposes, it seems necessary and sufficient to focus primarily on the nature of the core capability of interest – Human-Autonomy Teaming – from a cognitive capability perspective, seeking scientific evidence for understanding of the purposes, functions and processes of the critical components.

4.4.4.1 HTA Analysis

In the first instance, analysis was conducted of the content of the contributions for evidence of understanding of the functioning of the three basic system components, namely human engagement, autonomy operations, and teaming interdependencies. The contributions were examined for the presence and strength of content with reference to Human (H), Teaming (T) and Autonomy (A) component functioning.

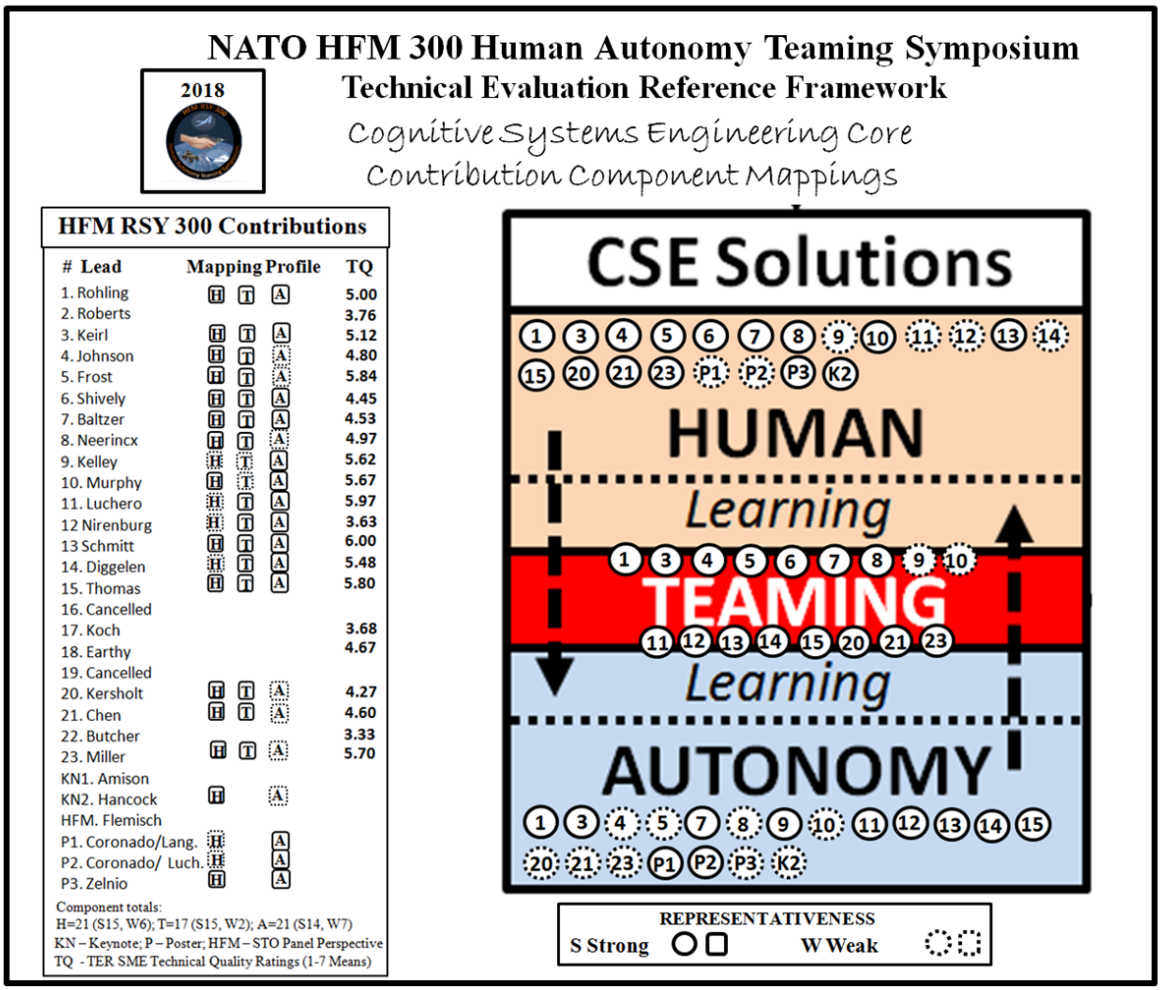


Figure 10: Human Autonomy Teaming Analysis

The results of the HTA analysis were mixed, and generally inconclusive, with limited diagnostic or prognostic power. They indicated that 17 of the 27 contributions provided consideration distinguishing the requirements of all three HTA elements, with a further four considering only human (H) and autonomy (A) functioning without explicit regard for teaming requirements (T). This equated to 21 Human (15 Strong, 6 Weak), 17 Teaming (15 Strong, 2 Weak), and 21 Autonomy (14 Strong, 7 Weak). Only six contributions (2, 17, 18, KN1, HFM) provided no explicit differentiation between these component requirements. Thus, whilst there was some evidence of a slightly reduced coverage of teaming requirements – indicative perhaps of a potential

weakness or gap in understanding - there was no clear evidence of bias towards either human or autonomy related content, both being covered well, in approximately equal measure. The Programme intent was to seek a balanced consideration of human, autonomy and teaming requirements. This evidence suggests that a good balance probably was achieved, albeit with a weakness in evidence of teaming requirements.

4.4.4.2 REMEDE Decisions Model Comparison

A second deeper analysis was conducted seeking further discrimination between the contributions content on the core problem solving and decision making capability. With the challenges of integrating AI in mind, it is considered that understanding models and methods of problem-solving and decision-making, analysing the operations of joint cognitive systems, and examining the application of cognitive systems engineering methods, define the initial scope of enquiry of primary interest.

Dr Chris Miller, in his excellent HAT Symposium Contribution on Displaced Transparency (23. Miller, TQ 5.70), provided an informative account of the theoretical concept of models comparison, specifically "mental models", including models reconciliation and synchronisation, characteristic of mentor critiquing systems [29, 30]. Good teaming is often considered to be "a meeting of minds". In practice, models comparison can be considered as involving consideration of mental models for specific processes and functions for problem solving and decision making. This can be achieved by analysis of information processing stages, discretized as a series of logical steps, typically representing information acquisition, analysis, decision making, and course-of-action plan execution. Additionally, the Teaming requirement necessitates consideration of interaction, communication and information exchanges for effective teaming models reconciliation and synchronisation, and for coordination, cooperation and collaborative teaming performance.

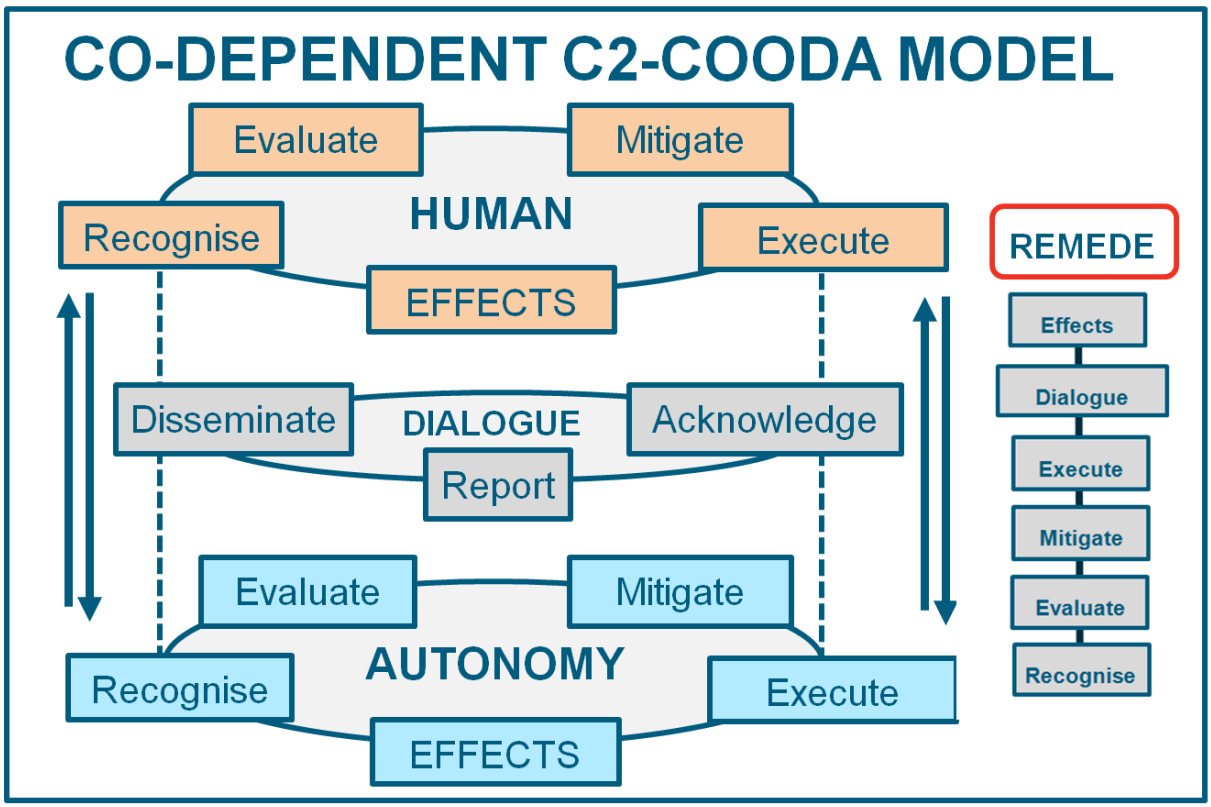


Figure 11: REMEDE Decision Comparison Model

Knowledge-based systems engineering uses a models framework, CommonKADS, to represent the knowledge content and structure needed to support decision making in knowledge based tasks [31]. CommonKADS comprises linked models representing organisation, task, agent, knowledge, communication and design components.

For the purpose of conducting a decision models comparison analysis, with performative effects as the core control component, the Dstl REMEDE framework was used, based on a co-dependent COODA Model for joint cognitive control systems and control task analysis [32, 33]. The REMEDE framework seeks to differentiate the representativeness, scope and content of the individual HFM SY 300 HAT contributions, based on a simple human cognition model, including teaming communication dialogue and resultant effects.

The REMEDE model distinguishes six core, logical, cognitive components of problem solving and decision making based on extant theory of joint cognitive systems, and as used previously for Cognitive Control Task Analysis:

- RECOGNITION: Awareness of changed situation. Observation of information and data. Identification of current actual mission and system state. Tasks, Targets, Threats, Tactics, Timings.
- EVALUATION: Interpretation & analysis of consequences, options, choices, performance implications & anticipated effects. Mission goals, opportunities, affordances, constraints, options, effects.
- MITIGATION: Identification of change to aims & objectives. Definition of new task and operating conditions. Formulation of task plan, procedures and sequence of actions. Revised mission goals, task re-definition, plan adjustment, and changed tactics, techniques and procedures (TTPs).
- EXECUTION: Enactment of adjusted re-planned tasks, procedures, manipulations and manoeuvres. Coordination, cooperation and collaboration. Tasks, TTPs, Coordination, Cooperation, Collaboration
- DIALOGUE: Primary message communication and feedback. Dissemination, Acknowledgement, Report.
- EFFECTS: Impact of actual effects on achievement of Mission and Task Objectives. Survivability, Effectiveness, Timeliness, Agility, Adaptability, Offensive and Defensive Performance, Probability of Mission Success

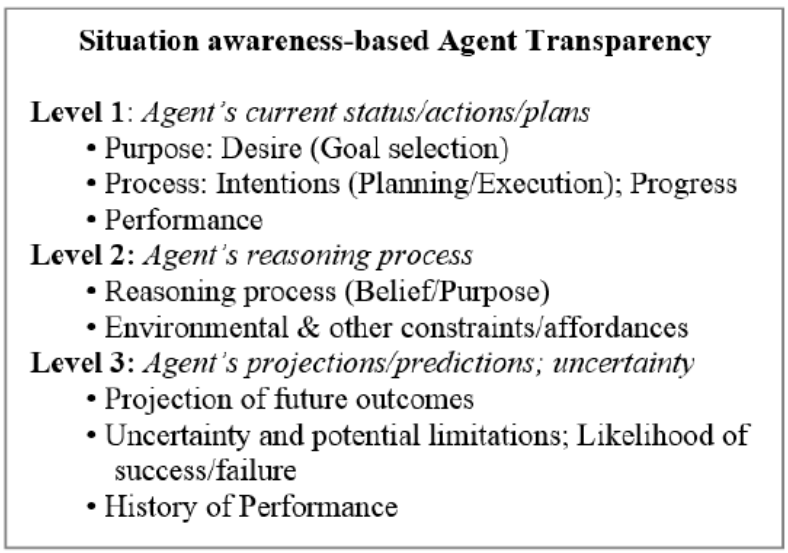


Figure 12: SAT – Situation Awareness-based Agent Transparency

Recognition, Evaluation and Mitigation components can be considered as equivalent to the three SAT situation awareness transparency levels reported in the Symposium proceedings (11, 21), based on the popular Perception, Comprehension, and Projection SA trilogy [34]. REMEDE adds Execution, completing the cognitive OODA Loop (Observe, Orient, Decide, Act). REMEDE includes Dialogue to capture the collaborative nature of Teamwork. Dr Miller points out in his contribution on Displaced Transparency (23) that in high performing, efficient teams, with well synchronised models, explicit communication dialogue is required less, and that dialogue focuses on models mismatch [35]. Finally, REMEDE captures the Effects of the processes functioning, for completeness, reflecting the functional purpose as a decision making model for dynamic cognitive control, or DOODA Loop, requiring feedback on performance [36]. Additionally, the model comprises essential, generic functional components of a cognitive framework for Artificial General Intelligence, Social Intelligence and mentor Critiquing Systems. It should be noted that under the Dstl work on Collaborative Adaptability Proficiency Test and Evaluation Assessment Methodology (CAPTEAM) adaptability proficiency performance effectiveness metrics (3), Human-Human Teaming collaboration is represented using teaming attributes from Crew Resources Management techniques, namely Shared SA, Leadership, Communication, and Support [37, 38].

The content of the contributors' documentation was analysed by the TER author using a Dstl REMEDE protocol. The protocol used is illustrated. It provided estimated ratings indicative of the strength of contribution of the 6 REMEDE facets represented in the work in relation to team joint decision making processes and performance. Ratings were obtained for the presence and influence of the 6 REMEDE facets, discretizing evidence of contributions from human and/or autonomy models, and including ratings of teaming models comparison dialogue and performative effects. Additional ratings were capture in relation to Trustworthiness and Capability Maturity.

NATO HFM 300 HUMAN AUTONOMY TEAMING – TECHNICAL EVALUATION REPORT							
Paper/Poster Number: Authors: Title:							
HAT Relevance **Contribution to Human-Autonomy Team Joint Decision Making Processes & Performance**	**Low**	**ASSESSMENT RATING**					**High**
	1	2	3	4	5	6	7
RECOGNITION Awareness of changed situation. Observation of information & data. Identification of current actual mission & system state. Tasks, Targets, Threats, Tactics, Timings							
Comments:							
EVALUATION Interpretation & analysis of consequences, options, choices, performance implications & anticipated effects. Mission goals, opportunities, affordances, constraints, options, effects.							
Comments:							
MITIGATION Identification of change to aims & objectives. Definition of new task & operating conditions. Formulation of task plan procedures & sequence of actions. Revised mission goals, task re-definition, plan adjustment, changed TTPs.							
Comments:							
EXECUTION Enactment of adjusted re-planned tasks, procedures, manipulations and manoeuvres. Coordination, cooperation & collaboration. Tasks, TTPs, Coordination, Cooperation, Collaboration							
Comments:							
DIALOGUE Primary message communication and feedback. Dissemination, Acknowledgement, Report.							
Comments:							
EFFECTS Impact of actual effects on achievement of Mission & Task Objectives. Survivability, Effectiveness, Timeliness, Agility, Adaptability, Offensive & Defensive Performance, Probability of Mission Success							
Comments:							
TRUSTWORTHINESS Dependability, predictability, reliability, resilience, safety, security, availability							
Comments:							
HAT Capability Maturity Model - Level of Maturity	0 Not Performed	1 Initial	2 Recognised	3 Defined	4 Managed	5 Optimised	6 Adaptive
Summary:							

Figure 13: Assessment Protocol for Decision Models Analysis

Differentiating the contributions was neither a simple nor straightforward task. It was heavily dependent on the information available in the technical descriptions of the autonomy processes and functioning, which varied greatly. It often required assumptions, logical reasoning and a degree of judgement to be exercised, based on understanding of the technology and of the nature of the problem-solving tasks involved in the reported work. Without further enquiry and validation, the findings should be regarded as tentative, and at best indicative of the probable shared balance of effort and capability. Without validation, they should be recognised as potentially inaccurate, erroneous and not necessarily reliable, consistent, or strongly representative. The ratings obtained in this manner were then classified broadly as either relatively strong or weak. The results were then set out in relation to the REMEDE model of the core cognitive components.

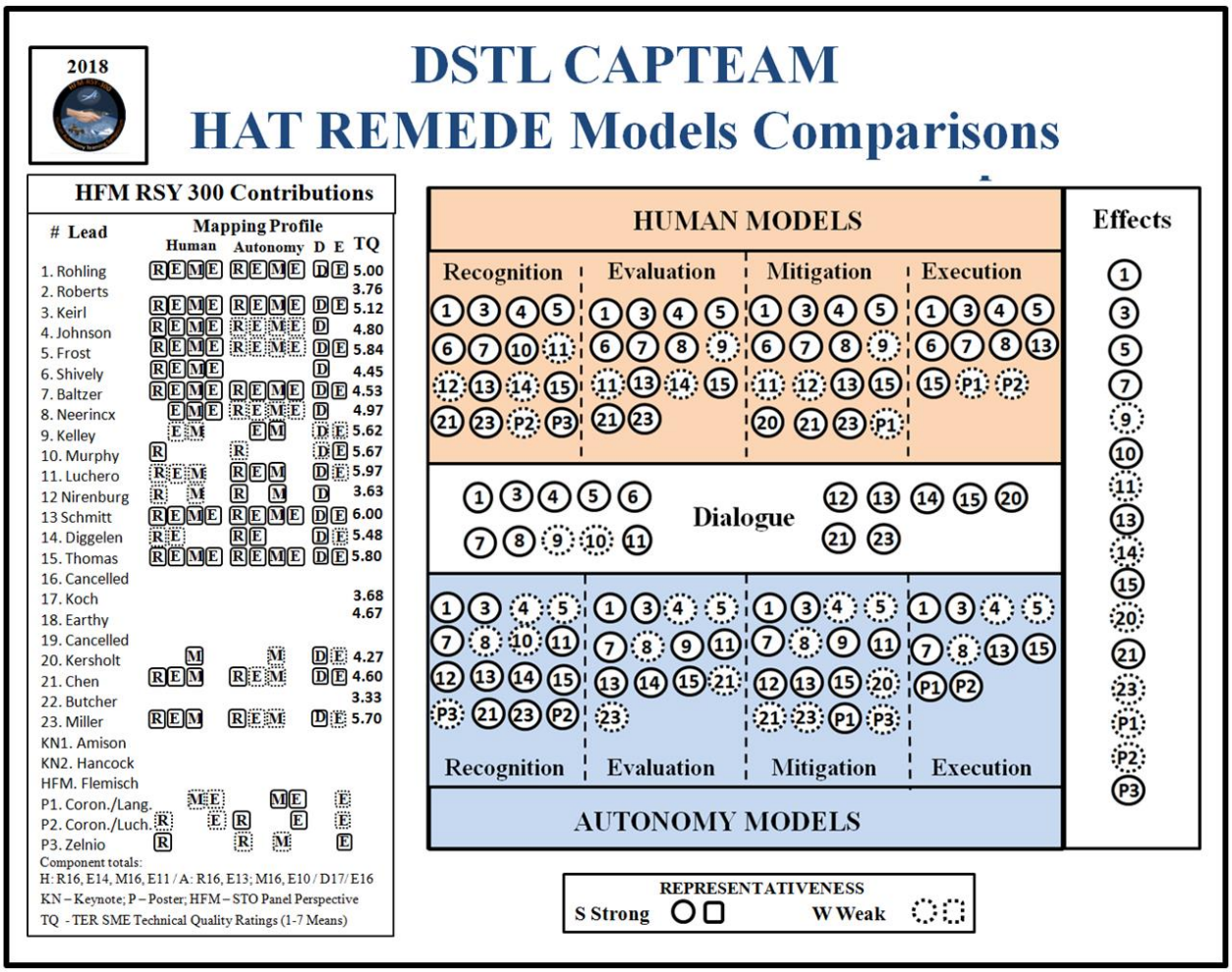


Figure 14: Results of Decision Models Analysis

This analysis shows evidence of model comparisons representativeness in the majority of the 20 qualifying contributions with analysable core cognitive component engineering content. The component mappings totalled 145 out of a possible 200, distributed as follows:

- Human Models - Recognition 16; Evaluation 14; Mitigation 16; Execution 11. Total 57 (44 Strong, 13 Weak).
- Autonomy Models - Recognition 16; Evaluation 13; Mitigation 16; Execution 10. Total 55 (35 Strong, 20 Weak).

- Dialogue 17 (15 Strong, 2 Weak)
- Effects 16. (9 Strong, 7 Weak)
- Total: 145

Models Comparison Top Ten
1. Rohling – Command management **REME REME DE** (TC 5.40)
3. Keirl – Effectiveness metrics **REME REME DE** (TC 5.40
7. Baltzer – Collision avoidance **REME REME DE** (TC 4.08)
13. Schmitt – Manned-Unmanned teaming **REME REME DE** (TC 6.20)
15. Thomas – Policy management **REME REME DE** (TC 6.00)
5. Frost – Functional division **REME** REME **DE** (TC 6.40)
4. Johnson – Interdependency analysis **REME** REME **D_** (TC 5.00)
21. Chen – Transparency trust **REM_ R**EM_ **DE** (TC 5.00)
11. Luchero – Information provenance REM_ **REM _ D**E (TC 6.00)
8. Neerincx – Patterns, work agreements **_EME** REME **D_** (TC 5.50)

Comparative analysis indicates that relatively strong Models Comparison profile work is associated with high Technical Quality. The results showed that 7 of the Top Ten Technical Capability (TC) contributions had strong Models Comparison profiles (5, 13, 11, 15, 8, 3, 1).

The results showed evidence of balance of human and autonomy component competencies, with human activity contributions more strongly represented, which is understandable naturally. Human cognition was a given, whereas autonomy cognition needed creating by engineering. There was some evidence of relatively stronger representativeness of Recognition (32) and Mitigation (32) components over Evaluation (27) and Execution (20). Dialogue (17) and Effects (16) were relatively well represented. Dialogue was a strong feature of most contributions. Effects were much more varied in strength, as reported earlier.

The REMEDE component Evaluation, equivalent to SAT Level 2 Reasoning Process, is a vitally important contributor to adaptive mitigation. This raises concern that the Evaluation/Reasoning component seems relatively weak in this small sample of Autonomy Model Components (13/20, 8 Strong, 5 Weak). Dr Miller reports in his contribution on Displaced Transparency (23), that there is evidence that the inclusion of mental modelling capabilities in the reasoning of a robot agent, where the robot was modelling the expected reasoning of a human operator and reacting accordingly, produced a 44-75% improvement in robot decision making in terms of avoiding resulting resource conflicts [29].

4.4.4.3 Capability Maturity Level Analysis

The Technical Capability assessments of the HFM SY 300 contributions were considered in relation to evidence of capability maturity. This was achieved using the recently developed Dstl HAT Capability Maturity Model and associated metrics [37, 38]. The Dstl HAT Capability Maturity Model (HAT CMM) approach is based on 1980's software CMM work at Carnegie Mellon University [39], modified with later more people-based dimensions (Systems Engineering, Usability, Organisation, People, Team Process). Capability Maturity Levels (CML) are calibrated in a manner similar to TRLs. Seven levels for HAT Capability Maturity are identified and defined, differentiated in terms of goals, processes, and behaviours. They progress from CML 0-Not Performed to CML 6-Adaptive, progressing through Level 1-Initial, Level 2-Recognised, Level 3-Defined, Level 4-Managed, and Level 5-Optimising. The Symposium contributions were nominally categorised as follows:

- CML 0 – Not Performed: 2, 17, 22, K1, K2, HFM
- CML 3 – Defined: 3, 5, 12, 18, 20, P1, P2

- CML 4 – Managed: 1, 4, 7, 8, 9, 10, 14, 15, 21, P3
- CML 5 – Optimised: 11
- CML 6 – Adaptive: 6, 13, 23

Using this form of classification, none were assessed as CML 1 or 2, Initial or Recognised. The majority of the contributions (17/27) were classified as being at relatively high, CML 3 Defined (7) and CML 4 Managed (10), adaptable rather than adaptive, HAT Capability Maturity Levels. Four of the contributions were judged higher, with one at CML 5 Optimised (11), and three at the highest level CML 6 Adaptive (6, 13, 23).

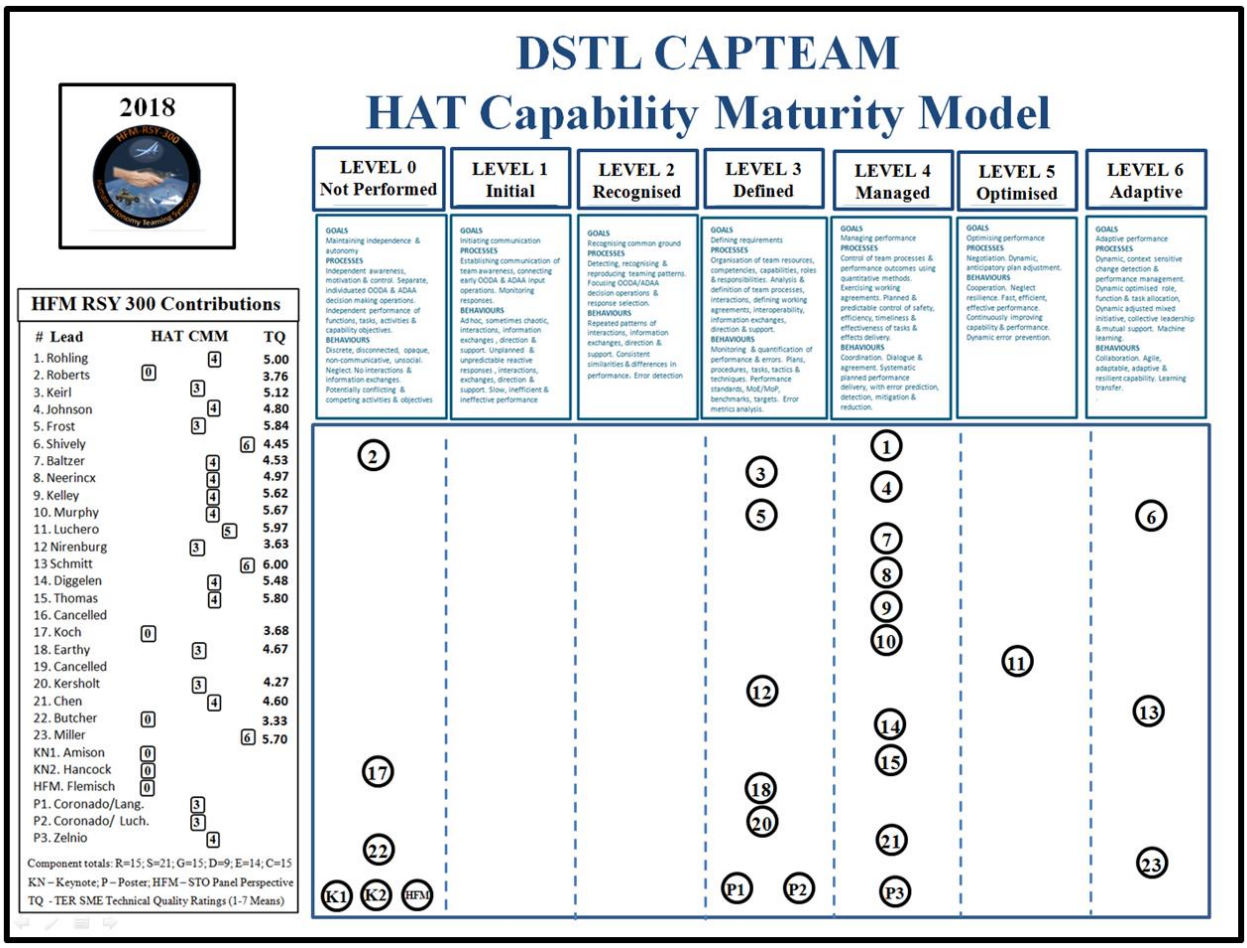


Figure 15: Results of Capability Maturity Level Analysis

High Capability Maturity Levels
CML6 Adaptive
6. Shively – Communication dialogue (TC 5.50)
13. Schmitt – Manned-Unmanned teaming (TC 6.20)
23. Miller – Displaced transparency (TC 5.00)
CML5 Optimised
11. Luchero – Information provenance (TC 6.00)
CML4 Managed
1. Rohling – Command management (TC 5.40)
4. Johnson – Interdependency analysis (TC 5.00)

7. Baltzer – Collision avoidance (TC 4.08)
8. Neerincx – Patterns, work agreements (TC 5.50)
9. Kelley – Search planning (TC 6.33)
10. Murphy - Imagery analysis (TC 5.33)
14. Diggelen – Social intelligence (TC 5.40)
15. Thomas – Policy management (TC 6.33)
21. Chen – Transparency trust (TC 5.00)
P3. Zelnio - Imagery analysis

Comparative analysis indicates that relatively high Capability Maturity is associated with work of good Technical Quality, with some exceptions. The results showed that nine of the Top Ten Technical Capability (TC) contributions obtained high CML4/5/6 Capability Maturity estimates (9, 13, 11, 15, 8, 14, 3, 1, 10). Most notably, the work on Manned-Unmanned Teaming - 13. Schmitt, CML 6, TC 6.20 - was classified as CML 6 Adaptive, with high TC 6.20. This was due to the integration of crew adaptive workload management techniques in teaming with an artificial cognitive agent associate system, including mixed initiative planning. The contributions on Communication Dialogue (6. Shively, CML 6, TC 5.00) and Displaced Transparency (23. Miller, CML 6, TC 5.00) were judged High CML 6, despite modest TC, and relatively low TRL and HRL attributions. Both provided strong and convincing logical arguments for adaptive concept capabilities, but offered no objective proof of concept through HAT system development, test and evaluation. The work on Functional Division (5. Frost, CML 3, TC 6.40) received a high TC 6.40, but modest CML 3 Defined appraisal. This work used a human confederate emulation for autonomous team-mate functioning.

The work on Measuring Effectiveness defines test criteria, methods and metrics for high levels of HAT Capability Maturity, including CML 6 Adaptive Proficiency (3. Keirl, CML 3; TC 5.40). However, the capability maturity of the HAT systems under test (C2 UAS, MUxV, MUM-T, Collaborative autonomy, Adaptable autonomy, Policy management, Negotiation) was judged CML 4 Managed (15. Thomas, CML 4, TC 6.33). Development of HAT Test & Evaluation methods and metrics for verification, validation, qualification and certification of autonomous systems employing non-deterministic algorithms was identified as a particularly significant challenge in the DoD DSB Task Force Report 2012 on The Role of Autonomy in DoD Systems [4]. In UK work under the Dstl/MoD Autonomy Research Programme [37], relatively mature methods and metrics for Adaptivity Proficiency have been developed, tested and applied to measuring Human Taskwork and Human Teamwork with autonomous systems e.g. Dstl CAPTEAM (CML 6). Methods and metrics for measuring Autonomy Taskwork (e.g. Dstl Trustworthiness) and Human-Autonomy Teamwork (e.g. Dstl HAT CMM) are defined and under development by Dstl for application testing V&V (CML 3). Identification and development of reliable and effective methods and metrics for measuring Autonomy-Autonomy Teamwork, where observability is a particular challenge, are currently under investigation.

The contributions on Communication Dialogue and Displaced Transparency are distinctive because they provide accounts of means by how HAT systems can achieve CML 6 Adaptive capability maturity. Dialogue is a critical path enabler of teaming effectiveness. Dialogue provides feed-forward and feedback needed for control of effects. For teaming, dialogue enables teaming models reconciliation and synchronisation. Historically, effective HAT dialogue is best exemplified by use of the short-cut language of “plays”, rapidly communicating intent and plans for coordinated tasks from set of tactics, manoeuvres and procedures using a Playbook GUI, based on American football practice [40].

Arguably, of all the contributions, only the Manned-Unmanned Teaming work provides a successful, proven implementation of CML 6 HAT Adaptive capability, albeit using a crew workload monitoring adaptive system. The MUM-T work provides an indication of the complex technical requirements for providing a proven Intelligent Adaptive HAT System, capable of adjusting and optimising Teaming with context sensitivity, content, structure and performance delivery matched to the wider system context, including mission and external situational variables, such as task, targets, threats and tactics.

4.4.5 Innovation

Innovation refers to pushing boundaries and production of something new and novel. HAT is a relatively recent re-formulation of the Human-Automation relationship dilemma, with Artificial Intelligence and machine agency added. In that sense, delivering effective HAT can be thought of as an inherently new and novel challenge. However, Innovation (In) was assessed as the relatively weaker attribute of the assessed five dimensions, with ratings averaging lowest at In 4.59, yet still above the scale mid-point. Innovation average ratings for individual contributions ranged from In 2.20-6.50, with 9 of 21 contributions rated greater than 5.00.

Innovation Top Ten
23. Miller – Displaced transparency
11. Luchero – Information provenance
9. Kelley – Search planning
14. Diggelen – Social intelligence
5. Frost – Functional division
13. Schmitt – Manned-Unmanned teaming
10. Murphy - Imagery analysis
20. Kersholt – Social relationships
15. Thomas – Policy management
12. Nirenburg – Natural language

The idea of Displaced Transparency, along with models synchronisation and reconciliation, is undoubtedly new, and profoundly simple, as is most good thinking (23. Miller, In 6.50). These ideas draw upon the good experimental work performed by USARL on trust and transparency [34], making AI systems more explainable, for purposes of GCS interface design, as applied to IMPACT MUxV C2 (21. Chen, TC 5.00). Displaced transparency presents the idea that teaming relationships, including trust, need not be dependent on cognitively demanding real-time interaction and transparency. Effective teaming relationships can evolve and build before, during and after action, through proactive, active, and reactive learning and interdependencies.

Understanding trust in the context of Human-Autonomy Teaming has attracted attention of researchers since the original workshops on Human-Electronic Crew Teamwork, 1988-1997 with trust and teaming as a linking theme. The problem with research on automation trust was once described insightfully to the author by Raja Parasuraman RIP, after publication of his seminal work in 1997 on automation use and abuse [41]. In summary, Raja said: *"Trust is an intervening cognitive variable between reliability and use".* Conceived of as such, trust is a complex, dynamic, unpredictable, subjective attitudinal variable. Like "*confidence*", trust is influenced by complex interactions with cognitive, affective and behavioural factors. Early human factors research on Air Traffic Control showed confidence to be poorly correlated with operator performance. Human-Machine Teaming systems engineering requirements are derived from understanding of factors directly affecting reliability and use, such as information veracity, provenance, reliance and trustworthiness (3. Keirl, Effectiveness metrics, TQ 5.12).

The scope for learning transfer between human-machine teaming programmes, and trust and influence efforts, needs further definition and improved understanding. Since 2012 USAF AFSOR have hosted a Trust and Influence research programme on kinetic and non-kinetic effects, leveraging work on Human-Machine Teaming and on trust in autonomous systems [42]. The programme covers cognitive and social factors in human reliance, socio-digital media influence and application of computational methods in social science. The work seeks better understanding of variables that influence social and cultural behaviour, including attitudes and beliefs, through ideas such as properly calibrated trust, and on how people maintain and repair trust in agents, with interest in developing novel man-machine interfaces and interaction techniques.

The USN SPAWAR work on Information Provenance, in collaboration with Kings College London, concerning algorithmic transparency and accountability, is deeply technical, and potentially ground breaking, unveiling the veracity AI/ML data analytics and decisions (11. Luchero, In 6.33). Other USN SPAWAR work on Search Planning seeks to couple human knowledge with AI planning algorithms to reduce search space and assist the machine agents' through constraints and goals dynamically during plan generation and execution (9. Kelley, In 5.67). This provides a reminder that the core AI technology is data analytics, data fusion, and task information planner and manager systems. The TNO work on adding Social Intelligence to collaborative AI systems, as a modular, pluggable layer is simple, original and attractive (14. Diggelen, In 5.60). It is backed up by a sound comprehensive appreciation of literature (20. Kersholt, In 5.00), and situated with a strong MUxV C2 programme *e-partners* approach TC focusing on ontologies, design patterns and working agreements (8. Neerincx, TC 5.50). Working agreements and agreement technologies are fundamental core components for cognitive control with HAT, such as UK DERA PACT Contractual Autonomy [43, 44]. Agreement technologies remain central also to the UK Dstl/QinetiQ work on COMPACT Policy Management and Negotiation, directed at delivery of assured MUM-T and MUxV C2 (15. Thomas, In 5.00).

In general, the evidence for Innovation in the contributions - pushing boundaries and producing something new and novel - can be considered to be more mixed. This may reflect the strong influence of applied research approaches in the set of provided contributions. Bridging science and operations often means applying proven, practical and pragmatic methods and techniques, compared with basic, foundational, investigative research, where opportunities for scientific innovation are more likely. Most contributions, seeking to prove concepts and develop solutions with operational relevance, have sought to use rigorous, proven and effective methods. The concepts and methods of cognitive systems engineering are not new.

As indicated in the earlier Exploitation analysis, the research contributions, including posters papers, but excluding Keynotes, exhibited a wide range of TRLs. Specifically, 13 were assessed as at TRL 3 or below (i.e. relatively Conceptual /Analytical), and 11 at TRL 4 or above (Validation /Test/ Demonstration). Four of the 13 low TRL contributions were assessed as relatively strong on Innovation, namely: Displaced transparency (23. Miller, In 6.50); Information provenance (11. Lucerno, In 6.33); Imagery analysis (10. Murphy, In 5.17); Natural language (12. Nirenburg, In 4.83). Innovation was considered also to be evidenced in some of the more constrained higher TRL work, most notably: Search planning (9. Kelley, In 5.67); Social intelligence (14. Diggelen, In 5.60); Functional division (5. Frost, In 5.20); Manned-unmanned teaming (13. Schmitt, In 5.20). It should be noted that Dr Chris Miller has been a distinguished thought leader in the HAT domain since the 1980s Human-Electronic Crew era, and it is most pleasing that his work continues to be seen to be highly valued.

4.4.6 Technical Quality

TQ refers to the aggregated ratings averages of all five assessment technical components, coupling Operational Relevance, Scientific Quality, Technical Capability, Exploitability and Innovation. TQ averaged TQ 4.90 with a range of TQ 3.33-6.00, and with 10 contributions judged equal or greater than TQ 5.00. 17/21 contributions were above the scale mid-point of TQ 4.00. This provides evidence of good technical quality across this highly diverse body of work.

Technical Quality Top Ten
13. Schmitt – Manned-Unmanned teaming
11. Luchero – Information provenance
5. Frost – Functional division
15. Thomas – Policy management
23. Miller – Displaced transparency
10. Murphy - Imagery analysis
9. Kelley – Search planning

14. Diggelen – Social intelligence
3. Keirl – Effectiveness metrics
1. Rohling – Command management

Manned-Unmanned Teaming (13. Schmitt, TQ 6.00) and Information Provence (11. Luchero, TQ 5.97) vie closely for the highest rated assessed contributions. Functional Division (5. Frost, TQ 5.84) and Policy Management (15. Thomas, TQ 5.80) follow closely. Displaced Transparency (23. Miller, TQ 5.70) completes the Top Five, marginally ahead of Imagery Analysis (10. Murphy, TQ 5.67). Manned-Unmanned Teaming is also covered in the UK work on Effectiveness Metrics (3. Keirl, TQ 5.12) and Policy Management (15. Thomas, TQ 5.80). MUM-T is a particularly powerful application and test-bed for developing HAT principles and techniques. It provides a context of use and testing with high risks and performance stress, particularly operating in the airborne environment, involving mission critical Measures of Effectiveness, associated with Survivability, Effectiveness and Timeliness.

Table 7: List of Contributions in Order of Technical Quality

TQ Rank Order	PAPER	Technical Quality TQ	SESSION	TG247
1	13 Schmitt - Manned Unmanned Teaming	6.00	5. HAT Interaction and Design	TG247
2	11 Luchero – Information provenance	5.97	4. Autonomy Capabilities for Teaming	TG247
3	5 Frost – Functional division	5.84	2. Human Autonomy Teaming Structure	TG247
4	15 Thomas – Policy management	5.80	5. HAT Interaction and Design	TG247
5	23 Miller – Displaced transparency	5.70	6. HAT Interaction and Design	TG247
6	10 Murphy – Imagery analysis	5.67	4. Autonomy Capabilities for Teaming	
7	9 Kelley – Search planning	5.62	2. Human Autonomy Teaming Structure	
8	14 Diggelen – Social intelligence	5.48	5. HAT Interaction and Design	
9	3 Keirl – Effectiveness metrics	5.12	1. Operational Requirements	TG247
10	1 Rohling – Command management	5.00	1. Operational Requirements	TG247
11	8 Neerincx – Patterns, work agreements	4.97	2. Human Autonomy Teaming Structure	TG247
12	4 Johnson – Interdependency analysis	4.80	1. Operational Requirements	
13	18 Earthy – ergonomics standards	4.67	Panel on HAT Institutional Integration	
14	21 Chen – Transparency, trust	4.60	6. HAT Interaction and Design	TG247
15	7 Baltzer – Collision avoidance	4.53	3. Human Autonomy Teaming Structure	TG247
16	6 Shively – Communication dialogue	4.45	2. Human Autonomy Teaming Structure	TG247
17	20 Kersholt – Social relationships	4.27	Panel on HAT Institutional Integration	
18	2 Roberts – HFI Organisation risks	3.76	1. Operational Requirements	
19	17 Koch – Ethical, Legal arguments	3.68	Panel on HAT Institutional Integration	
20	12 Nirenburg – Natural language	3.63	4. Autonomy Capabilities for Teaming	
21	22 Butcher – Trust attitudes	3.33	6. HAT Interaction and Design	

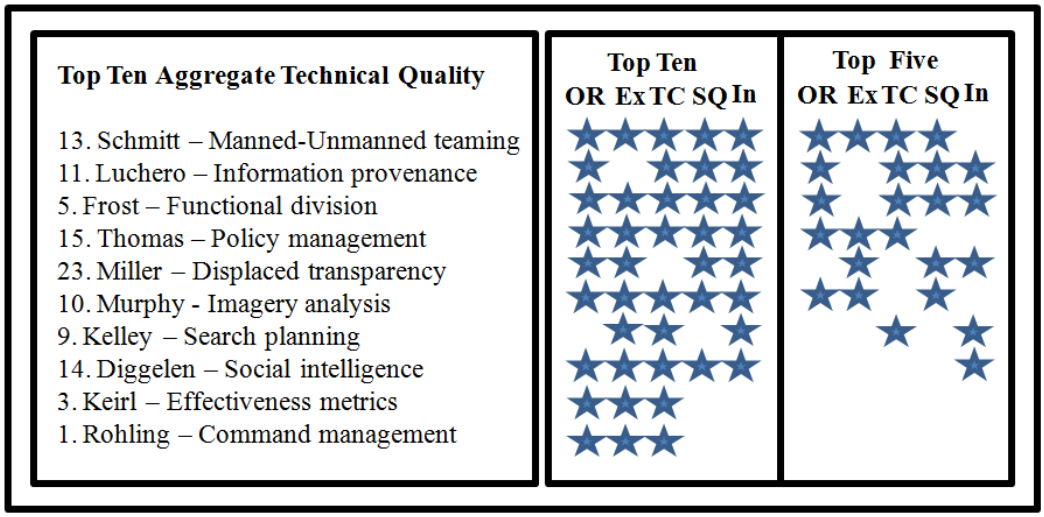


Figure 16: Top Ten TQ Obtaining Top Ten and Top Five Technical Component Ratings

Comparisons of the ratings for the five technical components with the aggregated Technical Quality show consistency in ratings across multiple technical components, in particular for Operational Relevance, Exploitability and Technical Capability. The Top Ten Aggregate TQ contributions featured frequently in the Top Ten for OR (9), Ex (9), TC (9), SQ (7) and In (8). The following contributions featured 5 times in the Component Set Top Ten:

- Manned-Unmanned Teaming (13.Schmitt, TQ 6.00)
- Functional Division (5.Frost, TQ 5.84)
- Policy Management (15. Thomas, TQ 5.80)
- Imagery Analysis (10. Murphy, TQ 5.67)
- Social Intelligence (14. Diggelen, TQ 5.48).

More stringently, the high TQ contributions featured less frequently in the Component Set Top Five with OR (5), Ex (4), TC (5), SQ (5), In (5), but with the following more sensitive tailored differentiation:

- Manned-Unmanned Teaming x4 (13.Schmitt, TQ 6.00),
- Information Provenance x4 (11. Luchero, TQ 5.97)
- Functional Division x4 (5.Frost, TQ 5.84),
- Policy Management x3 (15. Thomas, TQ 5.80),
- Imagery Analysis x3 (10. Murphy, TQ 5.67)
- Search Planning x3 (9. Kelley, TQ 5.62)
- Social Intelligence x1 (14. Diggelen, TQ 5.48).

The value of these comparisons is in indicating internal consistency based on evidence of relative status of the contributions, rather than a dependence on potentially unreliable precision in absolute values.

Table 8: Top Ten for Five Technical Components and Aggregated Technical Quality

Operational Relevance	Exploitability	Technical Capability
10. Murphy - Imagery analysis	18. Earthy – Ergonomics standards	5. Frost – Functional division
15. Thomas – Policy management	15. Thomas – Policy management	9. Kelley – Search planning
13. Schmitt – Manned-Unmanned teaming	13. Schmitt – Manned-Unmanned teaming	13. Schmitt – Manned-Unmanned teaming
5. Frost – Functional division	23. Miller – Displaced transparency	11. Luchero – Information provenance
11. Luchero – Information provenance	10. Murphy - Imagery analysis	15. Thomas – Policy management
3. Keirl – Effectiveness metrics	5. Frost – Functional division	8. Neerincx – Patterns, work agreements
1. Rohling – Command management	14. Diggelen – Social intelligence	14. Diggelen – Social intelligence
17. Koch – Ethical, legal arguments	3. Keirl – Effectiveness metrics	3. Keirl – Effectiveness metrics
23. Miller – Displaced transparency	1. Rohling – Command management	1. Rohling – Command management
14. Diggelen – Social intelligence	9. Kelley – Search planning	10. Murphy - Imagery analysis

Scientific Quality	Innovation	Aggregate Technical Quality
11. Luchero – Information provenance	23. Miller – Displaced transparency	13. Schmitt – Manned-Unmanned teaming
13. Schmitt – Manned-Unmanned teaming	11. Luchero – Information provenance	11. Luchero – Information provenance
5. Frost – Functional division	9. Kelley – Search planning	5. Frost – Functional division
10. Murphy - Imagery analysis	14. Diggelen – Social intelligence	15. Thomas – Policy management
23. Miller – Displaced transparency	5. Frost – Functional division	23. Miller – Displaced transparency
14. Diggelen – Social intelligence	13. Schmitt – Manned-Unmanned teaming	10. Murphy - Imagery analysis
15. Thomas – Policy management	10. Murphy - Imagery analysis	9. Kelley – Search planning
7. Baltzer – Collision avoidance	20. Kersholt – Social relationships	14. Diggelen – Social intelligence
8. Neerincx – Patterns, work agreements	15. Thomas – Policy management	3. Keirl – Effectiveness metrics
21. Chen – Transparency, trust	12. Nirenburg – Natural language	1. Rohling – Command management

Table 9: Results of Technical Quality Assessment and Analysis

PAPER	TQ	OR	Ex	TC	SQ	In	RSGDEC	HTA	REME/REME/DE	Effect	TRL	HRL	CML
13 Schmitt	6.00	6.20	6.20	6.20	6.20	5.20	RSGDEC	HTA	REME/REME/DE	Perf'd	6	5	6
11 Luchero	5.97	6.00	5.17	6.00	6.33	6.33	RSG_EC	(H)TA	(REM)_/REM_/D(E)	Reas'd	2	1	5
5 Frost	5.84	6.00	5.80	6.40	5.80	5.20	RSGDEC	HT(A)	REME/(REME)/DE	Perf'd	6	5	3
15 Thomas	5.80	6.33	6.33	6.00	5.33	5.00	RSGDEC	HTA	REME/REME/DE	Perf'd	7	5	4
23 Miller	5.70	5.50	6.00	5.00	5.50	6.50	_S _ _EC	HT(A)	REM _ /R(EM)_ /D(E)	Reas'd	2	2	6
10 Murphy	5.67	6.33	5.83	5.33	5.67	5.17	RS_DEC	H(T)A	R_ _ _ /(R) _ _ _ /(D)E	Perf'd	3	5	4
9 Kelley	5.62	5.33	5.33	6.33	5.00	5.67	_ S _ _EC	(HT)A	_ (EM) _ /_ EM _ /(DE)	Reas'd	4	2	4
14 Diggelen	5.48	5.40	5.60	5.40	5.40	5.60	_SG_ EC	(H)TA	(RE) _ _ /RE _ _ /D(E)	Reas'd	6	5	4
3 Keirl	5.12	5.80	5.60	5.40	4.40	4.40	RSGDEC	HTA	REME/REME/DE	Perf'd	7	5	3
1 Rohling	5.00	5.80	5.60	5.40	4.00	4.00	RSGDEC	HTA	REME/REME/DE	Perf'd	7	5	4
8 Neerincx	4.97	4.67	5.00	5.50	5.17	4.50	_SGDEC	HT(A)	_ EME/(REME)/D_	Recog'	2	2	4
4 Johnson	4.80	4.80	5.00	5.00	4.80	4.40	_S _ _E_	HT(A)	REME/(REME)/D_	Recog'	3	3	4
18 Earthy	4.67	4.67	6.33	3.33	4.67	4.33	_ _GD_ _		_ _ _ _/_ _ _ _/_ _	Reas'd	2	3	3
21 Chen	4.60	5.33	4.33	5.00	5.00	3.33	_SG_EC	HT(A)	REM _ /R(EM)_ /DE	Perf'd	4	5	4
7 Baltzer	4.53	4.50	4.75	4.08	5.25	4.08	_S_ DE_	HTA	REME/REME/DE	Perf'd	6	5	4
6 Shively	4.45	4.75	4.25	5.00	4.25	4.00	_S _ _ E _	HTA	REME/ _ _ _ _/D_	Recog'	2	1	6
20 Kersholt	4.27	4.00	4.00	3.67	4.67	5.00	_S _ _ E _	HT(A)	_ _M _ /_ _ (M)_ /D(E)	Reas'd	1	2	3
2 Roberts	3.76	4.80	4.40	3.20	4.20	2.20	_ _GD_C		_ _ _ _/_ _ _ _/_ _		1	2	0
17 Koch	3.68	5.50	3.50	2.50	2.00	4.00	R_ G_ _ _		_ _ _ _/_ _ _ _/_ _		1	1	0
12 Nirenburg	3.63	2.67	2.67	3.00	5.00	4.83	_ S_ _ _ _	(H)TA	(R)_(M)_ /R _ M _ /D_		3	1	3
22 Butcher	3.33	3.67	3.67	2.67	4.00	2.67	R_ _ _ _ _	HT(A)	_ _ _ _/_ _ _ _/_ _		1	1	0

Comparison of assessment ratings with the results of the analyses shows a broadly consistent pattern of associated relationships between assessment and analysis. Generally, high assessment ratings are associated with supportive analytical findings. Supportive analysis was obtained from consideration of the wider RSGDEC requirements context, inclusion of the three Human, Teaming and Autonomy key HAT elements, strong balanced REMEDE core cognitive model comparison components, evidence of performed and reasoned effects, and high TRL, HRL and CML classifications. Again, of particular note is the outstanding work on Manned-Unmanned Teaming (13. Schmitt, TQ 6.00), Functional Division (5. Frost, TQ 5.84) and Policy Management (15. Thomas, TQ 5.80). All three were judged to be consistently high across the spectrum of Technical Assessments and Analyses.

The excellent technical work on Information Provenance (11. Luchero, TQ 5.97), proved to be an exception, along with the thought provoking, original concept work on Displaced Transparency (23. Miller, TQ 5.70). Both were highly rated on TQ, but with low TRL and HRL classifications.

5.0 CONCLUSIONS

The conclusions could be considered with reference to the Technical Evaluation criteria or the Symposium Session Themes. The Technical Evaluation criteria are already deeply analysed, and using them to derive generalised conclusions risk further being overly analytic. Referencing the Symposium Session Themes is more direct and more likely to be sensitive to context. The Symposium comprised five themes, namely Operational requirements; Human-autonomy teaming structure; Autonomous capabilities that support teaming; HAT interaction and design; HAT institutional integration. Whereas the themes seem appropriate, sensible and carry some logical progression, in practice they are not entirely mutually exclusive and overlap. Many of the contributions fit more than one or many of the themes. For example, all the contributions should have at least some operational requirements relevance. Many conclusions from the themes are embedded in the detailed content and analysis, indicating where the work and findings are significant, robust and exploitable. To summarise, the key enabling HAT technologies identified include perceptual processing, planning, learning, interaction, natural language understanding and multi-agent coordination. The focus is on interactions and interfaces for reliable and trusted HAT collaboration, providing situational awareness to operate in a complex battle space, exploiting large-scale synergistic teaming of manned and unmanned systems.

Further generalisations risk becoming superfluous and redundant. To provide added value, and perhaps a simpler, more integrating and higher level synthesis, we seek to reference the conclusions with respect to the US DSB 2012 Task Force Framework for design and evaluation of autonomous systems, replacing the unproductive Levels of Autonomy approach, with three critical considerations, namely Cognitive Echelons, Mission Timelines Dynamics and Complex Human-Machine Systems Trades Space [4].

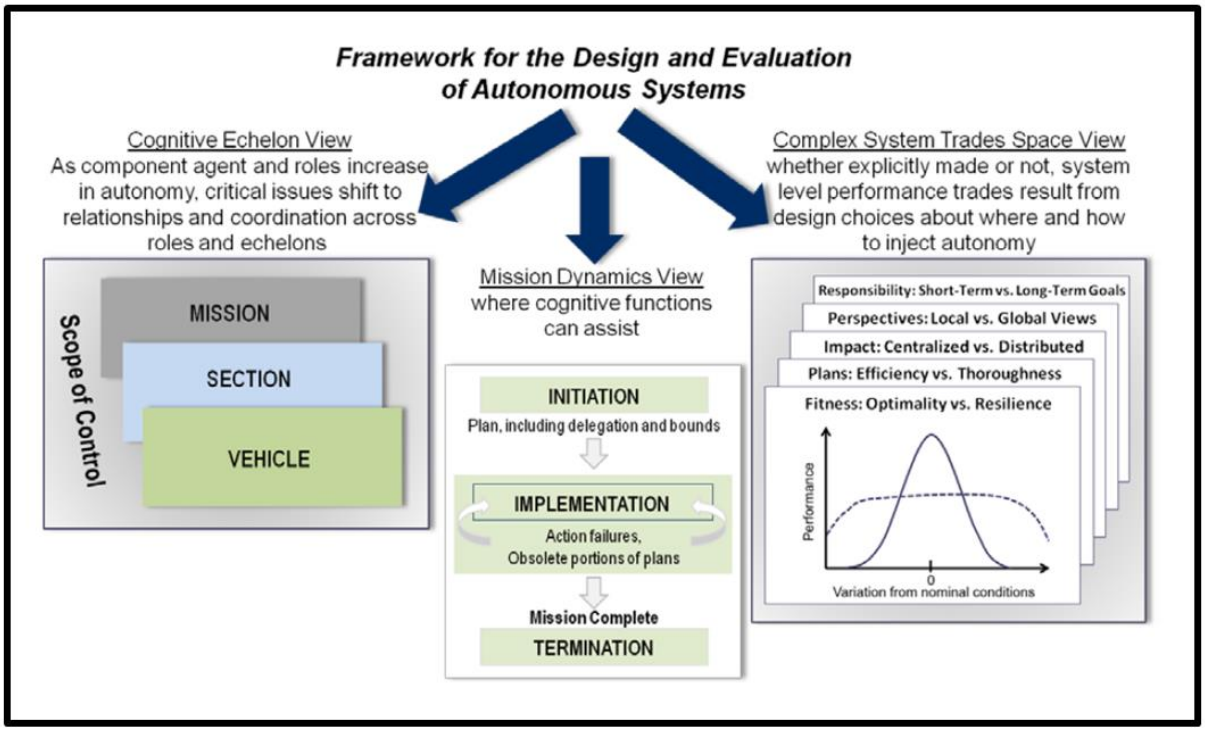


Figure 17: US Defense Science Board DSB 2012 Task Force Framework for the Design and Evaluation of Autonomous Systems [4]

5.1 Cognitive Echelons

The DSB 2012 cognitive echelon view refers to the perspective that "as component agent and roles increase in autonomy, critical issues shift to relationship coordination across the roles and echelons, as joint activity unfolds, with the scope of control involving mission, section and vehicle levels". HAT can support the scope of control for layered levels of users throughout the command structure, improving understanding, extending reach area depth, increasing SA for specific effects, and improving adaptation to unexpected events and surprises. There is extensive communication and coordination among commanders, team leaders and operators with the potential for each cognitive function to be allocated to, or shared between, the computer and the operator/supervisor.

The most extensive use of HAT has been at the lower echelon of vehicle/platform/sensor, supporting individual vehicle autonomy with less manpower. The scope of control involves controlling vehicle movement, sensor operations, communications and status monitoring. Typically, because of differing competency demands, and for reasons of operator workload management, vehicle control and sensor management/exploitation functions have been separately allocated as individual operator responsibilities. There is potential to reduce manpower with HAT autonomous capabilities. Anticipated autonomous capabilities include automated take-off and landing, waypoint navigation, automatic return to base upon loss of communications and path planning. Integration of these autonomy capabilities potentially frees up operator resources to focus on sensor management and real-time imagery exploitation.

Recent HAT research has focused on providing an individual operator with the ability to operate multiple platforms for many mission types and phases. This extends the scope of control and responsibility of the individual operator, evolving the role into a more supervisory control function for multi-vehicle/agent scheduling and collaboration, including mission planning and re-planning. At the higher mission

commander/executive officer level, the scope of control includes scenario assessment and understanding, scenario planning and decision making and contingency planning for off nominal conditions. At higher echelons, AI can autonomously fuse and abstract data, as well as manage, prioritize and route data, autonomously produce plans, anticipate failures and manage coordination, such as through policy management and negotiation, with other members in net-centric warfare.

DSB 2012 identified the following cognitive functions lacking in autonomous systems support provision to the mission commander/operator:

- Operator Level Situation Awareness
- Team Leader Level Failure Anticipation and Re-planning, and Multi-agent Communication, Collaboration
- Mission Commander Level Scenario Assessment and Understanding and Information/Network Management.

Other, under-utilised existing autonomy capabilities, well proven in laboratory and research settings included:

- Operator Level Fault Detection and Vehicle Health Management, and Communications
- Team Leader Level Mission Planning and Decision Making
- Commander Level Scenario Planning and Decision Making, and Contingency Management

The HAT Symposium reports evidence of significant progress in autonomous capabilities for Multi-agent Communication and Collaboration at Team Leader Level, and increased exploitation at Team Leader Level of capabilities in Mission Planning and Decision Making. Situational Awareness, as always, continues to be an Operator Level challenge working in a highly dynamic complex environment. This stretches the capabilities of human dynamic cognition, extended by the burden of supervising multiple autonomous agents, with weak mental modelling reasoning capabilities, and lacking transparency in trusted reasoning. The challenge of augmentation of Situation Awareness promises soon to be ameliorated in part by progress on AI sensor fusion capability and mission data fusion, with computer algorithms able to acquire, distil, organise and present otherwise disparate pieces of intelligence into a single picture. However, the picture of the problem still needs to be recognised, evaluated, and planned to be mitigated by an effective course of action decided involving human cognition.

5.2 Mission Dynamics

The DSB 2012 Mission Dynamics view refers to the perspective in which in which “autonomy may be employed in different ways for various mission phases and effects how different agents synchronize activities across mission phases, roles, and echelons as new events, disruptions, and opportunities arise”. The allocation of cognitive functions may vary over the course of a mission based on such factors as environmental complexity and required response time. The Symposium provides evidence of HAT with advanced autonomy technologies has been implemented in specific phases of missions, such as navigation route planning, search planning, imagery analysis, vehicle escort and overwatch, reducing operator workload, improving efficiency and effectiveness, and reducing error.

There is evidence of HAT contributing selectively across mission phases, with the Human-Autonomy balance tailored to the demands of tasks and to meet specific capabilities, including the ISR PED cycle (Processing, Exploitation, Dissemination), Time Sensitive Targeting F2T2EA (Find, Fix, Track, Target, Engage, Analyse), and MUM-T Strike FFS (Find, Fix, Strike).

The balance in the allocation of cognitive functions varies between the missions and phases, but also between the phases of cognition in decision making, as the REMEDE models comparison analysis shows. More data intensive tasks naturally tend to be associated with autonomous computational applications, improving efficiency and reducing manpower demands. This applies and to Intelligence Surveillance and Reconnaissance tasks, imagery analysis, and search and overwatch, seeking indications of certain patterns of activities and evidence of change. The analysis shows some indications that Autonomy bias applies to the Recognition and Evaluation phases of REMEDE.

The Symposium also provides evidence that HAT with increased autonomy capabilities can effectively support dynamic mission management and re-planning. This is achieved through techniques such as task monitoring and management, policy management and negotiation. These techniques provide the basis for adjustment, adaptation and flexibility during the mission in response to changes in tasks, targets, threats and tactics, including changed mission goals, airspace restrictions, and weather and vehicle performance parameters. HAT can contribute to efficiency and effectiveness in mission preparation and planning mission initiation phases, and to post mission de-briefing and analysis, mission termination phases. Importantly, through the application of Machine Learning, HAT can improve capability both on-line and off-line, with advantages of off-line ML for survivability, effectiveness and timeliness, as well as efficiency.

5.3 Complex Systems Trade Space

The DSB 2012 complex system trades space view refers to the perspective in which "design choices about where and how to inject autonomy changes how the larger system balances multiple performance trade-offs; the risk is that autonomy related improvements in one area can produce unintended negative consequences in other aspects of total system performance".

Five key level trades are identified that often dominating performance after system deployment, with different benefits and unintended consequences, namely:

- Fitness – Optimality versus Resilience
- Plans – Efficiency versus Thoroughness
- Impact – Centralised versus Decentralised
- Perspectives – Local versus Global view
- Responsibility – Short-term versus Long-term goals

DSB 2012 points out that system trades made without consideration of their implications can lead to many unintended consequences, such as higher manpower and training costs, avoidable collateral damage, and failures attributed to "human error" and underutilization.

The Symposium contributions are less readily appreciated in terms these trade spaces. Trade space reasoning is perhaps an aspect of the body of work that needs further consideration. The work on HFI organisation risks, and on Ergonomics Standards, tackles some of the human resources trades, risks and consequences. Arguably, HAT, by focussing on autonomous capabilities that support teaming, and optimising human and machine capabilities in teams that maximise strengths and mitigate weaknesses contributes to the balancing of the trades for Fitness – Optimality versus Resilience. This might apply for example in the work on collision avoidance, imagery analysis, search planning and on measuring effectiveness, and arguably on Ergonomics standards and HFI risks, Also, consideration of complementariness in capabilities balances the trades for Plans - Efficiency versus Thoroughness. This might apply particularly through the work on Information Provenance, Policy Management and Negotiation, Patterns and Work Agreements. Work on Manned-Unmanned Teaming, Multi UxV C2, Command Management and Functional Division, under Human-Autonomy teaming structure, provide means of balancing the trades for Impact – Centralised versus Distributed, and trades for Perspectives

– Local versus Global Views. The work on Social Relationships and Social Intelligence touches on the benefits of distributed functioning. For Perspectives, arguably, HAT Interaction and Interface Design, together with big picture AI sensor and data fusion, affords the possibility of affording both Local and Global views. The consequences of HAT for Responsibility – Short term versus Long term goals – are not immediately obvious and possibly covered in part by the work on ethics, patterns and work agreements. Responsibility is covered more directly by the benefits of HAT supporting C2 cognitive echelons, and perhaps by understanding interdependencies and the trust building dimension, including trust maintenance and repair, trust transparency and trust attitudes. This needs more consideration and analysis.

Probably the only trade analysis reported directly in the Symposium bearing on Responsibility, concerned the benefits of Transparency and Explainable AI for understanding the efficacy of autonomy functioning, and for trust and reliance, associated to some extent with the work on communication dialogue and natural language. The benefits need to be weighed against the costs of information for human cognition attention and workload, which allocation of cognitive functions to autonomy is intended to reduce. Displaced Transparency, emphasising reliability analysis before and after mission, is proposed as the efficient and probably effective Human Cognition solution, judged as most worthy of potential exploitation. Collectively, this analysis can be considered as providing evidence of the relevance, veracity and power of HAT as a means of resolving conflicts in systems level design trades for determining the role of autonomy in systems.

Table 10. DSB 2012 Autonomous System Reference Framework Trades Space

DSB 2012 AUTONOMY TRADE SPACE ANALYSIS				NATO SCO HFM HAT SY-300
TRADE SPACE	**TRADES**	**BENEFITS**	**UNITENDED CONSEQUENCES**	
Fitness	Optimality vs Resilience	More precise for understood situations	Increased brittleness	Collision avoidance; Imagery analysis; Search planning; Measuring effectiveness; Ergonomics; HFI risks.
Plans	Efficiency vs Thoroughness	Balanced use of computational resources	Locked into wrong plan/difficulty revising plan	Information provenance; Policy management; Search planning; Imagery analysis; Patterns. work agreements; Multi UxV C2.
Impact	Centralised vs Distributed	Ability to tailor actions to appropriate echelon	High cost of coordination	Functional division; Command management; Manned-unmanned teaming; Multi UxV C2; Social relationships; Social intelligence.
Perspectives	Local vs Global Views	Ability to balance scale/area of action with resolution	Data overload; reduced speed of decision making	Functional division; Command management; Multi UxV C2; Manned-unmanned teaming; Measuring effectiveness; (+ Sensor Data Fusion).
Responsibility	Short-term vs Long-term Goals	Builds trust by tailoring risk management to goals, priorities, context	Breakdown in collaboration and coordination	Interdependency analysis; Patterns, work agreements; Transparency, trust; Displaced transparency; Trust attitudes; Communication dialogue; Natural language; Ethics.

6.0 FUTURE DIRECTIONS

Before proceeding, it should be noted that at the time of writing, the HAT community at large await with anticipation the publication of the Final Report of HFM SCO Task Group 247 on Human Autonomy Teaming, 2016-2018. This Final Report on HAT should significantly inform thinking on view points for future directions, based on a more considered, multi-author analysis reporting in detail the state of the art, presumably along with views on possible futures. There is a greater risk of bias in trusting guidance for direction from the Symposium Technical Evaluation Report. The TER is based on a relatively random sample of the work, submitted to the Symposium, reasonably but not entirely representative of current National Research

Programmes on HAT. The Symposium is a relatively transient affair, providing a more momentary snapshot, compared with the veracity of full spectrum, collective advice from the Nations' 3-year standing team of SMEs and their Chairman leadership Dr Mark Draper. Nevertheless, the TER is directed by the SCO HFM Panel to provide a view on Future Directions. Though not reflecting directly the unseen HFM 247 Task Group Final Report, this projection inevitably will be influenced significantly from the TER author's participation in this most recent, and other previous Task Groups relevant to HAT and earlier prescriptions. Much like Autonomy Transparency, consideration of future directions can be approached also from displaced perspectives, adding to understanding from the current state of the art, view-points derived from existing retrospective and emerging prospective optics.

5.1 Retrospective

In 2016, NATO HFM Task Group 247, on a sub-task led by UK, conducted a survey of Nations' current activities on HAT. The survey compared progress against issues identified in the 2007 Final Report of HFM 078 UMV HF Issues in Augmenting the Force [15]. The technical details and results are illustrated in the attached graphic. Essentially, it was identified that progression from prioritisation in 2007 on MUxS C2 issues, to priorities in 2016 for HAT, indicated a socio-cognitive shift, with associated risks for implementation through System Engineering, and for application measured by Military Benefit.

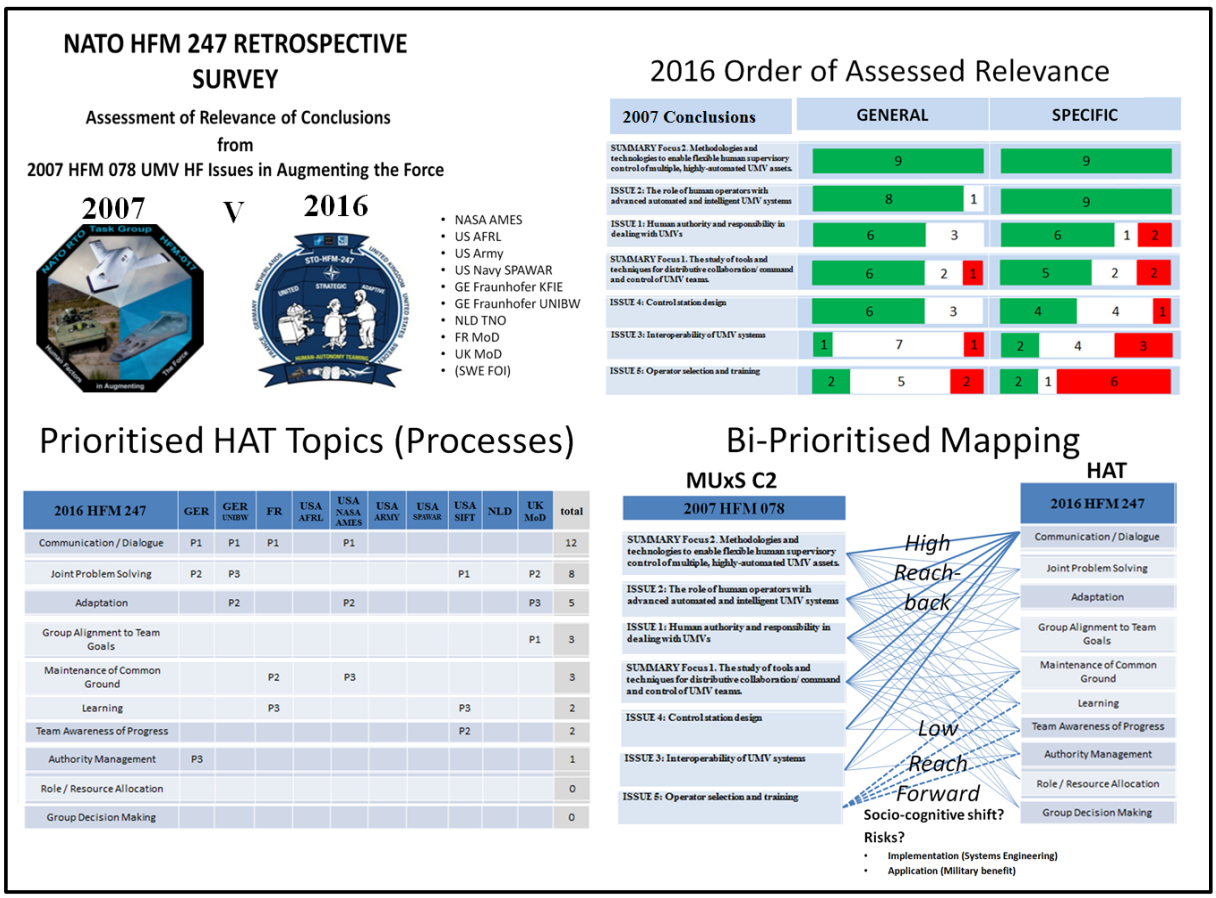


Figure 18: NATO HFM Task Group 2016 Retrospective Survey

In 2006, a socio-cognitive shift was perceptively anticipated in the Keynote from Professor Rene Amalberti, IMASSA France, presented to the Biarritz "UMV Force Multipliers" Symposium [45] as Capstone to HFM-

078/ TG017 [15]. Professor Amalberti identified three distinctive strategic research horizons needing to be addressed in parallel:

- A short term horizon centred on UAVs design related problems (level of automation Vs level of human supervision)
- A mid-term horizon (5 to 10 Yrs) centred on the system approach and the mission (role, articulation, selection, and education of all actors concerned with UAVs operations, design of operating protocols for normal and abnormal situations, design of distributed cooperative interfaces, interoperability),
- A long term horizon focused on the social impact of mass introduction of UAVs UCAVS in the military forces.

5.2 Current State of the Art

For consideration of the 2018 state of art for HAT, this can be simply summarised by referring to the evidently strong work and good progress reported at the NATO SCO HFM SY-300 HAT Symposium, identified by the following areas, balancing both domain agnostic fundamental work and domain specific applications:

- Manned-unmanned teaming
- Information provenance
- MUxV C2 Functional division
- Policy management
- Imagery analysis
- Search planning
- Social intelligence
- Displaced transparency
- Effectiveness metrics
- Command management

5.3 Prospective

Switching to a forward view, the analysis of the Symposium proceedings has provided a characterisation of status expressed in terms of capability maturity progression, clearly indicating short falls and the scope for further progression. The majority of the work can be considered as having progressed to, but not beyond, a Managed level of HAT capability maturity, adaptable perhaps, rather than adaptive. Thus, measured in these terms, there remains considerable scope for progression to Optimised, and ultimately to Adaptive levels of capability maturity, as defined and described in the illustration below.

The capabilities identified as priorities in the work of HFM TG 247 should be regarded as potentially transformative in evolving progression to these higher levels of HAT capability maturity.

- Communication/Dialogue
- Joint problem solving
- Group alignment to team goals
- Adaptation
- Maintenance of common ground

- Team awareness of progress
- Authority management
- Role/Resources allocation
- Group decision making

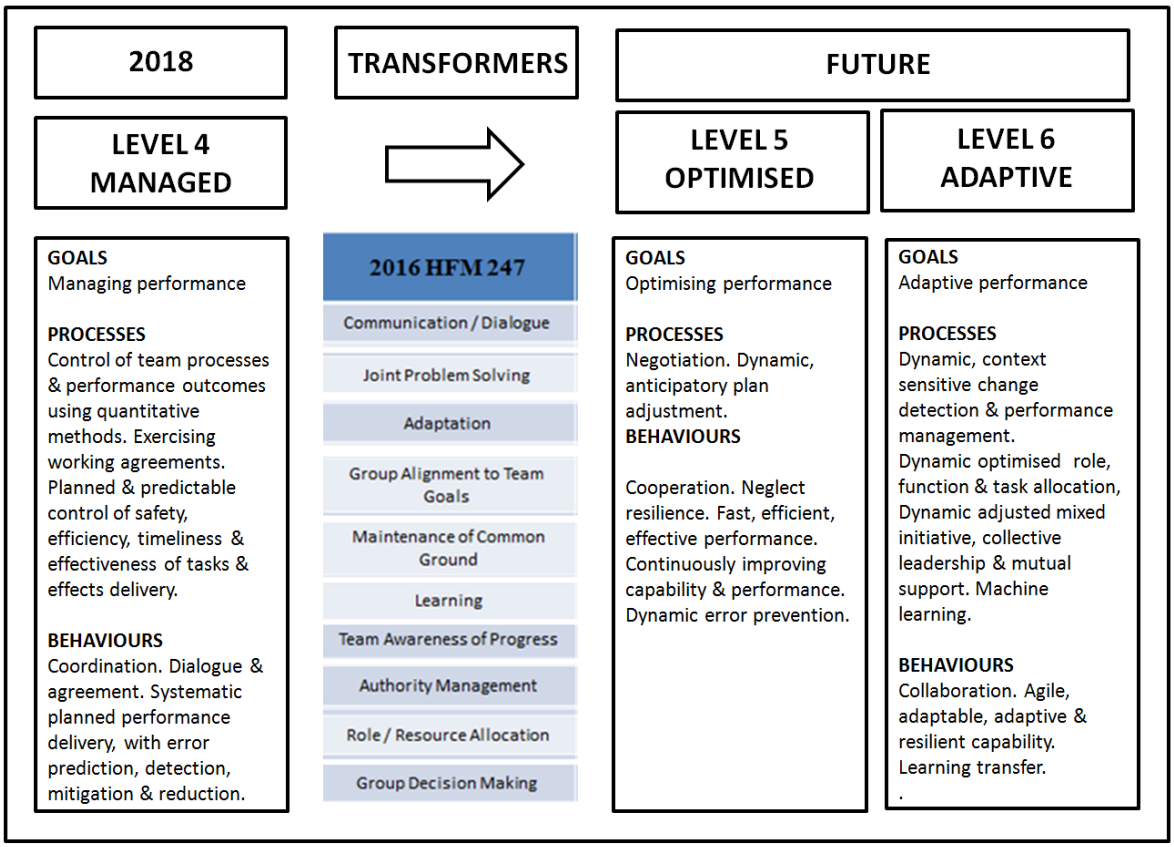


Figure 19: Means for Progression to Teaming Goal of Adaptive Performance

In addition to these capabilities, we can add now the concepts, principles, methods and technologies underpinning the leading work identified above in the 2018 State of the Art review, repeated below under Prospectives, for completeness.

- Manned-unmanned teaming
- Information provenance
- MUxV C2 Functional division
- Policy management
- Imagery analysis
- Search planning
- Social intelligence
- Displaced transparency

- Effectiveness metrics
- Command management

Dialogue Communication, covered in 2018 HFM SY 300 Symposium Programme, and believed to be strongly associated with CML 6, already features in the HFM TG 247 list.

These enabling capabilities should be regarded as proven, reliable and valid indicators of required lines for future development of adaptive teaming performance. Logically, the aim of future work should be to continue to coordinate, cooperate and collaborate successfully, as has been demonstrated under HFM TG 247, so as to make further progress, towards the ultimate aim and objective. That is to develop and share understanding on development and testing, of methods and techniques, technologies and systems, providing proof of efficacy and effectiveness for achievement of the ultimate goal of Intelligent Adaptive Collaborative Teaming defined by the following characteristics:

Processes

- Dynamic, context sensitive change detection and performance management
- Dynamic optimised role, function and task allocation,
- Dynamic adjusted mixed initiative
- Collective leadership and mutual support
- Machine learning

Behaviours

- Collaboration
- Agile, adaptable, adaptive and resilient capability
- Learning transfer

The principal research requirement remains to be able to identify the defining characteristics of effective teaming relationships in a human and autonomy capability force mix, optimising performance adaptively whilst mitigating cognition and automation bias. The aim is to achieve the Teaming goal of Adaptive Performance, as illustrated in Figure 20, using the results of the SY-300 REMEDE analysis (Fig 14) to indicate the current transitioning HAT models comparison balance (N/20). Achieving the correct balance of expertise between domain agnostic fundamental work and domain specific applications, in a joint collective learning endeavour, coupled with strong military advisor involvement, is critical for success in future NATO HFM Task Groups and activities working in this complex and challenging domain.

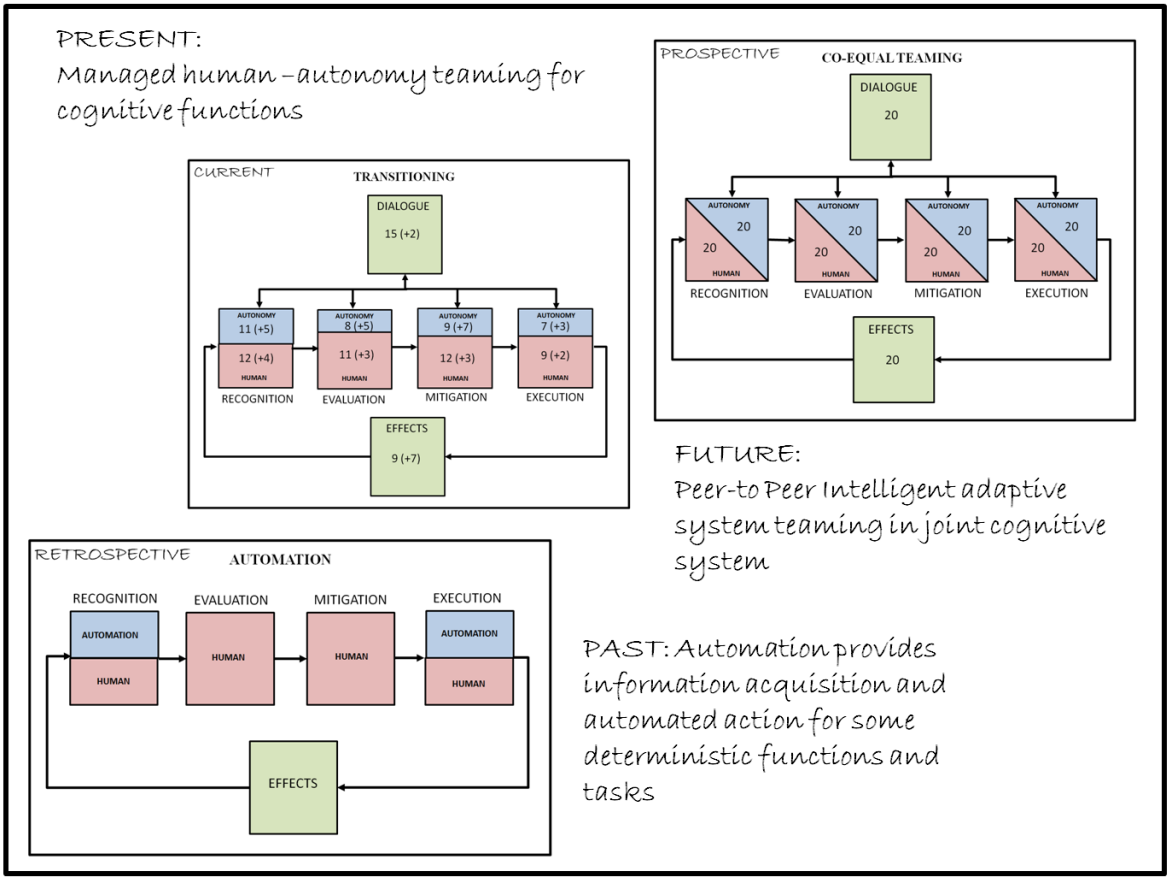


Figure 20: Past, Present and Future Perspectives

7.0 REFERENCES

[1] DCDC 2018. Human Machine Teaming. UK MoD Joint Concept Note 1/18. UK Ministry of Defence, Development, Doctrine and Concepts Centre. May 2018.

[2] Hancock, P. A. 2013. Automation: How much is too much? Ergonomics, *57*(3), 449-454.

[3] Hagel C. 2014. Secretary of Defense Speech: Regan National Defense Forum Keynote, Ronald Reagan Presidential Library, Simi Valley, CA. 15 March 2014. dod.defense.gov.

[4] DSB 2012. The Role of Autonomy in DoD Systems. DSB Task Force Report. United States Department of Defense, Defense Science Board Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, Washington D.C. 20301-3140, July 2012.

[5] Weiner N., 1948. Cybernetics: Or control of communication in the Animal or Machine. Cambridge, Massachusetts: MIT Press.

[6] Small R., Lizza C.S. and Zenyuh. 1989. The Pilot's Associate: Today and Tomorrow. In, *The Human-Electronic Crew: Can They Work Together?* Proceedings of the 1st Joint GAF/RAF/USAF Workshop on Human-Computer Teamwork.Stadttheater, Ingolstadt, FRG., 19-22 September 1988. Pp 133-138. Eds: Emerson T., Reising J., Taylor R.M., and Reinecke M. WRDC-TR-89-7008. AD-A211 871. Wright

Research and Development Center, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, 45433-6523, US. July 1989

[7] Rasmussen J. 1986. Information Processing and Human- Machine Interaction: An Approach to Cognitive Engineering. New York: North Holland

[8] Hollnagel E. 2007. Modelling Multi Layered Control: Application of the Extended Control Model to the Analysis of UAV Scenarios. In, *Uninhabited Military Vehicles (UMVs): Human Factors in Augmenting the Force.* Final Report of RTO Human Factors and Medicine (HFM) Task Group HFM-078/TG017. Chapter 7.7, Human Automation Integration, pp 73 to 96. AC/323(HFM-078)TP69. ISBN 978-92-837-0060-9. July 2007.

[9] Emerson T., Reising J., Taylor R.M., and Reinecke M. The Human-Electronic Crew: Can They Work Together? Proceedings of the 1st Joint GAF/RAF/USAF Workshop on Human-Computer Teamwork.Stadttheater, Ingolstadt, FRG. 19-22 September 1988. WRDC-TR-89-7008. AD-A211 871. Wright Research and Development Center, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, 45433-6523, US. July 1989

[10] Emerson T., Reinecke M., Reising J., and Taylor R.M. 1992. The Human-Electronic Crew: Is The Team Maturing? Proceedings of the 2nd Joint GAF/RAF/USAF Workshop on Human-Computer Teamwork.Stadttheater, Ingolstadt, FRG. 25-28 September 1990. WRDC-TR-89-3078. AD-A256 192. Wright Research and Development Center, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, 45433-6523, US. 10 July 1992

[11] Taylor R.M. and Reising J. 1995. The Human-Electronic Crew: Can We Trust The Team? Proceedings of the 3rd Joint GAF/RAF/USAF Workshop on Human-Computer Teamwork, Cambridge, UK. 27-30 September 1994. WL-TR-96-3039. DRA/CHS/HS3/TR95001/01. AD 19960524 018. Flight Dynamics Directorate, Wright Laboratory, Air Force Material Command, Wright-Patterson Air Force Base, Ohio, 45433-7562, US. December 1995

[12] Reising J., Taylor R.M., and Onken, R.. 1999. The Human-Electronic Crew: The Right Stuff Proceedings of the 4th Joint GAF/RAF/USAF Workshop on Human-Computer Teamwork. Hotel zur Post, Kreuth, Nr Tegernsee, Germany 23 - 26 September, 1997. Authors WL-TR-AFRL-HE-1999-0235. Human Effectiveness Directorate, Wright Patterson AFB, Ohio. December 1999.

[13] Taylor R.M. and Reising J. 1998. The Human-Electronic Crew - Human Computer Collaborative Teamwork. In, Collaborative Crew Performance in Complex Operational Systems. Proceedings of RTO HFM Symposium, Edinburgh, UK., 20-22 April 1998. NATO RTO MP-004. 22–1 to 17

[14] NATO RTO 2006. Human Factors of the Uninhabited Military Vehicles as Force Multipliers. RTO Meeting Proceedings MP-HFM-135. RTO Human Factors and Medicine Panel (HFM) Symposium, Biarritz, France, 9-11 October 2006. AC/323(HFM-135)TP/70. December 2006. NATO Research and Technology Organisation. BP 25, F-92201 Neuilly-sur-Seine Cedex, France.

[15] NATO RTO 2007. Uninhabited Military Vehicles (UMVs): Human Factors in Augmenting the Force. TR-HFM--078. Final Report of RTO Human Factors and Medicine (HFM) Task Group HFM-078/TG017. Ed T. Barry. AC/323(HFM-078)TP/69. ISBN 978-92-837-0060-9. July 2007. NATO BP25, F-92201 Neuilly-Sur-Seine, Cedex, France.

[16] NATO RTO 2012. Supervisory Control of Multiple Uninhabited Systems – Methodologies and Enabling Human-Robot Interface Technologies. TR-HFM-170. Final Report of RTO Human Factors and

Medicine (HFM) Task Group HFM-170. Ed L. Van Breda. AC/323(HFM-170)TP/451. ISBN 978-92-837-0167-5. December 2012. NATO BP25, F-92201 Neuilly-Sur-Seine, Cedex, France.

[17] NATO STO 2012. Supervisory Control of Multiple Uninhabited Systems Methodologies and Human-Robot Interface Technologies: A Workshop Examining the Human Factors of Multi-Uninhabited Vehicle Supervisory Control. NATO STO Human Factors and Medical Panel(HFM) Workshop, Prague, Czech Republic, 8-10 May 2012. STO-MP-HFM-217. AC/323(HFM-217)TP/500. ISBN 978-92-837-0184-2. December 2012. NATO BP25, F-92201 Neuilly-Sur-Seine, Cedex, France.

[18] Patzek, M., Zimmer, D. and Feitshans, G. 2012. Chapter 14 – US-1: Multi-UAV Supervisory Control Interface Technology (MUSCIT) Demonstration. In, *Supervisory Control of Multiple Uninhabited Systems – Methodologies and Enabling Human-Robot Interface Technologies*. TR-HFM-170. p14-1 to 13. Ed. Van Breda, L. AC/323(HFM-170)TP/451 NATO STO Human Factors and Medical Panel. ISBN 978-92-837-0167-5. December 2012. NATO Research and Technology Organisation. BP 25, F-92201 Neuilly-sur-Seine Cedex, France.

[19] Eggers J., Kelchner, R. 2012. Future Vision for Unmanned System Operations within NATO: Leveraging Autonomous Behaviours to Manage Complex Systems. In, *Supervisory Control of Multiple Uninhabited Systems Methodologies and Human-Robot Interface Technologies: A Workshop Examining the Human Factors of Multi-Uninhabited Vehicle Supervisory Control*. NATO STO Human Factors and Medical Panel(HFM) Workshop, Prague, Czech Republic, 8-10 May 2012. STO-MP-HFM-217. AC/323(HFM-217)TP/500. ISBN 978-92-837-0184-2. December 2012. NATO BP25, F-92201 Neuilly-Sur-Seine, Cedex, France.

[20] Taylor R., Grabham, A., Platts J., & Cottrell, R. 2012. Dynamic Airborne Mission Management: Technology Demonstration 3. In, *Supervisory Control of Multiple Uninhabited Systems Methodologies and Human-Robot Interface Technologies: A Workshop Examining the Human Factors of Multi-Uninhabited Vehicle Supervisory Control.* Prague, Czech Republic, 8-10 May 2012. AC/323(HFM-217)TP/500. STO-MP-HFM-217. NATO STO Human Factors and Medical Panel. ISBN 978-92-837-0184-2

[21] Schmorrow, D., Stanney K.M., Wilson G., and Young, P. 2006. Augmented Cognition in Human System Interaction. In, *Handbook of Human Factors and Ergonomics*, Third Edition, Ed Gavriel Salvendy, Chapter 52, pp 1364-1383. John Wiley and Sons, Hoboken, New Jersey.

[22] Osborn, K. 2017. Air Force Chief Scientist confirms F-35 will include artificial intelligence. Defense Systems, January 2017. McLean, VA. https://defensesystems.com/articles/2017/01/20/f35.aspx?m=2

[23] Eggers, J.W. and Draper, M.H. 2006. Multi-UAV Control for Tactical Reconnaissance and Close Air Support Missions: Operator Perspectives and Design Challenges. In *Human Factors of Uninhabited Military Vehicles as Force Multipliers.* Pp KN1-1 – KN1-20). Meeting Proceedings RTO-MP-HFM-135, Keynote 1. December 2006. NATO BP25, F-92201 Neuilly-Sur-Seine, Cedex, France.

[24] Draper M., H. and Van Breda L. 2012. Chapter 1 - Introduction. Chapter 18 – Conclusions. In, *Supervisory Control of Multiple Uninhabited Systems – Methodologies and Enabling Human-Robot Interface Technologies*. TR-HFM-170. Pp 1-1 to 6. Pp 18-1 to 4. Final Report of RTO Human Factors and Medicine (HFM) Task Group HFM-170. Ed L. Van Breda. AC/323(HFM-170)TP/451. ISBN 978-92-837-0167-5. December 2012. NATO BP25, F-92201 Neuilly-Sur-Seine, Cedex, France.

[25] Taylor R.M., Carver G., Kennett P. and Stensson P. 2007. Military Relevance. In *Uninhabited Military Vehicles (UMVs): Human Factors in Augmenting the Force*. RTO-TR-HFM-078-02 Chapter 2. pp 1-

52. Final Report of the RTO Human Factors and Medicine Panel (HFM) Task Group HFM-078/TG-017. July 2007. NATO RTO. F-9220 1 Neuilly-sur-Seine, CEDEX France. ISBN 978-92-837-0060-9.

[26] Keirl H, Taylor R, Grabham A, Thomas, M. and Halpin, D. Military Relevance. In, Final Report, Human Autonomy Teaming, NATO SCO Task Group 247. In press, January 2019.

[27] Acosta, H, 2010. Human readiness levels: Implementing HSI – Connecting some dots. Panel discussion presented at the 81st Annual Meeting of the Aerospace Medical Association, Phoenix, AZ.

[28] Endsley, M. 2015. Human readiness levels: Linking S&T to acquisition. Parer presented to the 2015 National Defense Industrial Association (NDIA) Human Systems Conference, Alexandria VA. (Human Systems Integration: Challenges and Opportunities. http://www.acq.osd.mil)

[29] Talamadupula, K., Briggs, G., Chakraborti, T., Scheutz, M., & Kambhampati, S. 2014. Coordination in human-robot teams using mental modelling and plan recognition. In: IROS 2014, pp. 2957-2962, IEEE

[30] Chakraborti, T., Sreedharan, S., Zhang, Y., & Kambhampati, S. 2017. Plan explanations as model reconciliation: Moving beyond explanation as soliloquy. In Proceedings of IJCAI, pp. 156-163.

[31] Scheiber, A. Th., Akkermans, J., Anjewierden, A., De Hoog, R., Shadbolt, N., Van De Velde, W., & Wielinga, B. 1999. Knowledge Engineering and Management: The CommonKADS Methodology. Cambridge MA: MIT Press.

[32] Vicente, K.J. 1999. Cognitive Work Analysis: Towards Safe, Productive, and Healthy Computer based Work. New Jersey: Lawrence Erlbaum.

[33] Sanderson, P., Naikar, N., Lintern, G., and Goss, S. 1999. Use of cognitive work analysis across the system life cycle: From requirements to decommissioning. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, Houston, TX. Santa Monica: HFES.

[34] Jessie Y.C. Chen, Shan G. Lakhmani, Kimberly Stowers, Anthony R. Selkowitz, Julia L. Write, and Michael Barnes. Situation Awareness-Based Agent Transparency and Human-Autonomy Teaming Effectiveness. In Theoretical Issues in Ergonomics Science, vol. 19, iss. 3, pp. 259—282, 2018.

[35] Entin, E., & Serfaty, D. 1999: Adaptive team coordination. Human Factors, 41, pp. 312–325.

[36] Brehmer, B. 2005. The Dynamic OODA Loop: Amalgamating Boyd's OODA Loop and the Cybernetic Approach to Command and Control. Assessment Tools and Metrics. In Proceedings of the10th International Command and Control Research and Technology Symposium. The Future of C2. pp 365-368.

[37] Taylor R.M. 2015. Human Autonomy Teaming: Verification, Benchmarking and Readiness. In, RPAS Achievements and Challenges. Proceedings of the 2015 Presidents Conference, Royal Aeronautical Society, London' 7-8 October 2015. SG UAS Event Code 772. ID P2PP2R-2015-09-28T 10:34:36. DSTL/TR91332 967705572. ISBN 1 8576831X.

[38] Taylor, R.M., Keirl, H., Thorpe E. and Grabham A. 2018. UK-1: HAT-CV&V Human Autonomy Teaming – Collaboration Verification and Validation. Final Report NATO HFM247. In Press, January 2019.

[39] Carnegie Mellon. 2002. Capability Maturity Model® Integration (CMMISM), Version 1.1. CMMISM for Software Engineering (CMMI-SW, V1.1). Continuous Representation. CMU/SEI-2002-TR-028.

ESC-TR-2002-028. CMMI Product Team, August 2002. Carnegie Mellon Software Engineering Institute, Pittsburgh, PA 15213-3890

[40] Miller, C. and Goldman, R. 1999. Tasking Interfaces that Know Who's the Boss. In, The Human-Electronic Crew: The Right Stuff Proceedings of the 4th Joint GAF/RAF/USAF Workshop on Human-Computer Teamwork. Hotel zur Post, Kreuth, Nr Tegernsee, Germany 23 - 26 September, 1997. Authors John Reising, Robert Taylor, and Reiner Onken. WL-TR-AFRL-HE-1999-0235. Paper 12, pp 97-102. Human Effectiveness Directorate, Wright Patterson AFB, Ohio. December 1999.

[41] Parasuraman R., and Riley V., 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors, 1977, 39(2), pp 230-253.

[42] Lyons, J. 2012. Trust and Influence. Air Force Research Laboratory, Wright Patterson AFB., OH 45433. 05 March 2012. https://apps.dtic.mil/dtic/tr/fulltext/u2/a563291.pdf

[43] Taylor R.M 2002. Technologies for Supporting Human Cognitive Control. In Human Factors in the 21st Century. NATO MP-077. 18:1-14. RTO Human Factors and Medicine Panel (HFM) Specialists' Meeting, Paris, France, 11-13 June 2001.

[44] Taylor, R.M 2007. Human Automation Integration with Contractual Autonomy. In, Uninhabited Military Vehicles (UMVs): Human Factors in Augmenting the Force. Final Report of RTO Human Factors and Medicine (HFM) Task Group HFM-078/TG017. Chapter 7, Human Automation Integration, pp 3-12 to 29. AC/323(HFM-078) TP69. ISBN 978-92-837-0060-9 July 2007.

[45] Amalberti, R. 2006. Points des recherches socio-cognitives à conduire sur les UAVs. Key Socio-Cognitive Research Issues Associated with Uninhabited Air Vehicles (UAVs). In, Human Factors of Uninhabited Military Vehicles as Force Multipliers. RTO Meeting Proceedings MP-HFM-135, KN 2:1 to 12. RTO Human Factors and Medicine Panel (HFM) Symposium, Biarritz, France, 9-11 October 2006. AC/323(HFM-135)TP/70. December 2006. NATO Research and Technology Organisation. BP 25, F-92201 Neuilly-sur-Seine Cedex, France.

8.0 ANNEX

A1. SUMMARY DESCRIPTIONS OF NATO SCO HFM SY300 HAT CONTRIBUTIONS

ASSESSMENT	ANALYSIS			
TQ: Technical Quality OR: Operational Relevance Ex: Exploitation TC: Technical Capability SQ: Scientific Quality In: Innovation TRL: Technology Readiness Level HRL: Human Readiness Level CML: Capability Maturity Level	RSGDEC R: Requirement Drivers S: CSE Solutions G: Functional Governance D: Development Enablers E: Performative Effects C: Enabled Capabilities	HTA H: Human T: Teaming A: Autonomy	REME/REME/DE R: Recognition E: Evaluation M: Mitigation E: Execution /D: Dialogue E: Effects REME/ : Human /REME : Autonomy	Moderators ( ) : Weak - : Not applicable

SESSION 1: OPERATIONAL REQUIREMENTS AND MEASURING HAT EFFECTIVENESS		
No.	AUTHORS	TITLE
1	Rohling.	Supporting Infantry Units with UGV and UAS
Presentation only. No Documentation. Description: C2 UxV collaboration autonomy technology systems; command chain management model, prototype, operational user command templates; AI assisted reporting interaction interfaces experimental test; air-land battle recce, operational user test environment; observed qualitative performative effects. Assessment: TQ 5.00; OR 5.80; Ex 5.60; TC 5.40; SQ 4.00; In 4.00 Analysis: [RSGDEC]; HTA; [REME/REME/DE]; TRL 7; HRL 5; CML 4; Performed effects.		
2	Roberts, Soper.	Identifying HF considerations associated with command, control, communication of autonomous vehicles in the maritime domain
Description: Maritime UxV C3 MAPLE architecture; task model, dataflow, functional de-composition; realistic (feasible, desirable), high autonomy; stakeholder HFI workshop analysis results; probability /impact of risks, issues, assumptions, dependencies, opportunities; HFI Social/Organisational, Manpower, Personnel, Training, HF Engineering, Health Hazards, System Safety. Assessment: TQ 3.76; OR 4.80; Ex 4.40; TC 3.20; SQ 4.20; In 2.20 Analysis: [- - GD - C]; [- - -; - - - -/- - - -/ - -]; TRL 1; HRL 2; CML 0		
3	Keirl, Thorpe, Taylor, Grabham.	Measuring effectiveness of HAT
Description: C2 UAS collaboration and MUM-T autonomy technology systems; defined, bench-marked, subjective, objective metrics for tasks, teaming, mission decision adaptivity effectiveness; HAT trustworthiness, capability maturity, risks; structured metrics model; participant, observer protocols; advanced decision models comparison prototype; operational user interaction, decision sensitive, validation testing; usability, reliability, sensitivity, discrimination, diagnostic, prognostic power; relevant simulation multi-mission air-land integration, layered C2/ISR environments; live flight trial, operational user test; full HAT spectrum subjective/objective performative effects. Assessment: TQ 5.12; OR 5.80; Ex 5.60; TC 5.40; SQ 4.40; In 4.40 Analysis: [RSGDEC]; HTA; [REME/REME/DE]; TRL 7; HRL 5; CML 3; Performed effects.		
4	Johnson, Vignati, Duran.	Understanding human autonomy teaming through interdependence analysis
Description: HAT generic, proof of concept, joint activity, formative design analysis; model, prototype development; human-automation interdependence support relationships; teaming diagnostics for design choices; human, technology factors to enhance/inhibit teaming effectiveness; Traffic Collision Avoidance System and Detect and Avoid System demonstration application. Assessment: TQ 4.80; OR 4.80; Ex 5.00; TC 5.00; SQ 4.80; In 4.40 Analysis: [- S - - E -]; HT(A); [REME/(REME)/D-] TRL 3; HRL 3; CML 4; Recognised effects.		

SESSION 2: HUMAN AUTONOMY TEAM STRUCTURE AND COMMUNICATIONS (1)		
No.	AUTHORS	TITLE
5	Frost, Bartik, Calhoun, Spriggs, Ruff, Behymer.	Creating a well situated human autonomy team: The effects of team structure
Description: C2 UxV autonomy technology systems managed collaboration; cooperative control planning optimisation algorithms; organisation resources tasking model definition, prototypes; simulated human-autonomy systems collaboration partnering; relevant air-land-maritime, base protection simulation, operational user test environment; operational user interaction experimental test plan; initial observed qualitative performative effects. Assessment: TQ 5.84; OR 6.00; Ex 5.80; TC 6.40; SQ 5.80; In 5.20 Analysis: [RSGDEC]; HT(A); [REME/(REME)/DE]; TRL 6; HRL 5; CML 3; Performed effects.		
6	Shively, Coppin, Lachter.	Bi-directional communication in human-autonomy teaming
Description: HAT generic, critical communication component for effective adaptive human-autonomy teaming; initial outline of purpose, objectives, requirements rationale and sub-component processes, functions for a bi-directional operator/agent communication manager; enabling comprehension and adaptiveness; simple application use cases with functional sub-components. Assessment: TQ 4.45; OR 4.75; Ex 4.25; TC 5.00; SQ 4.25; In 4.00 Analysis: [- S - - E -]; HTA; [REME/- - - - /D-]; TRL 2; HRL 1; CML 6. Recognised effects		

SESSION 3: HUMAN AUTONOMY TEAM STRUCTURE AND COMMUNICATIONS (2)		
No.	AUTHORS	TITLE
7	Baltzer, Lopez, Flemisch.	Human autonomy teaming using cooperative automation, interaction patterns and image schemes
Description: Vehicle automation assistance managed guidance and control technology sub-system; interaction mediation and cooperative guidance, model and prototype, operational user interaction experimental test; relevant land vehicle collision avoidance simulation, operational user test environment; subjective, objective performative effects. Assessment: TQ 4.53; OR 4.50; Ex 4.75; TC 4.08; SQ 5.25; In 4.08 Analysis: [- S – DE -]; HTA; [REME/REME/DE]; TRL 6; HRL 5; CML 4. Performed effects		
8	Neerincx, Diggelen, Stolk	Design patterns and ontologies for situated human agent collaboration, organisation, team and interaction level
Description: C2 UxV managed collaboration concept; *ePartners* agent autonomy technology; interaction design at organisation, team and individual levels; context adaptive, dynamic teamwork task allocation, coordination; cognitive engineering, perceptual cognitive framework; workload, SA situation cognition models; design patterns, ontologies, adjustable work agreements, harmonised interactive notifications; explanation model learning re-use capability evolution. Relevant, robot assisted disaster response application prototype simulation demonstration test environment. Assessment: TQ 4.97; OR 4.67; Ex 5.00; TC 5.50; SQ 5.17; In 4.50 Analysis: [- SGDEC]; HT(A); [-EME/(REME)/D-]; TRL 2; HRL 2; CML 4; Recognised effects.		
9	Kelley, Ouimet.	On human assisted multiagent-based planning
Description: C2 UxV managed collaboration; AI search planning optimisation sub-system technology; model and prototype; user-agent interaction interface; user experimental test plan; relevant maritime underwater search laboratory simulation operational user test environment; test plan includes subjective, objective performative effects. Assessment: TQ 5.62; OR 5.33; Ex 5.33; TC 6.33; SQ 5.00; In 5.67 Analysis: [- S- - EC]; (H)(T)A; [-(EM)-/-EM-/(DE]; TRL 4; HRL 2; CML 4; Reasoned effects.		

SESSION 4: AUTONOMOUS CAPABILITIES THAT SUPPORT TEAMING		
No.	AUTHORS	TITLE
10	Murphy	Agile design of an automated team mate for video intelligence exploration
Description: Imagery analysis AI/ML integration concepts for ISR/PED FMV exploitation; managed teaming interaction concept development; model, prototype operator-AI interaction work structure; laboratory based, pilot experimental test of plausible AI capabilities; operators observe FMV describing reportable events; test effects of AI teaming options, variable relevance AI "call outs" notifications, display prototypes; compare novice/expert operator-AI interaction; report mixed +/- qualitative performative effects; experts under-trust, novices over-trust AI. Assessment: TQ 5.67; OR 6.33; Ex 5.83; TC 5.33; SQ 5.67; In 5.17. Analysis: [RS – DEC; H(T)A; [R----/(R)---/)(D)E]; TRL 3; HRL 5; CML 4; Performed effects.		
11	Luchero, Huynh, Lange, Moreau.	Modelling provenance of decisions within the human-autonomy team
Description: HAT generic, information provenance, explainable autonomy technology optimization proof-of-concept, data model prototype; HAT interaction tracking, modelling, analytics; mitigates opaque AI decision models; data dependencies, responsibility work flow anomalies detection; standard generic, domain agnostic Web data model; domain interoperability; benefits user knowledge, memory unburdening; algorithmic transparency, improved SA, decision explanation; disaster response use case; proposed C2 UxV base protection demonstration; relevant operational simulation test environment; IMPACT play-calling planner, COMPACT policy management, rhetoric structure automated action narrative system. Assessment: TQ 5.97; OR 6.00; Ex 5.17; TC 6.00; SQ 6.33; In 6.33 Analysis: [RSG – EC]; (H)TA; [(REM)-/REM-/D(E)]; TRL 2; HRL 1; CML 5; Reasoned effects.		
12	Nirenburg, McShane	Two types of autonomous, language endowed learning agents for human autonomy teaming.
Description: HAT generic, natural language, speech understanding & generation dialogue communication; defines model prototype, simulated language endowed agent; computational cognitive architecture; ontological semantics knowledge base; domain, learning, context, intentions, roles, responsibilities. Application use cases: clinical medicine virtual patients simulation training; human robot team manufacturing. Assessment: TQ 3.63; OR 2.67; Ex 2.67; TC 3.00; SQ 5.00; In 4.83. Analysis: [- S - - - -; (H)TA; (R-M)-/R-M-/D-]; TRL 3; HRL 1; CML 3		

SESSION 5: HAT INTERACTION AND DESIGN (PART1)		
NO.	AUTHORS	TITLE
13	Schmitt, Brand, Rudnick, Schulte.	Experimental evaluation of a cooperative automation approach for manned-unmanned teaming in future military helicopter missions
Description: Multi air manned and UAV platform (MUM-T) coordination and guidance autonomy technology; crew adaptive mission management autonomy technology sub-system; model and prototype operational user and agent interaction experimental test; relevant air-land integration simulation operational user test environment; subjective and objective performative effects. Assessment: TQ 6.00; OR 6.20; Ex 6.20; TC 6.20; SQ 6.20; In 5.20. Analysis: [RSGDEC; HTA; REME/REME/DE]; TRL 6; HRL 5; CML 6; Performed effects.		
14	Diggelen, Barnhoorn, Peeters, Staal, Stolk, Vecht, Waa, Schragen	Pluggable social artificial intelligence for enabling human-agent teaming
Description: C2 UAS collaboration autonomy management technology, interface sub-system, independent additive layer, task and social, component technology; behavioural library model prototype demonstration; structured knowledge ontology; team structure interdependency interpredictabilities; soft directability work agreements; affective risk sensitivity; progressive teaming development; proactive time critical communication semantic anchors language; common ground, shared SA, anticipation, explanation AI functions; HAT training; application scenarios, use cases; relevant base protection surveillance swarm simulation, proof-of-concept demonstration; potential operational user test environment. Assessment: TQ 5.48; OR 5.40; Ex 5.60; TC 5.40; SQ 5.40; In 5.60. Analysis: [- SG – EC]; (H)TA; [(RE)- -/RE - - /D(E)]; TRL 6; HRL 5; CML 4; Reasoned effects.		
15	Thomas, Cottrell, Grabham.	Policy management and negotiation: Enabling effective human autonomy teaming
Description: C2 UAS effectiveness assurance autonomy, sub-system working agreement technology; policy management, negotiation model, prototype, development, user-agent interaction design patterns, operational user interface experimental testing; relevant multi-mission, layered C2, air-land integration simulation and operational live flight trial test environments; full HAT spectrum, subjective, objective performative effects. Assessment: TQ 5.80; OR 6.33; Ex 6.33; TC 6.00; SQ 5.33; In 5.00. Analysis: [RSGDEC]; HTA; [REME/REME/DE]; TRL 7; HRL 5; CML 4; Performed effects.		

PANEL: HAT INSTITUTIONAL INTEGRATION		
No.	AUTHORS	TITLE
17	Koch	Ethical challenges of lethal autonomous weapons systems and manned-unmanned teaming.
Description: Teaming moral philosophy advantage, human ruling, deciding imperatives; ethics impasses of autonomous military weapons systems technology, lethal applications; imperative reflective requirements; indispensible ethical considerations, normative, pre-conditions, assumptions, contingent actual empirical causal structure, non-contingent argumentations; responsibility gap, violation of human dignity; International Humanitarian Law, Human Rights standards; lethal force, combat risks; imponderable risks, precautionary principle; human involvement, risk reduction; partnership inequality, human ethical responsibility, human leading & controlling; skill fade, disempowerment; preservation of decision making capacity, human comprehension, understanding, reasoning. Assessment: TQ 3.68; OR 5.50; Ex 3.50; TC 2.50; SQ 2.00; In 4.00. Analysis: [R – G - - -]; - - -; [- - - -/- - - -/ - -]; TRL 1; HRL 1; CML 0.		
18	Earthy, Downs	Human-system interaction standards for robotic, intelligent, autonomous systems
Description: Review ISO TC159/SC4, Ergonomics of Human System Interactions, Technical Report ISO 9241-810 Human –system issues of robotic, intelligent and autonomous systems (RIAS); priorities for future standards work; three-dimensions research analysis framework: scale of system/criticality/stage in lifecycle; classification categories: RIAS Designed characteristics effects on individual humans, Human-RIAS Interaction, RIAS to RIAS, RIAS-Organisational, Social /Cultural, Emergent Societal; Hazards if ergonomics are not addressed; applicable standards; changes to develop RIAS ergonomics; HAT symposium papers; RIAS issues categories. Assessment: TQ 4.67; OR 4.67; Ex 6.33; TC 3.33; SQ 4.67; In 4.33. Analysis: [- - GD - -]; - - -; [- - - - /- - - -/ - -]; TRL 2; HRL 3; CML 3; Reasoned effects.		
20	Kersholt, Barnhoorn, Hueting, Schuilenborg	Automation as an intelligent teammate: Social psychological implications
Literature review, HAT outcome affecting factors; Input-Mediator-Output team effectiveness framework; cognitive, behavioural, affective mediators; taskwork, human appraisal, interaction outcomes; task performance, SA; relational, human appraisal neglected; agency increases social relational effects, task interactions; Human attentional control, spatial ability, gaming experience, self-efficacy; Robot reliability, effectiveness, personality, appearance; impacts trust; future collaborative decision making to impact social behaviours. Assessment: TQ 4.27; OR 4.00; Ex 4.00; TC 3.67; SQ 4.67; In 5.00. Analysis: [- S - - E -]; HT(A);[- - M-/- -(M) –/D(E)]; TRL 1; HRL 2; CML 3; Reasoned effects.		

SESSION 6: HAT INTERACTION AND DESIGN (PART 2)		
No.	AUTHORS	TITLE
21	Chen, Stowers, Wohleber, Barnes.	Agent transparency and human-autonomy teaming effectiveness in multi-robot management contexts
Description: C2 UxV autonomy technology collaboration management; plan play-calling; agent transparency, trust; user interface transparency SA level variable: current goal, actions, plans; reasoning process; projection outcome predictions; +/- uncertainty; variable communication critical, complimentary style; controlled experimental laboratory test environment; military mission collaboration plans (search patterns); map display route/text/acceptability options; variable aid accuracy; student participants bias; accept/reject effectiveness; subjective workload, trust; transparency effects accuracy, utility; style, trust interaction effects. Assessment: TQ 4.60; OR 5.33; Ex 4.33; TC 5.00; SQ 5.00; In 3.33. Analysis: [- SG – EC]; HT(A); [REM - /R(EM)-/D(E)]; TRL 4; HRL 5; CML 4; Performed effects.		
22	Butcher	Trust me I'm artificial intelligence
Description: Plan for investigation of social, legal, ethical factors of future AI systems use; assurance, training, education policy; outline intent for on-line questionnaire survey; MoD military/civilian target audience (Advanced Command Staff Course); aim to measure decision making tendencies (maximise, satisfice, minimise), attitudes, influences; beliefs, intentions, behaviours; effects of prior beliefs, attitudes, social & organisational norms, control level for trust, on AI system initial trust, decision confidence; comparison of MoD & public perception of opportunities, risks of AI ML; hypothetical Defence recruitment AI decision aid scenario proposed to identify automation trust challenge, utility, why, when; intended qualitative data thematic analysis. Assessment: TQ 3.33; OR 3.67; Ex 3.67; TC 2.67; SQ 4.00; In 2.67. Analysis: [R - - - - -]; - - -; [- - - -/- - - -/ - -]; TRL 1; HRL 1; CML 0		
23	Miller	Transparency over time: The importance of pre- and post-mission interactions in human autonomy teaming
Description: Basic principles analysis; autonomy information display negates workload savings; formulation of adaptive displaced transparency concept; displace display time to lower workload periods before/after execution; pre-mission planning, post mission explanation & de-briefing; off-line adaptive mental model synchronisation or reconciliation, SA comprehension & trust tuning; experimental paradigm; benefit predictions for SA, inferencing, teammate behaviour prediction, tempo, workload, communication bandwidth. Assessment: TQ 5.70; OR 5.50; Ex 6.00; TC 5.00; SQ 5.50; In 6.50. Analysis: [- S - - EC]; HT(A); [REM-/R(EM)-/D(E)]; TRL 2; HRL 2; CML 6; Reasoned effects.		

KEYNOTES		
No.	AUTHORS	TITLE
KN1	Amison	Human machine teaming: UK MoD Joint Concepts Note 1/18
Description: Presentation only. No documentation. Comprehensive, informed review of the military requirement. JCN 1/18 Synopsis: Human mental capacity is increasingly unable to cope with the data deluge involved in conflict, while computer algorithms are challenged by uncertainty and ambiguity in data and decision-making. Optimising human and machine capabilities in teams that maximise strengths and mitigate weaknesses is essential. Risk is assessed within context, and will remain a human responsibility. Mission command will change in an AI age and will demand variable autonomy in remote and automated systems. High-quality live and synthetic collective training and above all experimentation with AI systems will be essential for us to learn how to optimise our ability to create effective human-machine teams. The increasing array of capabilities of robotic and AI systems will be limited by not only what can be done, but also by what actors trust their machines to do. The MoD must continue to be proactive in considering legal, ethical and public concerns surrounding the use of robotics and AI.		
KN2	Hancock	Islands of Autonomy: Are human and machine autonomy a zero sum?
Description: Presentation only. No documentation. Imaginative, entertaining, erudite, academic keynote. Related indicative significant prior: The headlong rush to automate continues apace. The dominant question still remains whether we can automate, not whether we should automate. The suggestion offered is that unlimited automation of all technical functions will eventually prove anathema to the fundamental quality of human life. Arguably, certain tasks, pursuits and past-times should potentially be excused from the automation imperative. This deliberation leads us back to the question of balance in the cooperation, coordination and potential conflict between humans and the machine they create. (from: Hancock,2013. Automation: How much is too much. Ergonomics, 57(3), 449-454. September 2013. London, Taylor and Francis.)		
HFM	Flemisch	Human autonomy teaming: A STO HFM Perspective
Description: Presentation outline only. No documentation. Bringing HAT into reality - Open questions: Operational: What are requirements, hopes and concerns? Autonomous capabilities that support teaming? Function division and relationships/roles (who does what functions, and in what kind of relationship?). Human Machine interaction, communication & cooperation? Command and Control aspects of HAT? Ethical, Legal and Social aspects of HAT? Assessing risks and benefits of HAT?		

POSTERS		
No.	AUTHORS	TITLE
P1	Coronado, Lange.	Autonomic plan monitoring
Description: C2 UxV autonomy technology management system application formulation; autonomous agents monitoring, adjusting plans; monitoring tools, supervision detail, large vehicle numbers; IMPACT plan monitoring managed by Rainbow autonomics framework computing technique; manages high level goal-oriented plays translated by AI agents into plans; Rainbow evaluation, adaptation tools effecting plans & execution systems: Architecture Evaluator, Adaptation Manager, Strategy Executor, Model Manager; probes sense C2 environment; gauges build & update models; plans represented as networks; enables continuous plan quality evaluation via gauges; provides effectors automated adaptation strategies; notifications, executable actions (call play, restriction re-route); adaptations for automatic enactment or communication to operators; IMPACT/COMPACT/MAPLE AIM integration. Analysis: [RSG - -E-]; (H)- A ; TRL 2; HRL 2; CML 3; Recognised effects.		
P2	Coronado, Lucero, Lange.	Discretizing and managing the task environment
Description: C2 UxV autonomy technology task management sub-system application formulation; maintaining human engagement; discretizing task management environment, humans & machine agent understanding; enables more efficient HAT; human-in-the-loop/on-the-loop principles; autonomous agent for mission critical tasks, avoiding overload & repetition & trivia; abstract command conversion to machine & human-readable task & goal completion; digitizing task environment; modeling agent performance; tracking urgency, priority, situation changes; balancing workload; human authority visibility, control working agreements; interface design; operator task performance execution tools. IMPACT/AIM integration. Analysis: [RSG - E -]; (H)- A ; TRL 4; HRL 2; CML 3; Reasoned effects.		
P3	Zelnio, Fendley.	Automated aiding of image analysis as a function of sensor type
Description: Automated imagery analyst aiding decision support system DSS; concept formulation; experimental laboratory testing; prototype imagery analysis aiding variables; evaluation of simulated hypothetical automated decision support system/DSS aiding design; image interpretation target classification aiding; cognitive bias task structuring mitigation; variable DSS levels: none, classification recommendation, prioritization, confidence , reference images; variable imaging sensor types: electro-optical (EO), infrared (IR), and synthetic aperture radar (SAR); college student participants; significant DSS and sensor variable performative effects : DSS level increased operator target & type classification accuracy, interpretation time, fixation time eye tracking; DSS reference imagery increased confidence & trust; Accuracy: EO>IR,>SAR; Time: SAR>IR>EO; Heuristics use, response bias analysis, SAR anchoring, EO/IR confirmation; sensor literalness effects: EO requires less aiding than more non-literal IR, SAR sensors. Analysis: [RS - - EC]; H - A ; TRL 4; HRL 5; CML 4; Performed effects.		

A2. SUMMARY OF EVALUATION OF NATO SCO HFM SY300 HAT CONTRIBUTIONS

ASSESSMENT	**ANALYSIS**			
TQ: Technical Quality OR: Operational Relevance Ex: Exploitation TC: Technical Capability SQ: Scientific Quality In: Innovation TRL: Technology Readiness Level HRL: Human Readiness Level CML: Capability Maturity Level	RSGDEC R: Requirement Drivers S: CSE Solutions G: Functional Governance D: Development Enablers E: Performative Effects C: Enabled Capabilities	HTA H: Human T: Teaming A: Autonomy	REME/REME/DE R: Recognition E: Evaluation M: Mitigation E: Execution /D: Dialogue E: Effects REME/ : Human /REME : Autonomy	Moderators ( ) : Weak - : Not applicable

1. Rohling	**Command management** - Supporting Infantry Units with UGV and UAS						TG 247
ASSESSMENT	OR 5.80	Ex 5.60	TC 5.40	SQ 4.00	In 4.00	TQ 5.00	Performed effects
ANALYSIS	[RSGDEC]	[HTA]	REME/REME/DE	TRL 7	HRL5	CML 4	
2. Roberts, Soper.	**HFI Organisational risks** - Identifying HF considerations associated with command, control, communication of autonomous vehicles in the maritime domain						
ASSESSMENT	OR 4.80	Ex 4.40	TC 3.20	SQ 4.20	In 2.20	TQ 3.76	
ANALYSIS	[- - GD - C]	[- - -]	[- - - -/- - - -/ - -]	TRL 1	HRL2	CML 0	
3. Keirl, Thorpe, Taylor, Grabham.	**Effectiveness Metrics** - Measuring effectiveness of HAT						TG 247
ASSESSMENT	OR 5.80	Ex 5.60	TC 5.40	SQ 4.40	In 4.40	TQ 5.12	Performed effects
ANALYSIS	[RSGDEC]	[HTA]	REME/REME/DE	TRL 7	HRL5	CML 3	
4 Johnson, Vignati, Duran.	**Interdependency Analysis** - Understanding human autonomy teaming through interdependence analysis						
ASSESSMENT	OR 4.80	Ex 5.00	TC 5.00	SQ 4.80	In 4.40	TQ 4.80	Recognised effects
ANALYSIS	[- S - - E -]	[HT(A)]	[REME/(REME)/D-]	TRL 3	HRL3	CML 4	
5. Frost, Bartik, Calhoun, Spriggs, Ruff, Behymer.	**Functional Division** - Creating a well situated human autonomy team: The effects of team structure						TG 247
ASSESSMENT	OR 6.00	Ex 5.80	TC 6.40	SQ 5.80	In 5.20	TQ 5.84	Performed effects
ANALYSIS	[RSGDEC]	[HT(A)]	[REME/(REME)/DE]	TRL 6	HRL 5	CML 3	
6. Shively, Coppin, Lachter.	**Communication dialogue** - Bi-directional communication in human-autonomy teaming						TG 247
ASSESSMENT	OR 4.75	Ex 4.25	TC 5.00	SQ 4.25	In 4.00	TQ 4.45	Recognised effects
ANALYSIS	[- S - - E -]	[HTA]	[REME/- - - - /D-]	TRL 2	HRL1	CML 6	

7. Baltzer, Lopez, Flemisch.	**Collision Avoidance** - Human autonomy teaming using cooperative automation, interaction patterns and image schemes						TG 247
ASSESSMENT	OR 4.50	Ex 4.75	TC 4.08	SQ 5.25	In 4.08	TQ 4.53	Performed effects
ANALYSIS	[- S – DE -]	[HTA]	[REME/REME/DE]	TRL 6	HRL 5	CML 4	
8. Neerincx, Diggelen, Stolk	**Patterns, work agreements** - Design patterns and ontologies for situated human agent collaboration, organisation, team and interaction level						TG 247
ASSESSMENT	OR 4.67	Ex 5.00	TC 5.50	SQ 5.17	In 4.50	TQ 4.97	Recognised effects
ANALYSIS	[- SGDEC]	[HT(A)]	[-EME/(REME)/D-]	TRL 2	HRL 2	CML 4	
9. Kelley, Ouimet.	**Search planning** - On human assisted multiagent-based planning:						
ASSESSMENT	OR 5.33	Ex 5.33	TC 6.33	SQ 5.00	In 5.67	TQ 5.62	Reasoned effects
ANALYSIS	[RSGDEC]	[HTA]	[REME/REME/DE]	TRL	HRL 2	CML	
10. Murphy	**Imagery Analysis** - Agile design of an automated team mate for video intelligence exploration.						
ASSESSMENT	OR 6.33	Ex 5.83	TC 5.33	SQ 5.67	In 5.17	TQ 5.67	Performed effects
ANALYSIS	[RS – DEC]	[H(T)A]	[R----/(R)---/)(D)E]	TRL 3	HRL 5	CML 4	
11. Luchero, Huynh, Lange, Moreau.	**Information provenance** - Modelling provenance of decisions within the human-autonomy team						
ASSESSMENT	OR 6.00	Ex 5.17	TC 6.00	SQ 6.33	In 6.33	TQ 5.97	Reasoned effects
ANALYSIS	[RSG – EC]	([H)TA]	[(REM)-/REM/D(E)]	TRL 2	HRL 1	CML 5	
13. Schmitt, Brand, Rudnick, Schulte.	**Manned-Unmanned Teaming** - Experimental evaluation of a cooperative automation approach for manned-unmanned teaming in future military helicopter missions						
ASSESSMENT	OR 6.20	Ex 6.20	TC 6.20	SQ 6.20	In 5.20	TQ 6.00	Performed effects
ANALYSIS	[RSGDEC]	[HTA]	[REME/REME/DE]	TRL 6	HRL 5	CML 6	

14. Diggelen, Barnhoorn, Peeters, Staal, Stolk, Vecht, Waa, Schragen	**Social Intelligence** - Pluggable social artificial intelligence for enabling human-agent teaming						
ASSESSMENT	OR 5.40	Ex 5.60	TC 5.40	SQ 5.40	In 5.60	TQ 5.48	Reasoned effects
ANALYSIS	[- SG – EC]	[(H)TA]	[(RE)- -/RE - - /D(E)]	TRL 6	HRL 5	CML 4	
15. Thomas, Cottrell, Grabham.	**Policy management** - Policy management and negotiation: Enabling effective human autonomy teaming						
ASSESSMENT	OR 6.33	Ex 6.33	TC 6.00	SQ 5.33	In 5.00	TQ 5.80	Performed effects
ANALYSIS	[RSGDEC]	[HTA]	[REME/REME/DE]	TRL 7	HRL 5	CML 4	
17. Koch	**Ethical, legal arguments** - Ethical challenges of lethal autonomous weapons systems and manned-unmanned teaming.						
ASSESSMENT	OR 5.50	Ex 3.50	TC 2.50	SQ 2.00	In 4.00	TQ 3.68	
ANALYSIS	[R – G - - -]	[- - -]	[- - - -/- - - -/ - -]	TRL 1	HRL 1	CML 0	
18. Earthy, Downs	**Ergonomics standards** - Human-system interaction standards for robotic, intelligent, autonomous systems						
ASSESSMENT	OR 4.67	Ex 6.33	TC 3.33	SQ 4.67	In 4.33	TQ 4.67	Reasoned effects
ANALYSIS	[- - GD - -]	[- - -]	[- - - - /- - - -/ - -]	TRL 2	HRL3	CML 3	
20. Kersholt, Barnhoorn, Hueting, Schuilenborg	**Social Relationships** - Automation as an intelligent teammate: Social psychological implications						
ASSESSMENT	OR 4.00	Ex 4.00	TC 3.67	SQ 4.67	In 5.00	TQ 4.27	Reasoned effects
ANALYSIS	[- S - - E -]	[HT(A)]	[- - M-/- -(M)-/D(E)];	TRL 1	HRL 2	CML 3	

21. Chen, Stowers, Wohleber, Barnes.	**Transparency trust** - Agent transparency and human-autonomy teaming effectiveness in multi-robot management contexts						
ASSESSMENT	OR 5.33	Ex 4.33	TC 5.00	SQ 5.00	In 3.33	TQ 4.60	Performed effects
ANALYSIS	[- SG – EC]	[HT(A)]	[REM - /R(EM)-/D(E)]	TRL 4	HRL 5	CML 4	
22. Butcher	**Trust attitudes** - Trust me I'm artificial intelligence						
ASSESSMENT	OR 3.67	Ex 3.67	TC 2.67	SQ 4.00	In 2.67	TQ 3.33	
ANALYSIS	[R - - - - -]	[- - -]	[- - - - / - - - - / - -]	TRL 1	HRL 1	CML 0	
23. Miller	**Displaced transparency** - Transparency over time: The importance of pre- and post-mission interactions in human autonomy teaming						
ASSESSMENT	OR 5.50	Ex 6.00	TC 5.00	SQ 5.50	In 6.50	TQ 5.70	Reasoned effects
ANALYSIS	[- S - - EC];	[HT(A)]	[REM-/R(EM)-/D(E)]	TRL 2	HRL 2	CML 6	
P1. Coronado, Lange.	**Autonomics framework** - Autonomic plan monitoring Analysis: [RSG - -E-]; (H)- A ; TRL 2; HRL 2; CML 3; Recognised effects.						
ASSESSMENT							Recognised effects
ANALYSIS	[RSG - -E-]	[(H)- A]	[- - (ME)/- -M(E)/- (E)]	TRL 2	HRL2	CML 3	
P2. Coronado, Lucero, Lange.	**Task management** - Discretizing and managing the task environment						
ASSESSMENT							Reasoned effects
ANALYSIS	[RSG - E -]	[(H)- A]	[(R) - - (E)/R - -E /-(E)]	TRL 4	HRL 2	CML 3	
P3. Zelnio, Fendley.	**Imagery analysis** - Automated aiding of image analysis as a function of sensor type						
ASSESSMENT							Performed effects
ANALYSIS	[RS - - EC]	[H- A]	[R - - - /(R)-(M) - /- (E)]	TRL 4	HRL 5	CML 4	

Identifying Human Factors Considerations Associated with Command, Control and Communication (C3) of Autonomous Vehicles in the Maritime domain

Dr. M. H Roberts and Dr. C. S. Soper

QinetiQ
Portsdown Technology Park
Southwick Road
Cosham
Portsmouth
PO6 3RU
UNITED KINGDOM

mhroberts@qinetiq.com cssoper@qinetiq.com

ABSTRACT

As part of an ongoing program of research looking at autonomous vehicles in the maritime domain (i.e., Maritime Autonomous PLatform Exploitation (MAPLE) – Phase 4), this paper describes an approach to identifying human factors considerations associated with Command, Control and Communication (C3) of unmanned vehicles (UxV). This approach utilises the Human Factors Integration domains as a framework for conducting an Early Human Factors Analysis.

This paper describes the data capture process, as well as the context of the source / baseline data and the Early Human Factors Analysis (EHFA) workshop from which the main findings are derived. Additionally, this work provides the rationale for the EHFA tool selection, as well as the Human Factors Integration (HFI) domains which acted as a framework for the data collection. Following this, all Risks, Assumptions, Issues, Dependencies and Opportunities (RAIDO) are reported and summarised and a number of key recommendations and conclusions are provided, which will inform future work within the MAPLE programme, as well as future in-service C3 systems for autonomous vehicles.

The main findings are presented in terms of their impact per HFI domain. Specifically, this work identifies Human Factors considerations associated with 'Social and Organisational' (particularly for legislative and regulatory frameworks), as well as 'Manpower', 'Training' and 'System Safety' HFI domains. A key finding is that capability developers will need to overcome cultural resistance and inaccurate / inappropriate expectations before Unmanned Vehicles (UxVs) are fully deployed. Furthermore, there will be considerable effort involved in identifying training requirements, and incorporating new training technologies / methodologies, whilst not increasing the training burden (a key MoD requirement).

Additionally this paper will describe the next steps in the ongoing MAPLE research program. Specifically an upcoming series of experiments, which will further inform and guide human machine tasking / role allocation and elucidate the HF considerations associated with UxV C3.

1.0 INTRODUCTION

The MAPLE (Maritime Autonomous PLatform Exploitation) programme is focussed on the deployment and operation of Unmanned Vehicles (UxVs) in the maritime domain, in terms of a circa 2030 timeframe. The MAPLE programme is being considered in advance of established doctrine or regulatory requirements,

however ultimately, UxV deployment will be governed by the legislative / regulative frameworks developed specifically for their use. On completion of MAPLE 4, Defence Science and Technology Laboratory (Dstl) will have a roadmap describing a recommended approach for bringing UxVs into service, and integrating them within the Command Chain of a Royal Navy (RN) platform. This work contributes to the RNs understanding of the UxV Command, Control and Communication (C3) space and builds on research themes, which have developed throughout previous phases of the MAPLE programme.

1.1 Scope of analysis

The scope of this work covers the C3 associated with UxV management and control. Specifically, this work uses an Early Human Factors Analysis (EHFA) approach (described fully in section 2.2) to capture the Human Factors (HF) considerations (i.e., risks, issues, assumptions, dependencies and opportunities) associated with the C3 of UxVs. Notably, UxV operation and physical control (i.e., manual control of UxVs) are outside the scope of this current work.

1.2 Context (previous work in MAPLE programme)

The following subsections describe the research and analysis that has informed and guided this current work.

1.2.1 Autonomy Strategy and Roadmap

The Autonomy Strategy and Roadmap [1], developed under MAPLE 4, concentrates on identifying some of the more significant considerations that may assist (enable) or obstruct (constrain) the introduction of the capability implied by MAPLE. It also identifies the Dstl and MoD programmes that provide the exploitation and deployment opportunities to start to realise the capability envisaged. Future iterations of this work will further develop and refine the strategies and interactions between the various activities to form a roadmap for the deployment of autonomous capabilities. Relevant HF related material was extracted from [1] as an input to the materials used for the EHFA stakeholder workshop.

1.2.2 Functional decomposition

In order to model and capture the tasks, subtasks and enablers anticipated for UxV C3, MAPLE 4 (HF activity), extended the HF Elements of the Persistent Architecture [5] and thereby produced a functional de-composition and data flow depiction. Initially this effort was informed by work undertaken in MAPLE 2.5 [2], and further shaped by data collected during Unmanned Warrior 2016 (UW 16) [3].

During UW 16, observers were tasked with recording observed tasks / subtasks, equipment used to enable task and level of autonomy associated with each task (drawn from the MAPLE 2 and 2.5 workshops). The nature of the UxV Command and Control (C2), observed during UW 16, provided a unique opportunity for comprehensive data capture in relation to UxV activities undertaken in the current timeframe.

The MAPLE 4 work focussed on those areas that were outstanding from the MAPLE 2.5 work, as well as capturing any final changes required to the dataflow diagrams and functional de-compositions [4]. The revised depictions are provided for both realistic (feasible and desirable) and high autonomy in the circa 2030 timescale. These are being used in MAPLE 4 as the baseline for developing the experimental investigations such as those associated with level of autonomy, multiple assets and adaptable autonomy.

1.2.3 Wider HF issues

The wider HF issues that were incorporated in the Critical Architecture Review (CAR) at the end of MAPLE 2.5 included a number of slides describing the wider HF issues associated with truly autonomous systems

[5]. This material was also used as an input to develop the materials for the EHFA stakeholder workshop.

2.0 METHOD

This section describes the approach taken to data collection and analysis, as well as the rationale for selecting these methods. Additionally this section describes the context of the analysis and the Human Factors Integration (HFI) domains to which this current work applies.

2.1 Description of EHFA

EHFA is a key HFI tool, and is an essential part of initial project development activities. The main purpose of an EHFA is to identify high-level HF considerations, which are then classified by probability and impact, linked to an action plan where appropriate, and recorded in a Human Factors Issues Register (i.e., RAIDO log). The RAIDO log is intended to be updated iteratively as the programme develops in maturity, and thereby acts as an audit trail for HF governance.

The four explicit stages of the EHFA process, adapted from generic risk management protocols, are as follows:

- Identify – find, list and characterise HFI considerations;
- Assess – analyse and prioritise HFI considerations against agreed criteria;
- Plan – develop, analyse and recommend response actions to address HFI considerations; and
- Implement – implement, monitor, report and review response actions against objectives. Escalate risks where required.

The level of detail and content of an EHFA will depend on the level of information available (typically determined by the maturity of the project), as well as who performs the analysis and the end customer for whom the analysis is intended. As described in Joint Services Publication (JSP 912), the convening of a workshop is an effective way of conducting the EHFA together with an appropriately mature HFI baseline dataset acting as a preliminary input. Also, the EHFA process requires early engagement with relevant stakeholders, which greatly improves the quality and accuracy of the human-centric risks and potential inputs to the RAIDO. Given the level of maturity of this current work programme, which is at a research stage, a high level EHFA was conducted in accordance with the mandated EHFA methodology guide and toolset [6], but was appropriately tailored for the MAPLE programme. The wider HF issues identified under MAPLE 2 and relevant work within MAPLE 4 also provided an input to this work (see sections 2.3.2 – 2.3.4 for further detail).

The use of EHFA is recommended for the purpose of ensuring HF involvement at the earliest possible point in the Concept, Assessment, Demonstration, Manufacture, In-Service and Disposal (CADMID) lifecycle, and is used to identify, quantify and manage HF considerations, which influence safety, human performance and / or cost. Additionally, EHFA provides a framework through which to capture, monitor and review programme risks (as well mitigation strategies), thus ensuring that HF risks and issues are considered iteratively throughout the CADMID lifecycle.

2.2 HFI Domains

The EHFA approach sits within the broader context of HFI, and as such is bounded by the same HF domains. The HFI domains utilised as a discussion framework in this current work are as follows:

- Manpower – the number of military and civilian personnel required and potentially available to

operate, maintain, sustain and provide training for [C3 of unmanned systems];

- Personnel – the cognitive and physical capabilities required to be able to train for, operate, maintain and sustain [C3 of unmanned systems];
- Training – the instruction or education, and on-the-job or unit training required to provide personnel their essential job skills, knowledge, values and attributes;
- HF engineering – the integration of human characteristics into system definition, design, development, and evaluation to optimise human-machine performance under operational conditions;
- Health hazards – short or long term hazards to health occurring as a result of normal [C3 of unmanned systems];
- System safety – safety risks occurring when the system is functioning in either a normal or abnormal manner;
- Social and organisational – organisational practices (including doctrine), the culture associated with information sharing, centralised / distributed teamwork, information management and communication.

2.3 RAIDO Definitions

The definitions of the different elements that constitute a RAIDO are detailed below:

- An issue is a significant, unplanned event or situation that has already occurred, or will definitely happen;
- A risk is a significant uncertain event or situation;
- An assumption is a statement that is taken as being true for the purposes of planning or other decision making;
- A dependency defines the relationship between products or activities;
- An opportunity is, in effect, a risk with potentially beneficial consequences.

2.4 Scoring Criteria

The following subsections provide a breakdown of the descriptors, for probability of occurrence and potential impact, as well as breakdown of the scoring criteria and coding scheme.

2.4.1 Descriptors for likelihood of occurrence and impact

Figure 1 provides the descriptive criteria used in the EHFA tool to assess likelihood of risk / issue, as well as the impact following occurrence. Whilst the examples given in Figure 1 will not be relevant to all design / manufacturing programmes, they are provided here for reference purposes.

Likelihood of risk/issue occurring	Confidence in information on which cause is based (example guidance)
High	No evidence of risk/issue being addressed in previous system(s) or analysis/prototyping performed to understand or assess impact. New unproven technology, task or procedure is required.
Medium	Technology, task or procedure is similar to that used for previous system(s), but no evidence of analysis/prototyping performed for this system.
Low	Some evidence of risk/issue being considered in previous fielded system(s) or prior analysis/prototyping performed to demonstrate performance in an operational scenario.

Potential Impact	Description of impact in relation to HFI (example guidance)
High	Failure to resolve the issue or fully mitigate the risk will prevent the system achieving specified performance, safety requirements, usability and/or maintainability/ supportability requirements.
Medium	Failure to resolve the issue or fully mitigate the risk will prevent the system achieving specified performance, usability, safety requirements, and/or maintainability/ supportability requirements, but can be managed within the performance margin for the system.
Low	Failure to resolve the issue or fully mitigate the risk will impact system performance, usability, safety and/or maintainability/supportability requirements, but is contained within the performance margin for the system.

Figure 1: Descriptive criteria for probability of occurrence (top panel) and impact on programme, if risk / issue / opportunity occurs (bottom panel)

2.4.2 Breakdown of scoring criteria

Figure 2 shows each Probability + Impact (P+I) score, as well as its associated colour code, and a descriptor of the scoring criteria for each P+I score. As can be seen in the descriptors provided in Figure 2, issues are scored differently to risks and opportunities, in order that the certain probability of issues can be appropriately weighted. Additionally the lowest and highest scores are given nominal P+I scores (1 and 9 respectively) in order that a continuous (i.e. scalar) distribution is provided.

P+I Score	Descriptor
9	Only given to issues which have a high impact (i.e., for issues which are given an impact rating of high, a P+I score of 9 is assigned)
6	Either a risk / opportunity with high probability (P = 3) and high impact (I = 3), or an issue with medium impact (i.e., for issues which are given a medium impact rating (I = 2), a probability rating of 4 is assigned)

5	This P+I score can be achieved in several ways. First, either a risk / opportunity with high probability (P = 3) and medium impact (I = 2), or a risk / opportunity with medium probability (P = 2) and high impact (I = 3). Additionally, a P+I score of 5 is also given to an issue with a low impact rating (I = 1; P = 4)
4	A risk / opportunity with medium probability (P = 2) and medium impact (I = 2)
3	Either a risk / opportunity with medium probability (P = 2) and low impact (I = 1), or a risk / opportunity with low probability (P = 1) and medium impact (I = 2)
1	A risk / opportunity with low probability and low impact is given a nominal rating of 1

Figure 2: Overview of P+I scoring criteria, together with colour coding scheme

2.4.3 EHFA workshop

The majority of the content contained in the RAIDO log described in this work was derived from discussions with SMEs during an EHFA themed / HFI guided workshop. The format of which, as well as emergent HF considerations, are described in this subsection.

The overarching aim of the MAPLE 4 EHFA workshop was to highlight the key HF related considerations that will need to be addressed to enable UxV capabilities/technology to be introduced into service without increasing manning / training overheads. Additionally the output from the EHFA workshop informed development of the issues to be addressed within the MAPLE 4 HF experimental programme. A suitable agenda was agreed with the customer, which included top level discussions covering the HFI domains and their effect on the introduction of autonomous vehicle capabilities, assumptions and notable considerations. Discussions were recorded for the purposes of populating the EHFA report and RAIDO log and stakeholders were invited to provide further comments / suggestions and validate the workshop outputs.

The main output of this workshop was an endorsed and validated set of HF considerations, organised by HFI domain. The validation process took the form of recorded workshop comments and subsequent feedback, being incorporated into the HFI domains, and thereafter disseminated for further validation and final endorsement.

Attendees were a representative panel of domain experts and military advisors, which included SMEs from Dstl as well as senior members of the RN. All pertinent comments were transcribed and subsequently sent to attendees for validation and endorsement. All considerations and descriptors contained in this work are either direct quotes or very close paraphrasing of validated and endorsed commentary.

3.0 FINDINGS AND RESULTS

This section describes the main findings which emerged from the EHFA workshop.

3.1 EHFA Findings

The following subsections contain information pertaining to key risks, assumptions, issues, dependencies and opportunities, as well as a summary of the distribution of these HF considerations across the seven HFI domains (see Tables 2 – 8). Additionally, a summary table is provided which includes a Probability + Impact score for the key risks, issues and opportunities (see Table 8).

3.2 Key Programme Risks, Issues, and Opportunities

The following sub-section describes risks, issues and opportunities associated with the C3 of UXVs. The following subsections (3.4 – 3.6) groups the HF considerations thematically and provides an overview and descriptor for each risk, issue and opportunity.

3.2.1 Summary of issues

This sub section reports all of the issues captured during the EHFA workshop. According to [6], "an issue is a significant, unplanned event or situation that has already occurred, or will definitely happen, which is certain to affect at least one project or programme activity, or business objective. As such the probability of an issue occurring is scored as 'Certain' [achieving a score of 6 in the EHFA tool]".

3.2.2.1 Training considerations

There are a group of issues pertaining to the understanding of the training requirements for UxV C3 (associated with the 'Personnel' and 'Training' HFI domain), these issues are described in Table 1. Three of these issues (I-0001 – I-0003) were given the highest P+I score (certain probability + highest impact), and as such are a high priority consideration for the MAPLE programme.

Table 1: Summary of issues associated with training and personnel, together with P+I score

Issues (ID and title)		Descriptor	P+I
I-0001	No Target Audience Description (TAD)	"Currently it is unknown who (which roles and at what seniority) will be responsible for UxV C3, therefore there is no Target Audience Description (TAD), and a lack of understanding regarding tasks/roles, changes to competencies, and what Knowledge Skills and Attitudes will be required"	9
I-0002	Training requirement	"Bespoke training will be required for UxV C3. Entirely new operator tasks, skills and competencies will need to be understood. Planning and delivering this training requirement (from training needs analysis, to designing course content and establishing appropriate standards) will be a high priority"	9

I-0003	Training requirement (weaponisation)	As per I-0002 except with specific focus on issues associated with C3 of weaponised UxVs.	9
I-0004	Maintainer's KSAs	"Maintenance crews will need to be equipped with skills and knowledge to support C3 for UxV platforms and payloads"	5

3.2.2.2 Frameworks and protocols

There are also a group of issues related to the adoption of a common set of standards for UxV design (particularly for communication and networking protocols), associated with the 'HF Engineering', 'System Safety' and 'Social and Organisational' HFI domains. These issues are shown in Table 2, and as can be seen this group of issues also attracted the highest P+I score throughout (certain probability + highest impact), and similarly are a high priority consideration for the MAPLE programme.

Table 2 All issues associated with frameworks and protocols, together with P+I score

Issues (ID and title)		Descriptor	P+I
I-0005	Open architectures	"It will be crucial for system designers to adopt a modular approach, using open architectures (i.e. using principles of service oriented architectures) and standards (including for communication and networks)"	9
I-0006	Emergency recovery protocols	"Emergency recovery protocols must ensure that unmanned systems can be stopped and recovered at any time during operations"	9
I-0007	Legislative and regulative frameworks	"The development of legislative and regulative frameworks will be a high priority"	9

3.2.3 Summary of risks

This sub section reports all of the risks captured by the EHFA workshop. According to [6], "A risk is an uncertain event for which the probability of occurrence should be assessed".

3.2.3.1 Manning and training methods

A key group of risks, which emerged from the EHFA workshop, pertain to both manning levels and training methodologies (associated with the 'Manpower', 'Training' and 'Social and Organisational' HFI domains) are shown in Table 3. Only one of these risks attracted the highest P+I score (R-0001), however R-0017 (inappropriate expectation) was rated as high probability + medium impact (P+I = 5). Consequently, both of these risks are a priority for mitigation planning.

Table 3 All issues associated with frameworks and protocols, together with P+I score

Risks (ID and title)		Descriptor	P+I
R-0001	Current manning levels	"There is a risk that current manning levels may not support change to tasking"	6
R-0017	Inappropriate expectation	"There is a risk of inappropriate expectations that using UxV systems will reduce manning levels"	5
R-0002	Government directive to reduce manning levels	"There is a risk that UK government will issue a(nother) directive to reduce manpower, in which case manning levels will be insufficient"	4
R-0004	Training technologies	"Emerging training technologies (e.g., VR / AR / AI) may have a negative impact (e.g. through disconnection and disorientation) on the development of operator's knowledge, skills and attitudes (KSA)"	4

3.2.3.2 Human-machine interaction

There are a number of risks associated with human-machine interaction and the human machine interface (associated with the 'HF Engineering', 'Health Hazards' and 'System Safety' HFI domains), listed below in Table 4. Only one of these risks attracted the highest probability + impact score (R-0008), however a group of three risks (R-0005, R-0010 and R-0011) were rated as either high probability + medium impact or medium probability + high impact (P+I = 5). Consequently, these four risks are priorities for appropriate mitigation.

Table 4 All risks associated with human machine interaction, together P+I score

Risks (ID and title)		Descriptor	P+I
R-0008	Operator Overload	"There is a risk that managing large numbers of unmanned systems, particularly during novel scenarios, will result in overload"	6
R-0005	HMI development	"As unmanned systems are developed a corresponding development of HMIs may also be necessary (e.g., incorporating AR / VR / AI), with design features to enhance trust"	5
R-0010	Reduced situation awareness	"There is a risk of reduced SA, as well as impaired decision making capabilities, due to increased automation"	5

R-0011	Task disengagement	"There is a risk that long vigilance / monitoring periods will result in task disengagement"	5
R-0006	Trust	"There is a risk that trust / confidence in automated systems will be low, impacting deployment and efficacy"	3
R-0007	Over confidence	"There is a risk of over confidence in UxV systems, thereby impacting on mission delivery and safety"	3

3.2.3.3 Safety considerations

Table 5 illustrates two risks reported pertaining to the system safety HFI domain. One of these risks (R-0012) attracted the second highest P+I score, medium probability + high impact (P+I = 5), the other (R-0013) attracted the third highest P+I score, medium probability + medium impact (P+I = 4). R-0012 will be a priority for appropriate mitigation.

Table 5 All risks associated with system safety, together with P+I score

Risks (ID and title)		Descriptor	P+I
R-0009	Safety checks	"Safety critical factors (including weaponisation and use of sensors for targeting) will affect the time required to perform safety checks of the C3 system"	5
R-0012	Information accuracy	"There is a risk that the UxV provides incorrect information, driving inappropriate / incorrect response from C3 operators"	5
R-0013	Maintenance	"If UxV C3 systems are classified as safety critical then additional / higher levels of maintenance will be required"	4

3.2.3.4 Organisational structures

The final group of risks pertains to the potential impact on organisational structures as shown in Table 6. Two of these risks (R-0014 and R-0016) attracted the second highest P+I score, medium probability + high impact (P+I = 5), the other two (R-0015 and R-0018) attracted the third highest P+I score, medium probability + medium impact (P+I = 4). Clearly then R-0014 and R-0016 will be priorities for appropriate mitigation, as will R-0015 and R-0018 although with a slightly lower priority.

Table 6 All risks associated with system safety, together with P+I score

Risks (ID and title)		Descriptor	P+I
R-0014	Platform organisational structures	"The introduction of unmanned systems will impact on platform organisational structures"	5
R-0016	Cultural resistance	"There is a risk of considerable cultural resistance to UxV uptake"	5
R-0015	RN organisational structures	"There is a risk that the introduction of unmanned systems will impact on RN organisational structures / career pyramids"	4
R-0018	Job satisfaction and motivation	"There is a risk that using UxV systems will negatively impact job satisfaction and motivation"	4

3.2.4 Opportunities

This sub section reports all of the opportunities captured by the EHFA workshop. According to [6], "An opportunity is, in effect, a risk with the possibility of positive consequences. The process for managing opportunities is very similar to that for managing risks and issues".

As can be seen in Table 7, there are two opportunities in the training domain (O-0002 and O-0003), an opportunity to address the manning levels (O-0001), and a related opportunity to introduce operator monitoring protocols (O-0004) to improve effectiveness of team / role allocation management. Additionally there is an opportunity to improve operational procedures (O-0005) and job satisfaction (O-0008), as well as continue with autonomous vehicle research (O-0006), particularly for multi-platform coalitions (O-0007). The majority of these opportunities were rated as medium probability + medium impact (P+I = 4), however the highest rated opportunity (autonomous vehicle research O-0006) was rated as high probability + medium impact (P+I = 5), and consequently ought to be the primary focus for future business.

Table 7 All reported opportunities, together with P+I score

Opportunities (ID and title)		Descriptor	P+I
O-0006	Autonomy research	"The autonomous vehicles space is well funded and is a priority for UK defence spending"	5
O-0002	Emerging training technologies	"The potential of emerging training technologies (e.g. VR/AR/AI) to enhance understanding of a systems plan/decision/action should be harnessed and used to help develop Operator trust/acceptance of new technology"	4

O-0003	On-board training	"There is an opportunity to use emerging technologies (e.g., VR/AR) to conduct training on-board platforms in order to keep pace with rapid technological development"	4
O-0004	Operator monitoring	"Human psychophysiological monitoring technologies can be embedded into UxV control hardware, and used to monitor operator engagement/task loading/awareness and reaction time and optimise C3 human machine teaming"	4
O-0005	Operational procedures	"Incorporating UxV C3 capability will lead to changes to operational procedures, which will enable every ship to have access to all deployed assets / capabilities"	4
O-0007	Multi-platform coalitions	"There is an opportunity to improve targeting and sensor performance, through multi-platforms coalitions"	4
O-0008	Job satisfaction and motivation	"There is an opportunity that to use UxV systems to positively impact job satisfaction and motivation"	4
O-0001	Manning levels	"Additional study will be required in order to address overall manning issues (this should be included in the roadmap)"	1

3.2.5 Summary of key risks, issues and opportunities

Shown below in Table 8 are the highest rated risks, issues and opportunities, ranked according to P+I scores.

Table 8 Top seven rated risks, issues and opportunities. Ranked according to P+I score, with colour coding shown

Rank	Risks			Issues			Opportunities		
1st	R-0001	Current manning levels	6	I-0001	No Target Audience Description (TAD)	9	O-0006	Autonomy research	5
2nd	R-0008	Operator overload	6	I-0002	Training requirement	9	O-0002	Emerging training technologies	4

3rd	R-0009	Safety checks	6	I-0003	Training requirement (weaponisation)	9	O-0003	On-board training	4
4th	R-0002	Government directive to reduce future manning levels	5	I-0005	Open architectures	9	O-0004	Operator monitoring	4
5th	R-0005	HMI development	5	I-0006	Emergency recovery protocols	9	O-0005	Operational procedures	4
6th	R-0010	Reduced situation awareness	5	I-0007	Legislative and regulative frameworks	9	O-0007	Multi-platform coalitions	4
7th	R-0011	Task disengagement	5	I-0004	Maintainer's KSAs	5	O-0008	Job satisfaction and motivation	4

3.2.6 Assumptions and dependencies

The following sub-section describes assumptions and dependencies associated with the C3 of UXVs. Assumptions and dependencies are not given a P+I score, consequently no ratings or rankings are provided.

3.2.6.1 Assumptions

This sub section reports all of the assumptions captured by the EHFA workshop. According to [6], an assumption is “a consideration that is set as true to enable a project to proceed”. Table 9 shows all of the assumptions which will likely act as constraints on UxV capability development. As can be seen in Table 9, there are a number of training / manning assumptions (A-0001, A-0002, A-0004, A-0008), as well as group of assumptions pertaining to recruitment (A-0005), career structures / staff roles (A-0006 and A-0007) and technical / regulatory based assumptions (A-0009 – A-0013).

Table 9 All reported programmatic assumptions

Assumptions (ID and title)		Descriptor
A-0001	Current manning levels	There will be no increase in overall manning
A-0002	Manning levels (holistic)	Manning levels will be considered holistically (i.e. adding in new roles or changing the responsibilities of existing roles will impact multiple interconnected functions)

A-0003	C3 hub in operations room	In 2030 timeframe the Operations room will still be the C3 hub
A-0004	Skills and competencies	Skills and competencies will be developed to support UxV tasking
A-0005	Recruitment strategy	Recruitment strategy / levels will either change or will be made sufficient to support UxV operations
A-0006	Career pyramids	Career pyramids and structures will remain unchanged / unaffected by new capability
A-0007	Staff roles	Staff roles which currently cover C3 decision making tasks will remain the same. However operator level roles and responsibilities will change
A-0008	Training burden	It is assumed that it will be possible to broaden expertise of existing personnel without increasing training pipeline burden
A-0009	C3 HCI	It is assumed that interface for the C3 of UxVs, will conform to appropriate HF standards and guidance materials
A-0010	Safety case	A robust safety case is required that covers human-centred issues, risks and constraints
A-0011	Failure modes	User reactions to failure modes need to be considered during the design process
A-0012	Rules of Engagement (ROE)	ROE will need to keep pace with developing UXV system capabilities
A-0013	Regulation and legislation	RN will utilise UxVs in advance of regulation and legislation catching up with UxV operations

3.2.6.1 Dependencies

This sub section reports all of the dependencies captured by the EHFA workshop. According to [6], a dependency is "a consideration that must be met to enable a project's delivery. Dependencies must be identified and tracked as they will impact on the projects delivery". Table 10 shows all of the dependencies which will likely act as constraints on UxV capability development. As can be seen in Table 10, the dependencies captured during the EHFA workshop are distributed across the HFI domains (with the exception of the 'System Safety' domain, for which no dependencies were recorded). For examples, 'Manning' and 'Personnel' (D-0001 and D-0002 respectively), 'Training' (D-0003), 'HF Engineering' (D-0004), 'Health Hazards' (D-0005), and 'Social / Organisational' (D-0006 and D-0007).

Table 10 All reported programmatic dependencies

Assumptions (ID and title)		Descriptor
D-0001	Manning levels are dependent	Before manpower requirements can be specified it will be necessary to specify the level of automation of individual UxVs (i.e., minimum / mandatory levels of autonomy, legislative boundaries of autonomous decision making, maximum number of autonomous assets under command)
D-0002	Competencies and Knowledge, Skills and Attitudes (KSAs)	Personnel will need to be kept aware of the impact that the C3 unmanned systems may have on future roles, competencies and KSAs
D-0003	Training Audience Description (TAD)	Training requirements must be considered sufficiently early to ensure training pipeline needs are met. A Target Audience Description (TAD) is needed to support MAPLE, and this will enable a Training Needs Analysis to be developed
D-0004	Evaluation and assurance	There is a dependency on appropriately specified test, evaluation and assurance procedures, including those related to the human element
D-0005	Task loading	Task loading associated with UxV C3 will need to be understood in order to put appropriate mitigations in place to avoid overloading
D-0006	Tactical doctrine	New tactical doctrine will need to be produced to incorporate unmanned systems
D-0007	Expectation management	There will be a need to build trust with RN stakeholders, by providing education (to all levels of RN structure) on what autonomy does and does not provide

3.2.5 HFI considerations totals per HFI domain

For summary purposes, Figure 3 shows the distribution of HFI considerations per HFI domain. As can be seen in Figure 3, all considerations are summed together and percentage totals per HFI domain are shown. Whilst the HF considerations are (for the most part) evenly distributed, the largest proportion of HFI considerations was reported for the ‘Social and Organisational’ domain (25% of total HFI considerations). ‘Manpower’, ‘Training’ and ‘System Safety’ each accounted for ~15% of the total, whilst ‘Personnel’ accounted for 12%. The least frequently occurring HFI consideration pertained to ‘Health Hazards’, accounting for only 4% of the total.

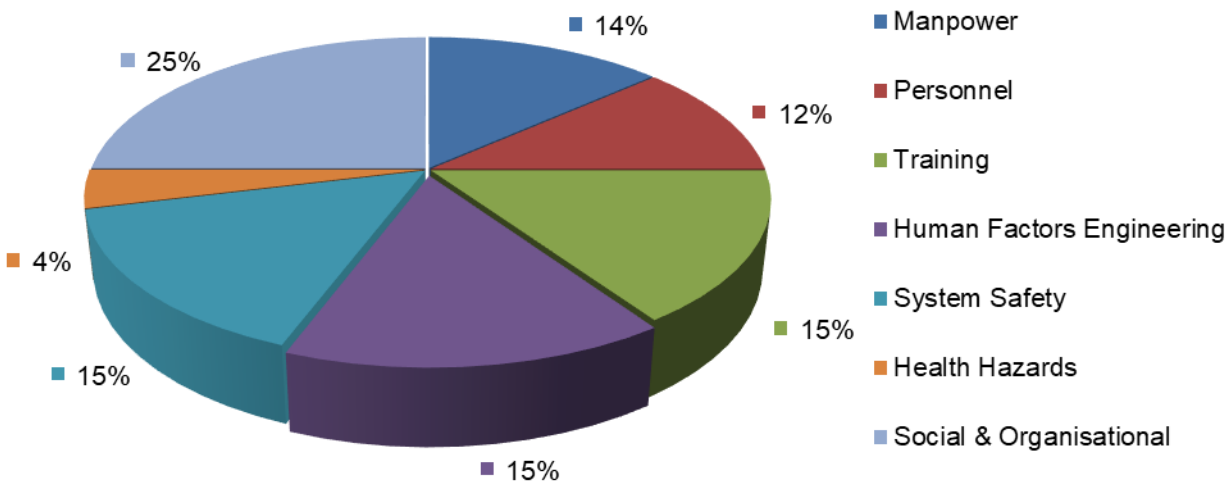


Figure 3 Distribution of HF considerations across HFI domains

4.0 Conclusions and recommendations

4.1 Conclusions

The main conclusion of this current work is that in order to achieve successful integration of UxVs into the maritime domain there will be a considerable amount of 'Social and Organisational' change needed. Essentially, this will include changes to legislative and regulative frameworks, platform organisational structures and RN organisational structures. Additionally, capability developers will need to overcome cultural resistance and inaccurate / inappropriate expectations (e.g., the expectation that UxVs will reduce manning levels, or at least not increase them).

In addition, the other key conclusion which emerges from this current work, and which aligns with previous findings, is there will be considerable effort involved in identifying training requirements, and incorporating new training technologies / methodologies, whilst not increasing the training burden. This may also qualify as an 'inappropriate expectation', particularly if (as seems likely) additional recruitment will be needed to support this new capability. A Target Audience Description (TAD) is needed to support MAPLE, and this will enable a Training Needs Analysis to be developed.

4.2 Recommendations

Taken together this mix of issues, risks, assumption and dependencies, associated with training and manning levels, as well as legislative / regulatory frameworks, indicates that further work is required to fully understand these HFI considerations and appropriately specify mitigation where required. This requirement for further research is captured by opportunity O-0006, which posits that: "the autonomous vehicles space is well funded and is a priority for UK defence spending".

It is recommended that RAIDO logs (i.e. risk registers) are considered live documents, and should therefore

be re-visited and updated with respect to any new considerations as they emerge throughout the CADMID lifecyle. For the MAPLE programme the RAIDO log will be updated again after a series of experiments investigating levels of automation and control of multiple assets in a synthetic environment. Once updated, the output of this work will provide a high level reference point for the procurement and integration of UxVs in the Maritime domain.

REFERENCES

[1] Autonomy Strategy and Roadmap MarCE/UxVC2/MP/438 QINETIQ/17/02661v2 October 2017

[2] MAPLE (UXV C2 TDP) Phase 2.5. Human Factors: Operator Autonomy levels for Functional Blocks, Task 7.40, Technical Report, Issue 1.0 MarCE/UxVC2/TS/379 November 2016

[3] Autonomous Control Exploitation and Realisation: Human Factors Quick Look Analysis, QINETIQ/16/04445 December 2016

[4] Human Factors: Extend the Human Factors Elements of the Persistent Architecture v2.0, QINTETIQ/17/02662, September 2017

[5] Persistent Architecture Phase 2 CAR pack, MarCE/UxVC2/TN/346, April 2016

[6] Early Human Factors Analysis Tool User Guide issue 2.1, Ministry of Defence, 2016

Measuring Effectiveness of Human Autonomy Teaming

Helen Lashley, Aaron Thorpe, Robert Taylor and Antony Grabham
Defence Science and Technology Laboratory
Dstl Portsdown West, Fareham, PO17 6AD
UNITED KINGDOM

hjkeirl@dstl.gov.uk

ABSTRACT

The aim of this work is to test and evaluate the effectiveness of HAT, specifically the interactions and dialogue between a human operator and autonomy system in joint, collaborative manned-unmanned operations.

Throughout the comprehensive UK MoD Dstl programme of R&D on Autonomy and Mission Systems, a number of different metrics have been developed predominantly focused on exploiting the critical decision methodology to address the different constituent parts of the complete HAT System.

The Human task work and teamwork elements of the HAT system, through CAPTEAM which is designed to estimate mission efficiency by the metrication of Reward and Effort associated with critical mission events and decision processes. The Autonomy Task Work component of the HAT system is addressed through a broken down multi-dimensional Trustworthiness scale. The entirety of the HAT system can be assessed through a combination of the HAT Capability Maturity Model and the Risk Assessment. The new component of the UK HAT System assessment methodologies is the REMEDE assessment protocol. The protocol itself is focused on the Human Autonomy Teamwork component of the entire system, and has been developed from a variety of information processing, communication and team work models.

At this stage, the REMEDE protocol has been selected for use with the STRATUS project under the UK MoD Dstl Autonomy Research programme, with a specific instantiation developed for the specifics of the trial. This trial will serve as the initial opportunity to undertake verification and validation testing of the REMEDE protocol, the outcomes of which will allow a greater understanding of the discrimination and sensitivity of the data captured.

1.0 INTRODUCTION

Human-Autonomy Teaming (HAT) - the creation of a mutually supportive, co-operative, collaborative partnership between the human user and advanced "intelligent" automation technology - has been recognised as the necessary strategic, human-system integration design objective for Unmanned Systems (UxS). HAT is aimed at delivering efficiently and effectively the required agile, adaptive, context sensitive, multi-mission capability, whilst enabling meaningful and effective Command and Control (C2).

Prior to the advent of Artificial Intelligence (AI) and before the introduction of decision making technology, the human commander/operator retained sole responsibility for providing cognitive decision making capability. The commander/operator uses sensing, thinking, values, reasoning, learning, knowledge and memory, coupled with judgement, vision and imagination, shared and developed through social communication and dialogue. Cognitive capability enables human judgement to be reasonable, dependable and reliable in dealing with uncertainties and ambiguity. But it also enables humans to be innovative and creative, thereby anticipating options, contingencies, risks, opportunities and threats. This creative cognitive

capability provides the essential context sensitivity, flexibility, agility and adaptability needed for decision superiority in the complex, dynamic military environment.

Concurrently, the provision of effective test and evaluation methodology for proving HAT efficacy, with verification and validation of human-system performance and effectiveness, has been identified as presenting a significant technical challenge as described by R Taylor [1], [2].

The aim of this work is to describe the test and evaluation methodologies used in determining the effectiveness of HAT, specifically the interactions and dialogue between a human operator and autonomy system in joint, collaborative manned-unmanned operations.

2.0 MILITARY CONTEXT (RATIONALE)

Remote and automated systems employing unmanned and remotely operated platforms, and systems employing embedded automation and self-governing autonomous functionality, are being increasingly employed as enablers of military capability. However, it is widely recognised that in the complex, dynamic and uncertain environment of military operations, human effectiveness and HAT is needed, as an axiomatic design imperative, for autonomous systems to be assuredly safe and effective.

Test methods are needed that are designed to achieve the levels of proof required to support evidence-based decision making in UK MoD Defence Equipment procurement. Selection of test methods involves balancing requirements for innovation and demonstration (e.g. use case testing), compared with the need for robust performance-based trade-space analysis, valuing discrimination power and optimisation (e.g. multi-variate performance testing). Innovation test methods can be less rigorous experimental design with changes to technology, adaptations to Tactics, Techniques, Plans and Procedures (TTPs) and learning encouraged throughout the duration of a trial. The data and outcomes are used as indicators for future research and lessons identified rather than statistically significant results. Discrimination test methods are more robust and rigorous with dependent and independent variables. The trials outputs are more defined and lessons and conclusions can be drawn directly from the data. Figure 1 illustrates the distinction between innovation and discrimination test methods, (R Taylor 2015 [1]). A balanced approach using both test methods has been employed by UK MoD for HAT research, resulting in quantitative and qualitative data as well as validated assessment methodologies for testing HAT effectiveness, as described in this paper.

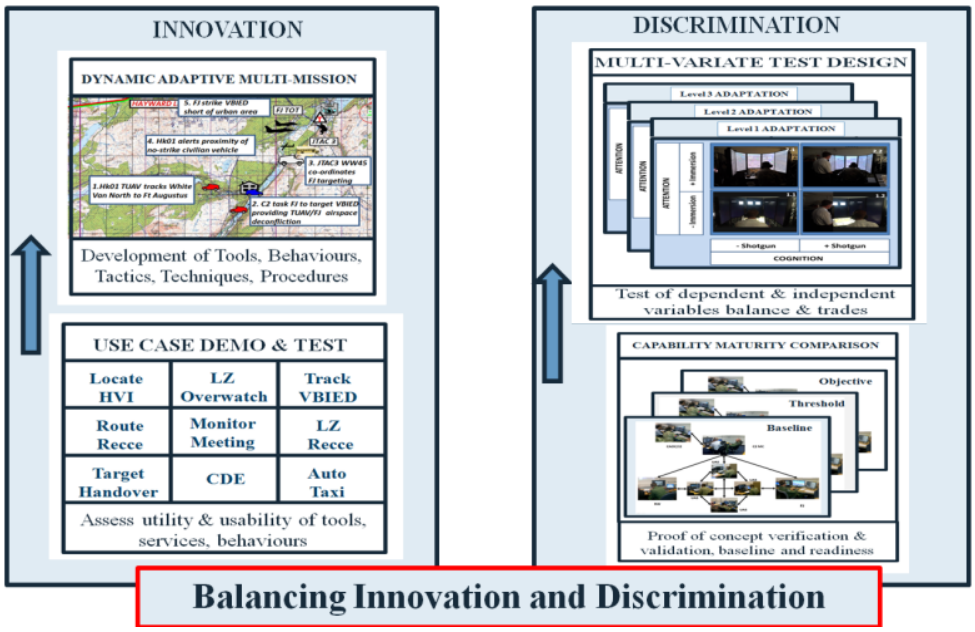


Figure 1: Applicable Test Methods

2.1 Critical Decision Methodology

Underpinning the test methodology, and providing the required human-effectiveness centric focus, Dstl has developed a cognitive capability testing approach, based on dynamic Mission Critical Decision Making (MCDM) methodology. MCDM focuses the locus of testing on Course of Action (CoA) adjustment responses to dynamic mission critical events. Uncertainties in the military combat environment affect understanding of risks and opportunities associated with delivering the command intent and in the performance of associated tasks, with dynamic targets and threats, resulting in adjustments and changes to CoA tactics, techniques, plans, procedures, all affecting the delivery of effects. (Taylor and Grabham, 2012 [3])

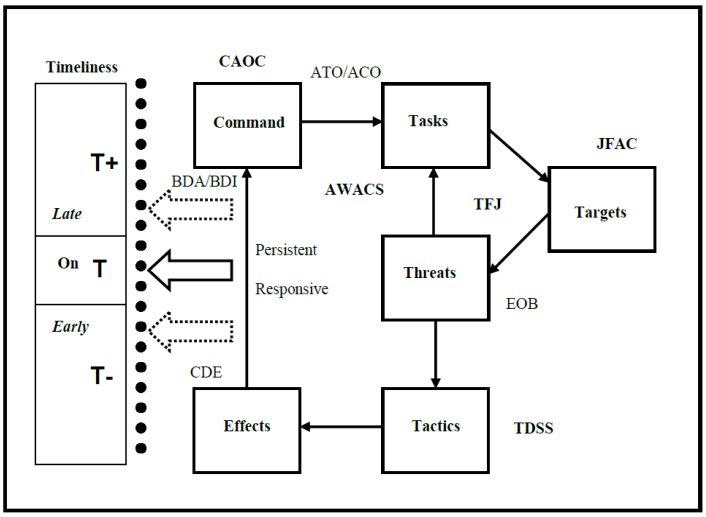


Figure 2: Model of Dynamic Air Mission Management

MCDM employs objective, task analysis and effects-based Measure of Effectiveness and Measures of Performance (MoE/MoP), coupled with subjective decision-based metrics for estimation of human and mission effectiveness (Collaborative Adaptability Proficiency Test Evaluation Assessment Methodology (CAPTEAM)). Essentially, the aim is to capture the agility (speed) and adaptability (quality) of mission commander/operator decision-making in dynamic mission timeline context as illustrated in Figure 3.

Subsequently, under the UK Autonomy and Mission Systems Research Programme, the MCDM methodology was further developed to measure collaborative adaptability proficiency with autonomous systems concepts and technologies. This has included investigating C2-MM decision making by varying and controlling demands for levels of conflict resolution, and by observing collaborative mitigation/adaptation responses, considered in terms of coordination (de-conflicting acts), cooperation (de-conflicting means) and collaboration (de-conflicting goals).

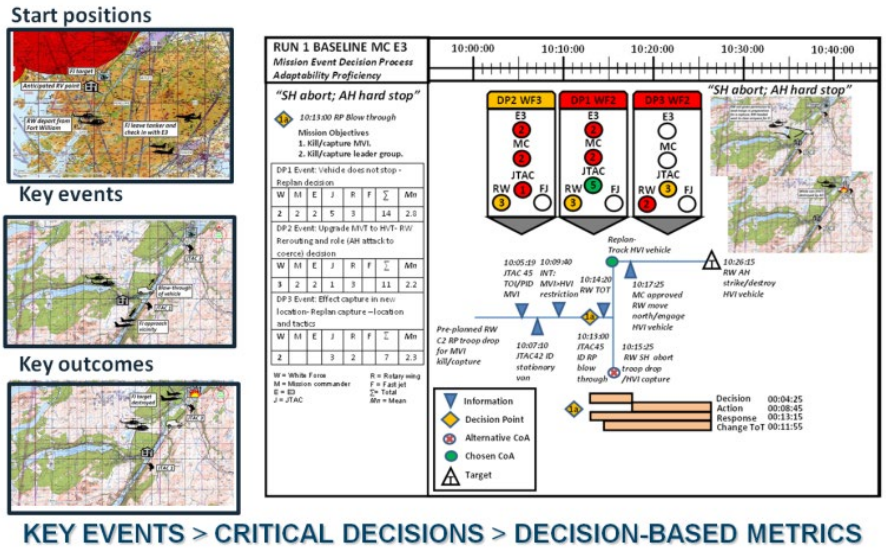


Figure 3: Dstl MCDM Methodology

When designing experiments to measure the effectiveness of HAT in a military context it is important to have militarily representative scenarios which require military operators or autonomous systems to make realistic decisions. Key events must be designed into the scenario to ensure operators are making critical decisions, with effects, in a layered C2 environment, to support our decision based assessment approach.

3.0 HAT METRICS

3.1 Introduction

The technical challenge of proving HAT efficacy, with verification and validation of human-system performance and effectiveness can be broken down in to a number constituent components, the first being the task work performed by a human operator, and the interactions that a human has with other human members of the team. To further add to the HAT system, it is necessary to consider the task work that is undertaken by the autonomy, and the teaming that occurs between autonomous components of the wider team. The final component is the human autonomy teamwork. The combination of the components is illustrated in Figure 4.

Throughout the comprehensive UK MoD Dstl programme of R&D on Autonomy and Mission Systems, a

number of different metrics have been developed predominantly focused on exploiting the critical decision methodology to address the different constituent parts of the complete HAT System. These components are described in the following sections, with the human autonomy teamwork metrics being described in detail.

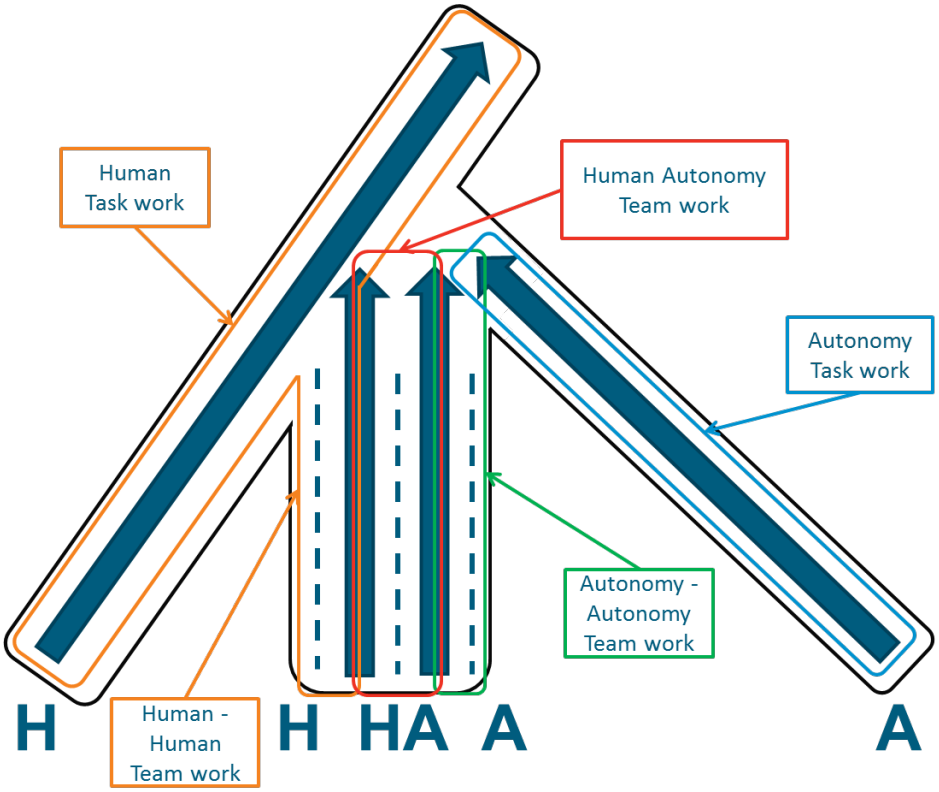


Figure 4: Dstl HAT Component Breakdown

3.2 Human System Teaming

Collaborative Adaptability Proficiency Test Evaluation Assessment Methodology (CAPTEAM) is an independent assessment provided by Dstl, which is designed to estimate mission efficiency by the metrication of Reward and Effort associated with critical mission events and decision processes [3].

The metrics factors measured in CAPTEAM are Workload, Re-plan Task Load, Situational Awareness, Decision Quality, Teamwork, Performance, Tools and Technical System. The metrics are split into Human Task work and Human-Human Teamwork, Task work is assessed in the Participant Protocol and Teamwork in the Subject Matter Expert Observer Protocol using Likert 7-point, Low-High, anchored rating scales.

LOW ANCHORS	DIMENSION Low 1	2	3	4	5	6	High 7	HIGH ANCHORS
WORKLOAD								
Slow/Leisurely/No time pressure	**WL Time Pressure**							*Rushed/Rapid/Frantic*
Very little/Extremely easy	**WL Mental Effort**							*Extremely hard/Stretched*
Relaxed/Placid/Tranquil/Calm	**WL Stress**							*Anxious/Worried/Up tight/Harassed*
REPLAN TASK LOAD								
Forgiving/Simple/Easy	**Replan Decision** Recognise–Evaluate-Mitigate							*Difficult/Demanding/Complex/Exacting*
Forgiving/Simple/Easy	**Replan Action** Disseminate–Acknowledge–Execute-Report							*Difficult/Demanding/Complex/Exacting*
SITUATION AWARENESS								
Stable/Simple/Fixed	**SA Demand on Attentional Resources**							*Unstable/Complex/Variable*
Relaxed/Inattentive/Spare Capacity/Singular	**SA Supply of Attentional Resources**							*Concentrated/Full capacity/Divided*
Uninformative/Meaningless/Unfamiliar	**SA Understanding**							*Informative/Meaningful/Familiar*
DECISION QUALITY								
Indecisive/Doubtful/Guess	**DQ Confidence**							*Decisive /Confident /Obvious*
Dangerous/Vulnerable/Risky	**DQ Survivability**							*Safe/Unthreatening/Secure*
Ineffective/Un-useful/Unproductive	**DQ Effectiveness**							*Effective/Capable/Useful*
Behind/Late/Pressurised	**DQ Timeliness**							*Ahead/On-time/In control*
PERFORMANCE								
Unsatisfactory/Unacceptable/Failure	**Task Performance**							*Perfect/Successful/Satisfied*
Useless/Complicated/Difficult	**Tools Utility**							*Easy/Intuitive/Effective*
No backup plans/ Insensitive/Unresponsive	**Adaptability Proficiency**							*Sensitive/Responsive/Timely effective backup plans*
Fails to meet any mission objectives/0%	**Probability of Mission Success**							*100%/Successfully achieved all mission objectives*

LOW ANCHORS	DIMENSION Low 1	2	3	4	5	6	High 7	HIGH ANCHORS
DECISION QUALITY								
Dangerous/Vulnerable/Risky	**DQ Survivability**							*Safe/Unthreatening/Secure*
Ineffective/Useless/Unproductive	**DQ Effectiveness**							*Effective/Capable/Useful*
Behind/Late/Pressurised	**DQ Timeliness**							*Ahead/On-time/In control*
TEAMWORK								
Silent/Unclear/Confused/Slow/Uninformative	**Communication**							*Clear/Timely/Coherent/Concise/Rapid/ Informative*
Uninformative/Meaningless/Unfamiliar	**Shared SA**							*Informative/Meaningful/Familiar*
Indecisive/Unassertive/Confused/ Uncommunicative	**Leadership**							*Commanding/Directing/Guiding/Initiating management*
Ignoring/Hindering/Blocking/Criticising/ Conflicting	**Support**							*Monitoring/Advising/Correcting/Assisting*
Light/Quiet/Unpressured	**Team Workload**							*Heavy/Busy/Pressured*
PERFORMANCE								
Unsatisfactory/Unacceptable/Failure	**Task Performance**							*Perfect/Successful/Satisfied*
Separate/Independent/Divided/Conflicting	**Collaboration**							*Joint/Shared/Coordinated/Cooperating*
Powerless/Weak/Irrelevant/Not Involved	**Influence Power**							*Influential/Powerful/Forceful/Decisive/Critical*
No backup plans/ Insensitive/Unresponsive	**Adaptability Proficiency**							*Sensitive/Responsive/Timely effective backup plans*
Fails to achieve any mission objective/0%	**Probability of Mission Success**							*100%/ All mission objectives successfully achieved*

Figure 5: Dstl 2011 CAPTEAM Participant Taskwork and Observer Teamwork Assessment Protocols

This assessment methodology relies on military operators to make critical decisions in response to dynamic mission events, with measurable effects, within a realistic military scenario in either a live or synthetic environment. Note that this early 2011 Dstl CAPTEAM protocol was designed for assessing human performance in UK MoD research on Dynamic Air Mission Management (DAMM). Here, the maturity of the automation and autonomy of the technical system was not yet ready to be demonstrated or evaluated as an explicit, discrete capability. In Dstl DAMM research, the focus was on evaluating human effectiveness with distributed networked, C2, at operational and tactical mission levels, with exploitation of human teaming collaboration capability supported by networked mission enabling technologies. The DAMM technical system was treated as being designed to provide tools for human use, in the sense of user-centred mission enabling technologies and operator decision support, with assessment focused on measures of tool usability and utility. Tools usability and utility assessment techniques contrast with the challenges of assessing the efficacy for human use of automatic or intelligent agent technical systems. Such systems are potentially capable of non-deterministic, fully autonomous functioning of operations (sensing, decision making, behaviour), completely independent of human control, with properties such as agency, self-

determination and autonomy associated with advanced Artificial Intelligence (AI), Big Data processing and Machine Learning (ML) technologies.

The Dstl CAPTEAM assessment protocol was used for operator assessments in the 3rd US-UK STRIKE WARRIOR III Joint SE Trial, July 2011 (Cottrell 2011 [4]). Statistical analysis by Dstl of data obtained from application of CAPTEAM indicated the extent to which CAPTEAM component metrics associated with Effort, or Reward are statistically correlated with Probability of Mission Success (PMS) (Taylor and Grabham 2012 [3]). The results showed that relatively beneficial Reward factors, such as improved SA and Decision Quality, were significantly highly correlated with PMS. In comparison, Effort factors and sub-components, such as Workload Stress and Mental Effort a exhibited relatively low, non-significant correlations with PMS. The inter-correlations of the Reward and Effort factors for both Task Work and Team Work in relation to PMS are shown in Figure 6. This illustration uses a structural equation model technique (Castor 2009 [5]). This technique seeks to provide a meaningful, evidence-based, visual data representation of the relationships between the assessment components. The structure of the CAPTEAM components is based on analysis of the data principal components and inter-correlations with PMS The positioning of the sub-components is arranged to be indicative of the progressive strength of the relationships with the intended high level effects. In this graphical representation, improved Adaptability Proficiency and raised PMS are shown to be the outputs, products, or positive effects, arising from the pluri-potential benefits of successful Collaboration in Task Work with Team Work.

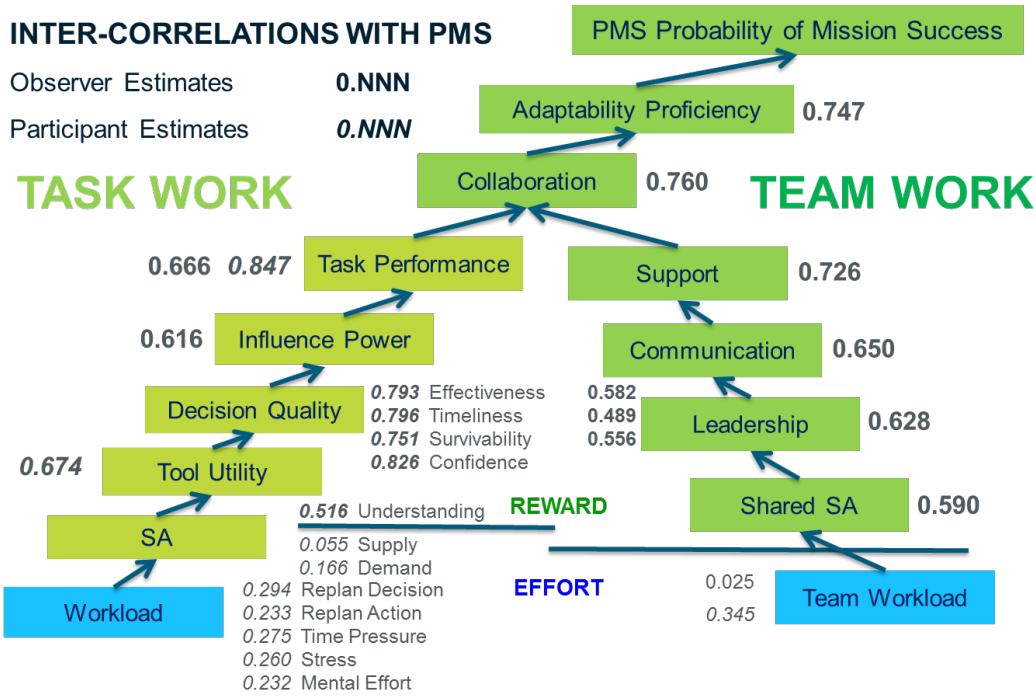


Figure 6: Inter-correlation of the Task work and Team work, Reward and Effort CAPTEAM metrics

An additional benefit from re-using MCDM/CAPTEAM methodology for Autonomous Systems Research T&E consistently over time has been to enable a set of recognised benchmarks to be established, and identified for comparative re-use, in particular for AP and PMS. These benchmarks help frame the assessment context, They act as validated human-system performance standards for comparison, gauging the relative strength of collaborative adaptability proficiency in development prototypes, and providing estimates of human, technical and autonomous system readiness (Taylor 2015 [1]).

3.3 Autonomy Task work

3.3.1 Trustworthiness

Selection of assessment components of autonomy task work seems particularly problematic since much of the detailed functioning of algorithms lacks observability at most commander/operator interfaces. This is reflected in common need for explanation and increased transparency at autonomous systems user interfaces. However, recent work under UK MoD Autonomous Systems Underpinning Research programme, SATLAOP[1] (Searle 2007 [6]) has investigated improving understanding of the requirements for trustworthiness of unmanned systems, based on Publicly Available Specification 754 Software Certification [7]. The results of this work helps guide selection of components of trustworthiness to include safety, reliability, availability, resilience and security components. With the addition of a component for dependability based on human factors literature on trust (Yagoda 2011 [8]), and using the notion of software operation predictability, defined in the sense of adherence to planning in the execution of operations, Dstl have developed and tested both a single and a multi-dimensional subjective rating scale protocol for estimating autonomy task work trustworthiness. The more mature multi-dimensional DSTL HAT Trustworthiness rating scale, with dimension definitions and anchors, is shown below in Figure 7.

HAT TRUSTWORTHINESS ASSESSMENT PROTOCOL								
Low Anchors	DIMENSION							High Anchors
	Low 1	2	3	4	5	6	High 7	
Unreliable; Uncertain; Inconsistent	**Reliability** The ability of the system to operate on missions, perform tasks and deliver effects as specified							Reliable; Certain; Consistent
Undependable; Fickle; Defective	**Dependability** The ability of the system to be relied upon to operate on missions, perform tasks and deliver effects as specified							Dependable; steady; loyal; faithful; responsible; unfailing
Unpredictable; Unlikely; Unexpected	**Predictability** The ability of the system to respond to events and to operate, perform and deliver effects consistently and reliably as planned and anticipated							Predictable; Likely; Expected
Unavailable; Unattainable; Unready	**Availability** The ability of the system to operate on missions, perform tasks, and deliver effects when requested							Available; Attainable; Unready
Brittle; Weak; Slow	**Resilience** The ability of the system to transform, renew and recover in timely response to events							Resilient; Robust; Agile
Unsafe; Unprotected; Unstable	**Safety** The ability of the system to operate without harmful states							Safe; Protected; Stable
Unsecure; Vulnerable; Risky	**Security** The ability of the system to remain protected against accidental or deliberate attacks							Secure; Protected; Assured

Figure 7: Dstl HAT Trustworthiness Protocol

The trustworthiness protocol has been used by the Autonomy Programme in various synthetic environment trials with front line military operators. Data from those trials has shown that operators and software developers are able to complete the protocol understanding the dimensions without difficulty (Taylor, Keirl, Thorpe and Grabham, 2018 [2]). The results indicate that the seven trustworthiness components potentially have useful sensitivity, discriminatory and diagnostic power in assessing HAT. Figure 8 shows the Trustworthiness Ratings obtained on the DAY17 C2 Autonomy Trail from individual software developer SMEs assessing the five individual software component capabilities contributing to the operator UxV GCS system, with comparison WITS Trial Min-Max benchmarks. Data from those trials also indicate that the trustworthiness assessments are not solely dependent on training on a system but on the usability and TTPs associated with the system use.

[1] Self-Aware Trustworthiness Levels and Assurance with Operational Policy

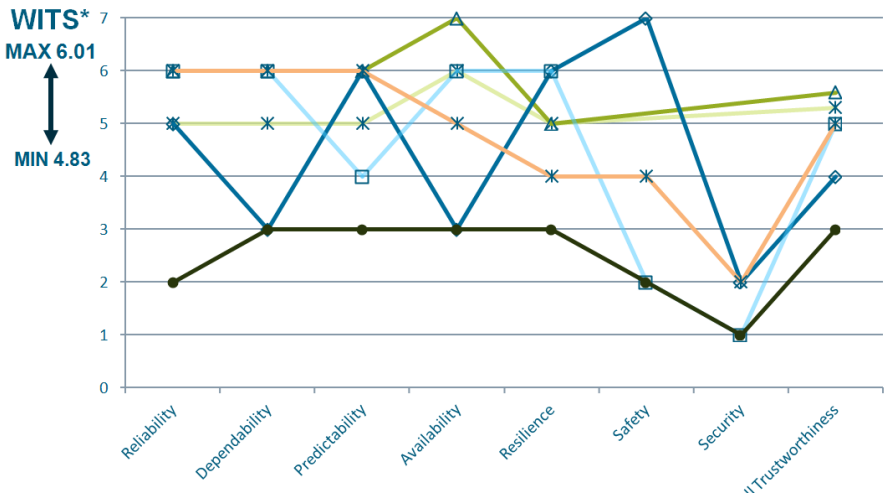


Figure 8: Results of Dstl Trustworthiness Ratings for 5 GCS Software Tools Components from DAY17 C2 Autonomy SE Trial with WITS Trial Comparison Benchmarks

3.4 Human Autonomy Teaming System (Whole System)

3.4.1 Capability Maturity Model

To help guide UK MoD trials in the assessments of autonomy teaming dimensions (human-autonomy and autonomy-autonomy), Dstl have proposed, developed and tested a model with metrics for assessing HAT capability maturity (Taylor, Keirl, Thorpe and Grabham, 2018 [2]). The Dstl HAT Capability Maturity Model (HAT CMM) approach is based on the 1980s software CMM work, originating from Carnegie Mellon University(Carnegie Mellon, 2002 [9]), with consideration of properties of later CMM derivatives, including more people-based dimensions (Systems Engineering, Usability, Organisation, People, Team Process). The underpinning CMMs for Dstl HAT CMM are derived from the basic five software CMM levels (initial through to optimising) with two additional extending, relevant levels, conveniently providing compatibility for integration with the seven level CAPTEAM metrics. The levels progress from Level 0-Not Performed to Level 6-Adaptive, passing through Level 1-Initial, Level 2-Recognised, Level 3-Defined, Level 4-Managed, and Level 5-Optimising. The seven levels of Dstl HAT CMM are each discretely calibrated and defined, with differentiation in terms of goals, processes, and behaviours.

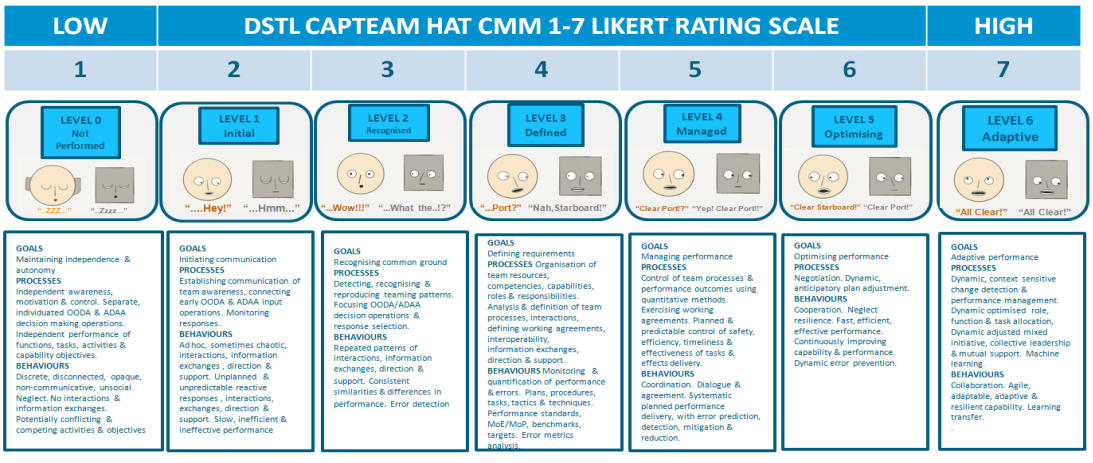


Figure 9: Dstl CAPTEAM HAT Capability Maturity Model Rating Scale

The Dstl HAT CMM Protocol has been applied successfully in recent Dstl Autonomy Research Programme SE and LVC trials. The protocol was used to assess the maturity of UxV GCS technology concepts and systems for assured C2 of Autonomy with military operators. It was found that the maturity assessment concept, including the dimensionality, calibration and scoring method, are novel, complex and needing familiarity and training to understand. Experience and training in relevant and representative military operations and in the proposed technical system use are essential to achieved stability and reliability in HAT maturity identification and classification. Notwithstanding, the capability maturity approach has been found to have both usability and utility for HAT C2 assessment purposes with military operator participant testing in both the DAY 17 SE and CB17Air LVC C2 of Autonomy Trials, The data obtained has indicated that Dstl HAT CMM has the potential to provide both sensitivity and discrimination power, with diagnostic utility, in relation to multiple dimensions of systems readiness and maturity, The scope of potential applications ranges across a diversity of relevant trial and system quality subcomponents, including the maturity of the military operations tested and the maturity of the system mission enabling autonomy technologies.

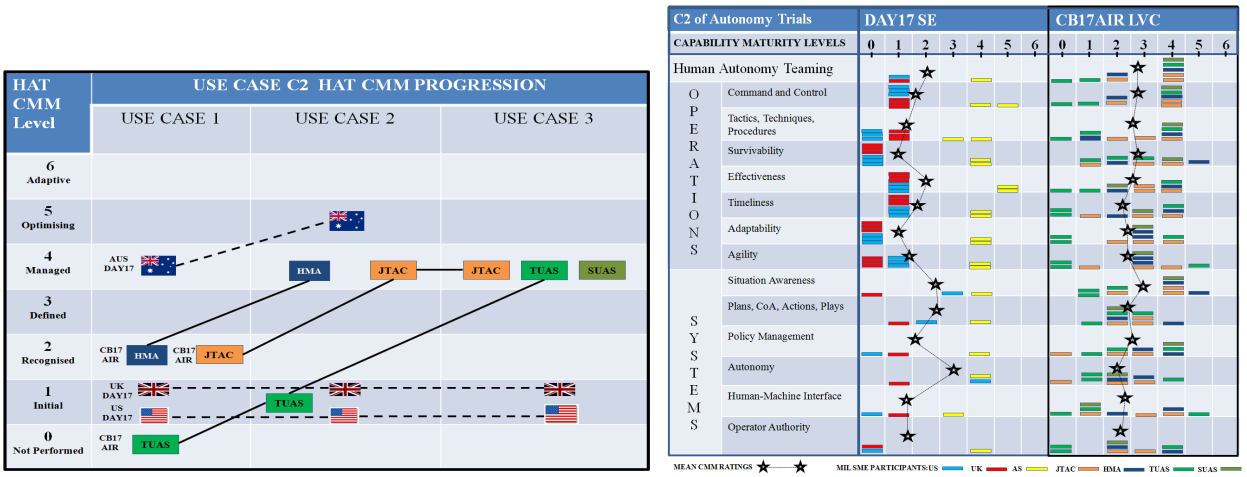


Figure 10: Results of Dstl HAT CMM from DAY17 SE and CB17Air UAS LVC C2 of Autonomy Trials.

3.4.2 Risk Assessment

In Synthetic Environment (SE) and Live, Virtual Constructive (LVC) Trials of Mission Enabling Technologies, assessment methods and protocols typically focus on operator and system performance and effectiveness, necessarily set within the conditions, assumptions, constraints and affordances of the SE/LVC operational trials design environment. To understand the Human Autonomy Teaming system from a broader perspective and identify wider potential requirements and issues, including Defence Procurement Lines of Development (DLODs) that might otherwise be missed, a Dstl Risk Assessment protocol was developed for the WITS[2] trial in March 2017 [10]. The Dstl WITS Risk Assessment protocol was an adaptation of the risk matrix approach used widely in hazard and safety risk assessment. The risk matrix provides identification of calibrated levels of risk probability and impact severity. The Dstl WITS Risk Assessment protocol modification sought to use this framework to address teaming risks for missions. The specific aim of this Dstl Mission Teaming Risk Assessment protocol was to identify the risks of teamwork and the implemented enabling technologies with the focus on the effects on mission success. It is necessary to note, that the capturing of mission teaming risks is focused away from specific platform risks which are typically used in risk based safety assessments of aircraft systems.

[2] Wildcat ISTAR Teaming for Strike, trial in March 2017 was designed to assess the benefits of teaming with potential future platform upgrades

Teaming Risk Assessment Protocol

TEAMING RISK					
NAME..... ROLE/POSITION.....					
DATE..... RUN NUMBER.....					
Mission Critical Decision Teaming Event:					
Description of Risk:					
Mitigation:					
PROBABILITY/ SEVERITY	Extremely Improbable	Improbable	Remote	Occasional	Frequent
Catastrophic					
Severe					
Major					
Minor					
Negligible					

Teaming Risk – Probability and Severity

RISK PROBABILITY OF OCCURRENCE					
Classification	Frequent	Occasional	Remote	Improbable	Extremely Improbable
Quantitative Definition	Likely to occur many times during mission	Likely to occur sometime during mission	Unlikely but possible to occur during mission	Very unlikely to occur during mission	Almost inconceivable to occur during mission

RISK CONSEQUENCE SEVERITY					
Classification	Catastrophic	Severe	Major	Minor	Negligible
Definition	Total failure to achieve mission objectives	Severe reduction in achievement of mission objectives	Major reduction in achievement of mission objectives	Minor reduction in achievement of mission objectives	No effect on mission

Teaming Risk Acceptability Matrix

RISK ACCEPTABILITY					
PROBABILITY/ SEVERITY	Extremely Improbable	Improbable	Remote	Occasional	Frequent
Catastrophic	Review	Unacceptable	Unacceptable	Unacceptable	Unacceptable
Severe	Review	Review	Unacceptable	Unacceptable	Unacceptable
Major	Acceptable	Review	Review	Review	Unacceptable
Minor	Acceptable	Acceptable	Acceptable	Acceptable	Review
Negligible	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable

Figure 11: Dstl Mission Teaming Risk Assessment – Protocol, Definitions and Acceptability Matrix

Risk is defined as a combination of the probability of occurrence and the severity of that occurrence if it should happen. A qualitative probability of occurrence between frequent and extremely improbably and a severity between Negligible and Catastrophic are assigned to each identified risk. The combination of probability and severity results in a matrix of combinations, with catastrophic/frequent the worst outcome and negligible/extremely improbable the best outcome.

The military SMEs (both participants and observers) are required to identify possible risks either that occurred during the scenario or that they could identify as possibly happening as a result of the implemented concepts during the trial. The implemented concepts could either be team orientated, or system/ technology orientated. As a part of the Protocol, the SMEs are able to suggest potential mitigations that may reduce the risk.

Experience with the application of the Dstl Mission Teaming Risk Assessment protocol initially on the WITS trial, and in its development and application subsequently for other Dstl Autonomy SE Research trials, indicate that the protocol purpose is familiar and popular with military operators, it provides a relevant, focusing, useable and useful facility, and a most welcome addition to the Dstl HAT assessment toolkit.

4.0 HUMAN AUTONOMY TEAM WORK - REMEDE

4.1 Introduction

Early Human information processing and team work models were developed in the 1940's, which were expanded upon in the proceeding decades. In the late 1980's MM Taylor developed communication layered protocols which introduced the basis protocol loop. Control abstraction layers were developed in the late 1980's and introduced to support systems for effective C2. All of these approaches model different elements

of human autonomy team work. It is necessary to develop a hybrid model that takes elements from each of the different models to create an assessment methodology that accurately covers Human Autonomy Team work. Each of the different methods will be described in detail, before detailing the hybrid approach that has been developed.

4.1.1 Human Information Processing Phases

In literature on cognitive systems engineering, the functioning of automation is commonly decomposed into a series of four sequential phases of machine information processing, namely Information Aquisition>Information Analysis>Decision Selection >Action Implementation, or AADA (Parasuraman et al, 2000 [11]). This widely used AADA framework mirrors the structure of the simple and popular human decision making OODA loop model (Observe>Orient>Decide>Act) based on understanding of pilot decision making in air combat (Boyd, 1986 [12]; Boyd 2001[13]). The OODA framework, and by implication AADA, has been criticised as a critically deficient over-simplification for conceptualisation of C2, when judged in terms of understanding of control systems principles used in cybernetics (Brehmer, 2005 [14]). This is because the OODA/AADA frameworks fail to identify the functions necessary for effective C2 dynamic decision making, A more complete and efficacious dynamic decision loop, or DOODA loop, includes functions identifying what needs to be accomplished (e.g. information collection, sense-making, and planning), and in particular the provision of feedback on products. Feedback on products is essential for dynamic control and effective C2. The key role in C2 of feedback on performance of critical functions - the products of effective functioning - necessitates representation of the Effects of actions, or Results, in models of C2 dynamic decision making e.g. Information> Decision>Action>Results/Effects.

In CAPTEAM, the Task Work model (Workload>SA>Decision Quality>Task Performance) substantially reflects the OODA Loop structure, with raised AP and PMS as high level Results/Effects, augmented by Teamwork Collaboration. PMS can be further usefully decomposed into attributes of Offensive and Defensive Performance (Castor 2009 [5]). In MOD Air Systems applied research on Mission Enabling Technologies, improved Survivability, Effectiveness and Timeliness (SET) has been long regarded as key performance trade attributes and evaluation assessment criteria for Decision Support Systems (DSS) for combat decision making, It is noteworthy that SET closely mirrors the Cost/Quality/Delivery tradespace commonly used elsewhere in the design of business services. In CAPTEAM, SET are identified as key performance assessment attributes of the Decision Making Task Work function.

4.1.2 Communication Dialogue Protocols

The CAPTEAM Team Work model, based on core attributes of Crew Resource Management (CRM) training, identifies Communication as a key factor in effective Team Work, enabling Shared SA and Leadership Support, and critically underpinning good Coordination, Cooperation and Collaboration. Cybernetics research indicates that communication in natural language, and for effective C2, involves multi-modal dialogue loops between actors with layered protocols for primary messaging and feedback (Taylor, M.M [15]). This dialogue protocol model, with feedback on functioning as a core attribute, applies equally to the design and conduct of effective human-computer or human-autonomy interaction in effective C2 operations.

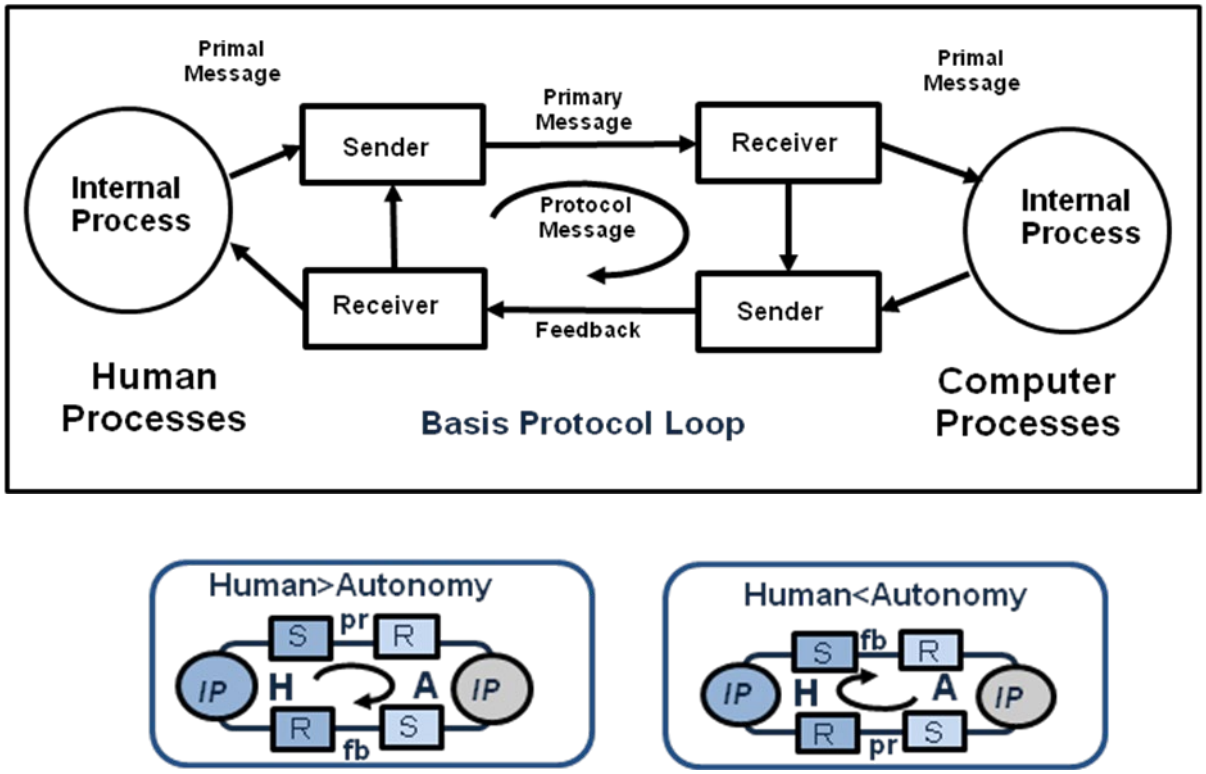


Figure 12: Communication Dialogue Protocols

4.1.3 Control Abstraction Layers

Cognitive systems engineering has shown that systems for effective C2 involve control abstraction layers, modelled largely on the Skills, Rules, Knowledge (SRK), goals-means-acts, hierarchical framework for human behaviour control (Rassmussen [16][17])

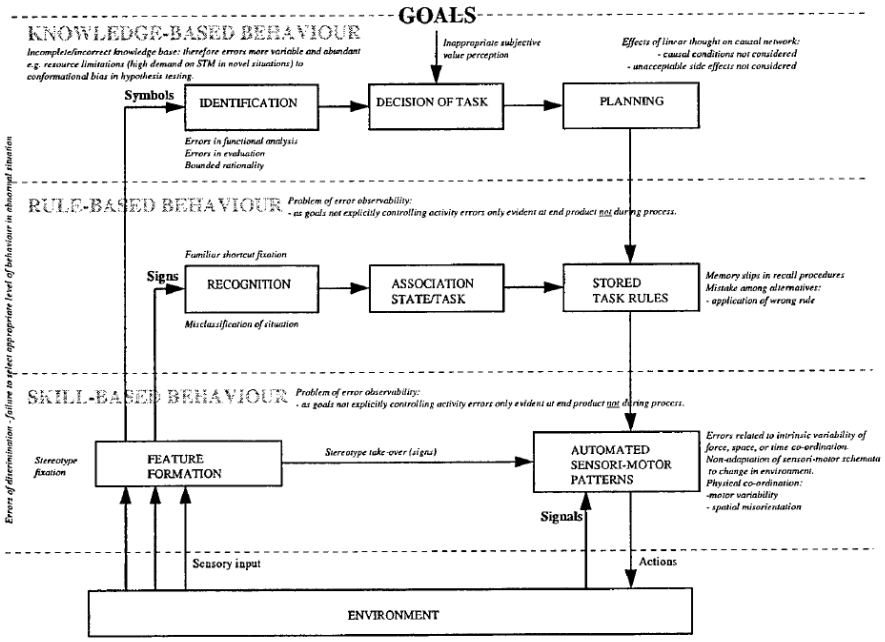


Figure 13: Hierarchical framework for Human behavior control

Subsequent research on Cognitive Work Analysis led to the proposal of a Decision Ladder model with layered components spanning Situation Analysis, Value Judgement, and Planning and Execution: (Observation>Identification>Evaluation > Decision >Planning >Execution) (Vincente, 1999 [18]). The Decision Ladder approach has subsequently been used to represent the collaborative interactions between pairs of actors/agents (Sanderson [19]). This has included control task analysis of user interactions involving PACT levels of automation/autonomy, performed under the DERA Cognitive Cockpit research programme (Taylor, [20], [21], and [22])

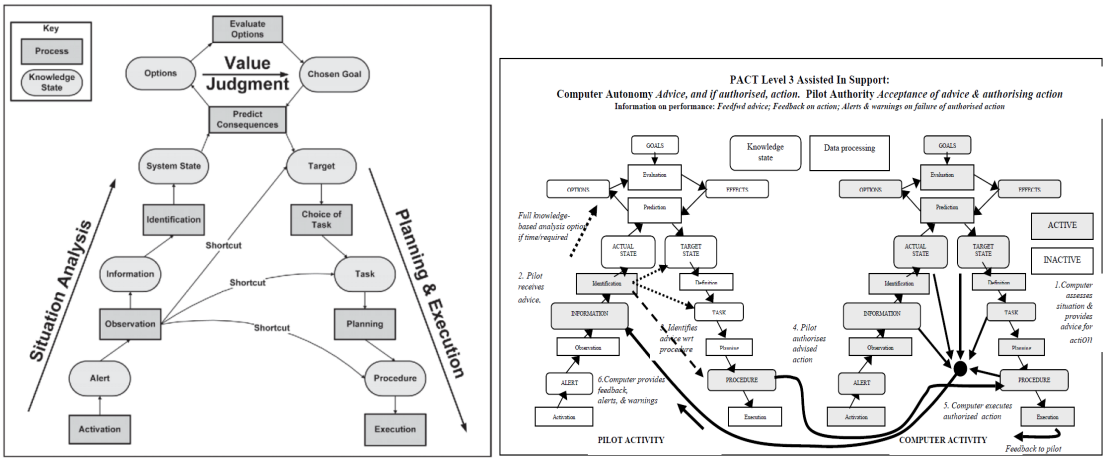


Figure 14: Decision Ladder model and Control Task Analysis of User Interactions

Research on Joint Cognitive Systems (Hollnagel and Woods, 1983 [23], Hollnagel, 1993 [24]) is particularly pertinent to Human-Autonomy Teaming (Taylor 2002 [25]). A multi-layered cognitive control model has been proposed for representation of the control of multiple UxVs (Hollnagel, 2007 [26]). Synchronised,

interacting layers of control identified include Targeting (Governing) > Monitoring (Directing) > Regulating> Controlling (Operating/Tracking).

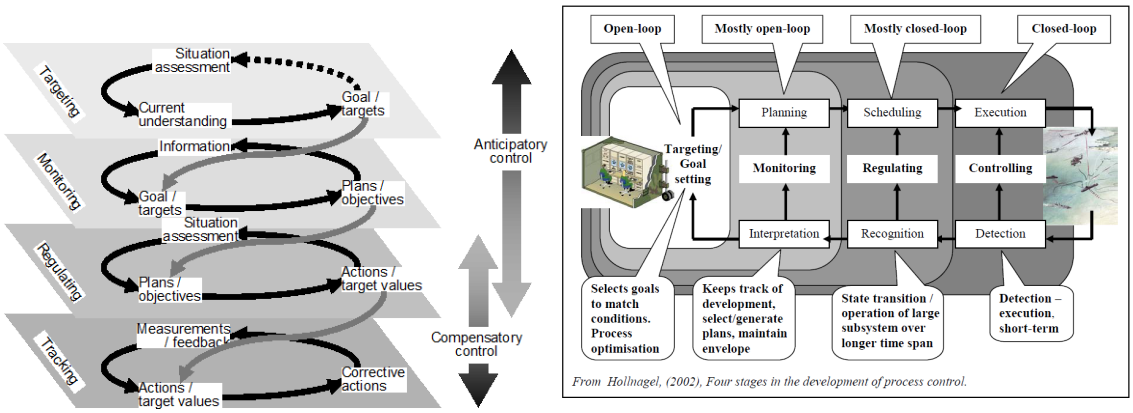


Figure 15: Multi-layered cognitive control model

Informed by this work on cognitive systems engineering and on the concepts of joint cognitive systems, cognitive work analysis and decision ladder control task analysis, work under the Dstl DAMM project identified the applicability of the REMDAER model (Recognise > Evaluate > Mitigate > Disseminate > Acknowledge > Decide > Execute > Report) for representing C2 decision making in a distributed, highly networked military operating environment (Taylor and Grabham 2012 [3]). This framework provided the components for multi-player, distributed or team, decision making cycle, within the operational and tactical C2 architecture C2 OODA or "COODA" layered control system and is illustrated in Figure 16.

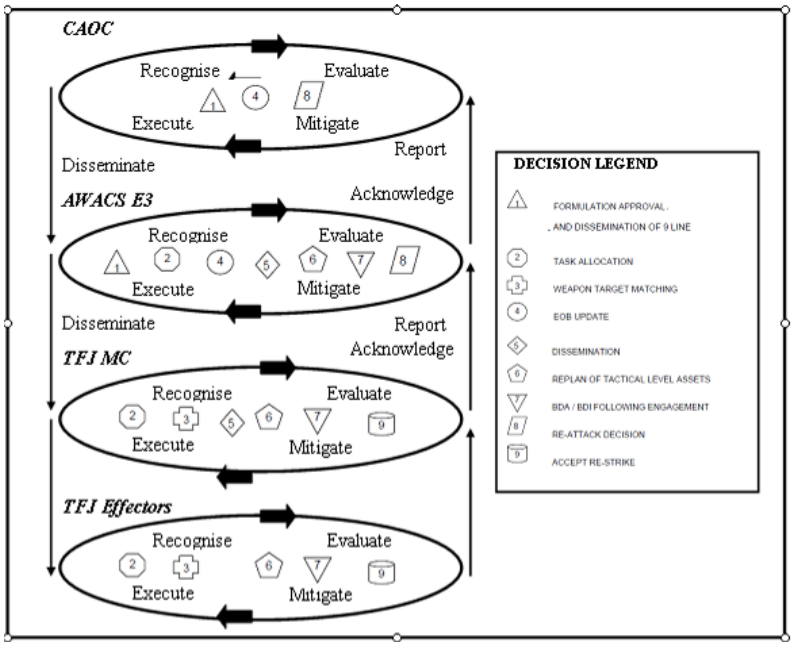


Figure 16: REMDAER Decision Model

4.2 Hybrid Approach

The REMDAER model for the layered control system for airborne mission management was used as an initial step to develop an assessment Protocol for HAT effectiveness. By acknowledging that HAT effectiveness is critically dependent on team member inter-communication, in a similar way to the COODA layered control system, the REMDAER elements were broken down into those that the human and autonomy can complete and those that are dependent on communication between the two. Both humans and autonomy can Recognise, Evaluate, Mitigate and Execute to cause Effects, with Disseminate, Acknowledge and Reporting dependent on the communication Dialogue between the team members.

For assessment purposes, significantly contributing components need to be characteristic and observable; measurable, sensitive and discriminating; and logical, meaningful and evidential. The focus has been on interaction essential information exchange, (and) where the key interactions happen during decision processes, and on the effectiveness of the action taken by the human or the autonomy. In particular the effectiveness of Recognising, Evaluating, Mitigating and Executing (REME) as a team within a mission, and effectiveness of the Communication Dialogue between the human and the autonomy (disseminating, acknowledging and reporting), and the consequences of HAT in terms of specific intended dynamic Effects. Specific measures have been identified to evaluate the effectiveness of the teaming framing around REMEDE.

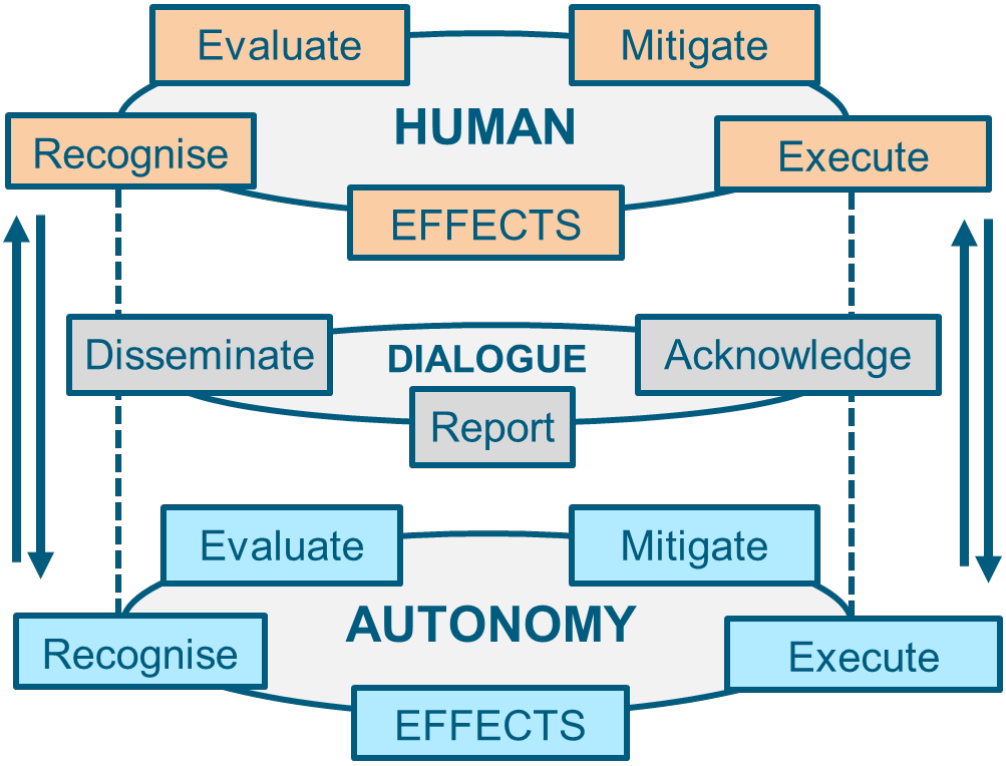


Figure 17: REMEDE Co-Dependent C2 COODA-Model

4.3 REMEDE Assessment Protocols

Seven individual assessment protocols have been proposed initially for assessing the effectiveness of HAT, for validation testing. Six of these protocols are focused on the individual components of the REMEDE model, Recognition, Evaluation, Mitigation, Execution, Dialogue and Effects, with the final protocol being a composite of the REMEDE components. This structure is similar to the Trustworthiness protocol, where the

development and availability of individual protocols and a composite protocol allows for the tailoring of the assessment to the type of experimentation that is being undertaken. Currently neither the composite nor the individual protocols has undergone validation and verification testing, but is scheduled to be utilised in the autumn 2018.

With the exception of the Dialogue protocol, which is unique amongst the different protocols, the remaining six protocols utilise a Likert rating, which ensures compatibility amongst HAT metrics including CAPTEAM, to allow SMEs (operators and/or observers) to score the human functioning component, the system autonomy component and the joint shared communication of each of the REMEDE components. Joint shared communication comprises both the human to autonomy and autonomy to human communication. The Dialogue protocol offers additional granularity for each of the composite REMEDE components in relation to human to autonomy communication and autonomy to human communication, with an individual rating available for each.

EFFECTS	Dstl REMEDE Human Autonomy Teaming Protocol - Effects																				
Impact of actual effects on achievement of Mission and Task Objectives.	Human Functioning Contribution							Joint Shared Communication *H>A & A>H*							System Autonomy Contribution						
	Impact							Impact							Impact						
	Low 1	2	3	4	5	6	High 7	Low 1	2	3	4	5	6	High 7	Low 1	2	3	4	5	6	High 7
Survivability																					
Effectiveness																					
Timeliness																					
Agility																					
Adaptability																					
Offensive Performance																					
Defensive Performance																					
Probability of Mission Success																					
EFFECTS COMPOSITE																					
EFFECTS COMMENTS																					

Figure 18: Dstl HAT REMEDE Effects Component Protocol

5.0 REMEDE VALIDATION AND VERIFICATION TESTING

The REMEDE HAT Assessment Protocol has not yet undergone validation and verification testing However, REMEDE has been adapted for, and will be used in the upcoming STRATUS[3] trials in September 2018 assessing human and system contribution and communication as well as an additional feature asking operators to assess the layered ISTAR contribution towards the REMEDE components. An example assessment protocol from the Stratus evaluation is shown below in Figure 19.

[3] The STRATUS project is assessing the benefits of a layered Intelligence, Surveillance, Target Acquisition, and Reconnaissance (ISTAR) approach in an urban environment with novel autonomy and decision making tools and technologies.

EFFECTS	Dstl REMEDE Human Autonomy Teaming Protocol - Effects																												
Impact of actual effects on achievement of Mission and Task Objectives.	Human Functioning Contribution							Joint Shared Communication							System Autonomy Contribution							Layered C2 ISTAR Collaboration						
	Impact							Impact							Impact							Impact						
	Low 1	2	3	4	5	6	High 7	Low 1	2	3	4	5	6	High 7	Low 1	2	3	4	5	6	High 7	Low 1	2	3	4	5	6	High 7
Raised Understanding																												
Reach Area Depth																												
SA for Specific Effects																												
Survivability																												
Effectiveness																												
Timeliness																												
Agility																												
Adaptability																												
Offensive Performance																												
Defensive Performance																												
Probability of Mission Success																												
EFFECTS COMPOSITE																												
Comments																												

Figure 19: Dstl HAT REMEDE Effects Stratus Specific Component Protocol

6.0 CONCLUSIONS

The assessment of HAT efficacy, while technically challenging can be broken down into a number of individual components of the Human Autonomy Teaming System to simplify the challenge. The understanding of the various components, Human Task work and Teamwork, Autonomy Task work and Teamwork and Human Autonomy Teaming, of the HAT system has been built up over more than 18 years of UK MoD Dstl R&D on Autonomy and Mission Systems. While it is possible, and in some cases easier to understand the HAT system in its entirety, it is often not possible to identify the locus of success or failure at such a high level depending on the system that is being tested.

The Human task work and teamwork elements of the HAT system, through CAPTEAM which is designed to estimate mission efficiency by the metrication of Reward and Effort associated with critical mission events and decision processes. CAPTEAM has been extensively utilised for trials within the UK research programme, with statistical analysis indicating the inter-correlation of the assessment components with relatively beneficial Reward factors, such as improved SA and Decision Quality, were significantly highly correlated with PMS. In comparison, Effort factors and sub-components, such as Workload Stress and Mental Effort a exhibited relatively low, non-significant correlations with PMS.

The Autonomy Task Work component of the HAT system is addressed through a broken down multi-dimensional Trustworthiness scale. While only used in a limited number of synthetic trials, the results indicate that the seven trustworthiness components potentially have useful sensitivity, discriminatory and diagnostic power in assessing HAT, with data from the trials also indicating that trustworthiness is not solely dependent in training on a system but on the usability and TTPs that are associated with the system in use.

The entirety of the HAT system can be assessed through a combination of the HAT Capability Maturity Model, which was derived and developed from software CMMs developed by Carnegie Mellon University and the Risk Assessment. The HAT CMM has been used successfully in recent Dstl Research Programme SE and LVC trials, assessing UxV technology concepts and systems for assured C2 of Autonomy with military operators. It was found that the concept of the maturity assessment, are novel, complex and needing familiarity and training to understand. With experience and training in relevant and representative military

operations and in the proposed technical system use are essential to achieve stability and reliability in HAT maturity identification and classification. Notwithstanding, the approach to capability maturity has been found to have both usability and utility for HAT C2 assessment purposes, with the obtained data indicating that the Dstl HAT CMM has the potential to provide both sensitivity and discrimination power. The Risk Assessment protocol has been used in both synthetic and LVC trials recently, and provides a broader perspective of the HAT system and allows for the potential identification of wider requirements and issues, including those associated with the Defence Procurement Lines of Development that might otherwise be missed. The focus of risks is away from the traditional platform safety risks that are typically addressed in risk based safety assessments of aircraft systems, but instead focused on the risks associated with teamwork and the implemented enabling technologies with respect to mission success.

The new component of the UK HAT System assessment methodologies is the REMEDE assessment protocol. The protocol itself is focused on the Human Autonomy Teamwork component of the entire system, and has been developed from a variety of information processing, communication and team work models. Previously these had been combined to form the REMDAER decision model, as part of previously reported work for the DAMM project, which was focused on a multi-player, distributed or team, decisions making cycle within the operational and tactical C2 architecture "COODA" layered control system. With the critical dependency of team member inter-communication being similar across the HAT system and the REMDAER layered control system it is possible to break down the REMDAER elements into what the Autonomy can complete and what the Human can complete and what dialogue is required between them to be effective. This results in the components of Recognise, Evaluate, Mitigate, Execute to cause effects being applicable to the Autonomy and the Human, with Dissemination and Reporting and Acknowledging forming the dialogue between them. The application of this to the assessment of HAT has been proposed in two different ways, from a high level composite protocol that only looks at high level of the REMEDE components, to an individual protocol for each of REMEDE components. This breakdown of the individual components allows for an increased granularity of the underlying aspects of the different components. At this stage, the REMEDE protocol has been selected for use with the STRATUS project under the UK MoD Dstl Autonomy Research programme, with a specific instantiation developed for the specifics of the trial. This trial will serve as the initial opportunity to undertake verification and validation testing of the REMEDE protocol, the outcomes of which will allow a greater understanding of the discrimination and sensitivity of the data captured.

Autonomy Teamwork has not been addressed within the UK research programme at this time and remains an area of interest in the context of Human Autonomy Teaming and the contribution that Autonomy Teaming has to the understanding the efficacy of the entirety of the HAT System.

7.0 REFERENCES

[1] Taylor R.M. 2015. Human Autonomy Teaming: Verification, Benchmarking and Readiness. In, RPAS Achievements and Challenges. Proceedings of the 2015 Presidents Conference, Royal Aeronautical Society, London' 7-8 October 2015. SG UAS Event Code 772. ID P2PP2R-2015-09-28T 10:34:36. DSTL/TR91332 967705572. ISBN 1 8576831X.

[2] Taylor, R.M., Keirl, H., Thorpe E. and Grabham A. 2018. UK-1: HAT-CV&V Human Autonomy Teaming – Collaboration Verification and Validation. Final Report NATO HFM247. November 2017.

[3] Taylor R.M. and Grabham A. 2012. UK-1: Dynamic Airborne Mission Management. In, Supervisory Control of Multiple Uninhabited Systems – Methodologies and Enabling Human-Robot Interface Technologies. Chapter 13, pp 1-27. Lessons Learnt. Annex A, pp 1-14. HFM-TG-170. AC/323(HFM-170)TP/451. December 2012

[4] Cottrell, R.J. 2011. RE314 July 2011 Strike Warrior III Joint Experimental Trial Report. QinetiQ/AEG/TI/CR1102929/V0.1. 14 Dec 2011. Cody Technology Park, QinetiQ Ltd., Farnborough, UK.

[5] Castor, M. 2009. The use of structural equation modelling to describe the effect of operator functional state on air-to-air engagement outcomes. Linkoping Studies in Science and Technology. Dissertation No. 1251. Linkoping University Institute of Technology. Linkoping, Sweden. ISBN 978-91-7393-657-6 ISSN 0345-7524.

[6] Searle T. 2017. ASUR Self-Aware Trustworthiness Levels and Assurance with Operation Policy 1014_C3_Ph2_060/073/110, 49048-45595R Issue 2, Frazer-Nash Consultancy Ltd.

[7] British Standards Institution, PAS 754:2014, 2014.

[8] Yagoda, R.E. 2011. WHAT! You want me to trust a ROBOT? The development of a human robot interaction (HRI) trust scale. MSc. Psychology Graduate Thesis. North Carolina State University, Raleigh, North Carolina 2011.

[9] Carnegie Mellon. 2002. Capability Maturity Model® Integration (CMMISM), Version 1.1. CMMISM for Software Engineering (CMMI-SW, V1.1). Continuous Representation. CMU/SEI-2002-TR-028. ESC-TR-2002-028. CMMI Product Team, August 2002. Carnegie Mellon Software Engineering Institute, Pittsburgh, PA 15213-3890

[10] Keirl, H, Hennessy, DKR, Thorpe, A, Taylor, R, Halpin, D. 2017. Wildcat ISTAR Teaming For Strike – Phase 1 Experimentation. DSTL/CR101870 Issue 1. March 2017. Defence Science and Technology Laboratory, UK MoD.

[11] Parasuraman, R., Sheridan, T. and Wickens, C. 2000. A Model for Types and Levels of Human Interaction with Automation, IEEE Transactions on Systems, Man, and Cybernetics. – Part A: Systems. and Humans, Vol. 30, pp. 286-297.

[12] Boyd, J.R. 1986. Organic Design for Command and Control. Briefing slides, October 1986

[13] Boyd, J.R. 2001. “An essay on winning and losing”. http://defence and the national interest. d-n-i-net, 2001

[14] Brehmer, B. 2005. One Loop to Rule Them All. In Proceedings of 11TH ICCRTS Coalition Command and Control in the Networked Area: C2 Analysis, C2 Modelling and Simulation, Cognitive Domain Issues. Department of War Studies, Swedish National Defence College, Stockholm, Sweden.

[15] Taylor M.M. 1989. Response Timing in Layered Protocols: A Cybernetic View of Natural Language. In, The Structure of Multi-Modal Dialogue. M.M. Taylor, F. Neel, and D.G.Bouwhuis (Eds). Chapter 11, pp 159-172 Amsterdam: Elselvier Sciences B.V. North Holland.

[16] Rasmussen J. 1986. Information Processing and Human- Machine Interaction: An Approach to Cognitive Engineering. New York: North Holland.

[17] Rasmussen, J., Pejtersen, A.M., and Goodstein, L.P. 1994. Cognitive Engineering: Concepts and Applications. New York: Wiley.

[18] Vicente, K.J. 1999. Cognitive Work Analysis: Towards Safe, Productive, and Healthy Computer based

Work. New Jersey: Lawrence Erlbaum.

[19] Sanderson, P., Naikar, N., Lintern, G., and Goss, S. 1999. Use of cognitive work analysis across the system life cycle: From requirements to decommissioning. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, Houston, TX. Santa Monica: HFES

[20] Taylor, R.M. 2001. Cognitive Cockpit Control Task Analysis. DERA Memo, DERA/CHS3/6.3/14/7, 07 March 2001.

[21] Taylor R.M 2002. Technologies for Supporting Human Cognitive Control. In Human Factors in the 21st Century. NATO MP-077. 18:1-14. RTO Human Factors and Medicine Panel (HFM) Specialists' Meeting, Paris, France, 11-13 June 2001.

[22] Taylor, R.M 2007. Human Automation Integration with Contractual Autonomy. In, Uninhabited Military Vehicles (UMVs): Human Factors in Augmenting the Force. Final Report of RTO Human Factors and Medicine (HFM) Task Group HFM-078/TG017. Chapter 7, Human Automation Integration, pp 3-12 to 29. AC/323(HFM-078)TP69. ISBN 978-92-837-0060-9 July 2007.

[23] Hollnagel, E. and Woods, D. D. 1983. Cognitive Systems Engineering: New wine in new bottles, International Journal of Man-Machine Studies, 18, pps

[24] Hollnagel E. 1993. Human Reliability Analysis: Context and Control. London: Academic Press. 583-600.

[25] Taylor, R.M. 2002. Capability, Cognition and Autonomy. RTO HFM Symposium on "The Role of Humans in Intelligent and Automated Systems", Warsaw, Poland, 7-9 October 2002, NATO RTO-MP-088. KN3, 1 to 26.

[26] Hollnagel, E. 2007. Modelling Multi Layered Control: Application of the Extended Control Model to the Analysis of UAV Scenarios. In, Uninhabited Military Vehicles (UMVs): Human Factors in Augmenting the Force. Final Report of RTO Human Factors and Medicine (HFM) Task Group HFM-078/TG017. Chapter 7.7, Human Automation Integration, pp 73 to 96. AC/323(HFM-078)TP69. ISBN 978-92-837-0060-9. July 2007.

Design Patterns and Ontologies for Situated Human-Agent Collaboration at Organization, Team and Interaction Level

Mark A. Neerincx[1,2] and Jurriaan van Diggelen[1], Martin L. Stolk[1]
[1]TNO Perceptual and Cognitive Systems, Kampweg 5, 3769 DE Soesterberg
[2]Delft University of Technology, Faculty of EEMCS, Van Mourik Broekmanweg 6, 2628 XE Delft
THE NETHERLANDS

<mark.neerincx, jurriaan.vandiggelen, martin.stolk>@tno.nl

ABSTRACT

For the integration of robots into the operations of rescue, reconnaissance, security and defence teams, there is a need for theoretical and empirical grounded human-agent collaboration models. This paper proposes the development of organization, team and interaction design patterns with corresponding ontologies, as re-usable building blocks of human-agent collaboration. A situated Cognitive Engineering (sCE) methodology supports the incremental development and implementation of these building blocks (i.e., ePartner abilities and behaviours) into a perceptual-cognitive agent framework. Core cognition concepts are formalized as ontology design patterns (e.g., "workload" and "situation awareness") and related to the concerning domain ontology (e.g., disaster response with concepts like "point of interest" and "urgency"). In parallel, relevant design patterns are identified or constructed with the corresponding ontologies, at the organization (e.g., norms), team (e.g., work agreements) and interaction level (e.g. explanation). The ontologies are implemented into the prototype and tested on their expected outcomes (i.e. the claims). The European TRADR project (www.tradr-project.eu) provides an example in which an agent-based system for robot-assisted disaster response is developed, where response team dynamically adapt the human-robot teamwork (i.e., task allocation and coordination) to the momentary context.

1.0 INTRODUCTION

Ongoing progress in artificial intelligence, robotics, sensors and network technology feeds the development of software agents that can act as teammates or so-called *ePartners.* These ePartners are applied in defense, safety and security missions [1],[2],[3] and operate robustly in hazardous and chaotic environments. They are supposed to show responsible, agile and flexible behaviors in response to foreseen *and* unforeseen events. Furthermore, their behaviors should be fluently integrated into the overall human-agent teamwork, and harmonized with the activities and expectations of the human teammates. To establish the envisioned human-ePartner collaboration, ePartners require an understanding of human's social, cognitive, affective and physical behaviors, and the ability to engage in partnership interactions (such as explanations of task performances, and the establishment of joint goals and work agreements).

To cope with the design complexity of these systems, theoretically and empirically grounded models are needed that capture these social, cognitive, affective and physical processes. This paper proposes organization, team and interaction design patterns with corresponding ontologies, that can be used as re-usable building blocks for human-agent collaboration in the design process. These patterns are abstractions and generalizations from successful individual design solutions. In our approach, humans and ePartners are part of a joint cognitive system that operates in a specific set of situations. In the system development phase, relevant *situation cognition* models have to be identified or constructed as core concepts of the human-ePartner collaboration. A *domain* model is constructed to explicate how cognition operates in the concerning situated collaborations. Subsequently, the *design patterns* are selected or constructed, which explicate how the actors of the joint cognitive system collaborate successfully in this domain.

As a running example, we will regard an agent-based system for robot-assisted disaster response, as used in the European TRADR project (www.tradr-project.eu). Within this application a response team needs to dynamically adapt the task allocation and coordination to the momentary context [4]. Figure 1 shows the Unmanned Ground Vehicles (UGV) and Unmanned Aerial Vehicles (UAV) that were deployed in the TRADR-project (the photos were taken during a realistic field evaluation of the TRADR-system).

Figure 1: Unmanned Ground Vehicle (UGV; left) and Unmanned Aerial Vehicle (UAV; right) during a field evaluation in the TRADR-project.

Figure 2 presents a part of the disaster response team and its structure (in this example with three human team-members, their ePartners and two robots). The team leader and her ePartner have the lead of two "units", each consisting of an operator, ePartner and robot. The domain model and cognition models of, for example workload, underpin the specification of design patterns at three levels. At the highest level, *organization* design patterns are specified about normative behaviors, such as the work and rest schedules of an organization to comply with a workload norm. At the intermediary level, successful collaborative behaviors of specific *teams* are specified, such as successful work agreements on the task (workload) allocations of a team leader, the two operators, the three ePartners and the two robots in a specific situation. At the lowest level, the *interaction* design patterns specify how the information is communicated between the actors, such as the user interfaces for sharing workload information.

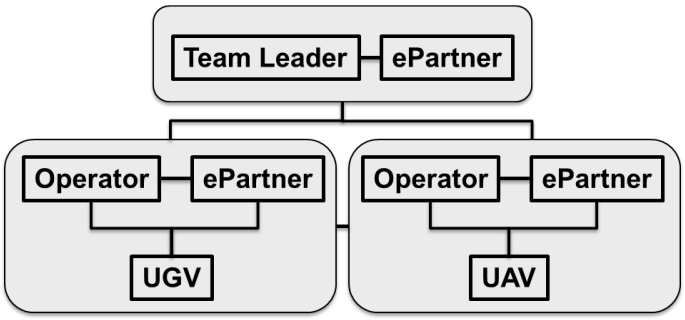


Figure 2: Part of the TRADR disaster response team with a UAV and a UGV.

Section 2 gives a brief overview of a situated Cognitive Engineering (sCE) method for the incremental development of the design patterns with the corresponding domain and cognition models (e.g., "workload" and "situation awareness"), which underpin system's functionality to collaborate. Section 3 discusses a perceptual-cognitive framework to implement this functionality. Subsequently, section 4 elaborates on the

development and application of the design patterns, entailing an extendable, theory and empirical grounded, knowledge base of the ePartner functions and behaviors. Section 5 contains the conclusions of this paper.

2.0 SITUATED COGNITIVE ENGINEERING OF E-PARTNERS

Knowledge about the *operational demands*, *human factors* and *technology* (see Fig 3), underpin the requirements specification of the human-agent system. The sCE methodology supports the acquisition, construction and usage of knowledge from these three foundation components.

First, *ontologies* are used as formal, explicit specifications of the knowledge that the agents need to have about the domain ("operational demands") and the human factors (see Fig 3). An ontology is an explicit representation of declarative knowledge. It is structured around concepts, properties, and relations, which allows for automated reasoning (about this structure). For our purposes, the ontology encompasses consensual knowledge that is important for humans-agent collaboration. Also here, we aim at an extendable set of building blocks, i.e., Ontology design Patterns (OP; [5]). They distinguish knowledge at the upper ('classes'), and lower level ('individuals'). To systematically incorporate human factors knowledge, relevant concepts of situated cognition are formalized as an ontological model, such as situation awareness [6] and workload [7]. For example, Harbers and Neerincx [7] provide a workload model, which distinguishes cognitive load, affective load and mental effort, and can reason about task allocations. The upper level workload ontology (or parts of the ontology) has been studied in and applied to several domains: disaster response [8], space missions [9], ship control [10], and train control [7]. During operation, the domain ontology is added to instantiate the concerning concepts, e.g. for the computational reasoning (such as task allocation based on "urgency" and "workload"). Part of the domain ontology for disaster response is shown in figure 4 (for more information on the domain ontology, see [11]). Note that the ontology refers to human factors concepts (like workload) that are specified in the "Situated Cognition" ontologies.

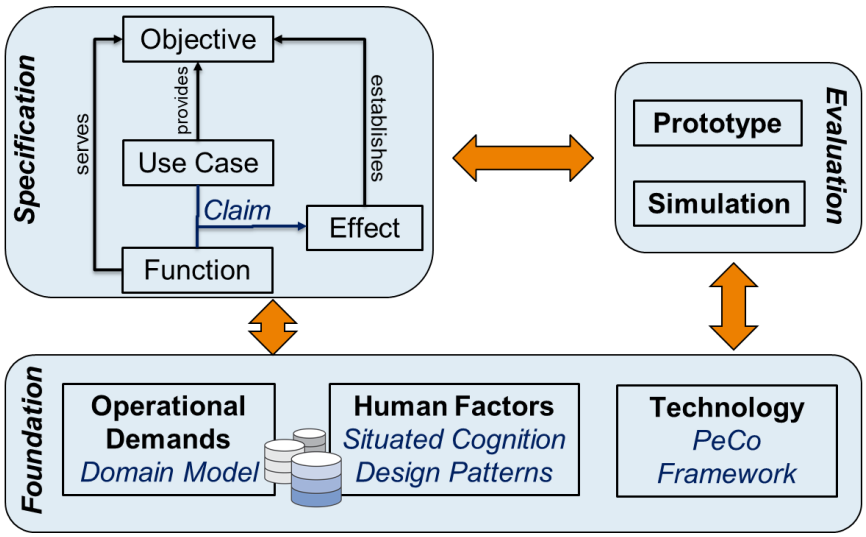


Figure 3: Situated Cognitive Engineering entails an iterative grounding ("foundation"), requirements derivation with its rationale ("specification"), and prototype and simulation testing ("evaluation").

Second, to build on past and current agent systems, and to relate the concerning agent behaviors to each other, it is important to understand and share agents' *design rationales*. In the specification or the functional requirements (see Fig 3), *claims* explicate the design rationale of a specific function: The expected effect of this functionality in a (set of) use case(s). In other words, claims describe the effects of agents behavior on

the condition and performance of (the ensemble of) team actors; the effects can entail an expected positive trade-off of the advantages (upsides) over the disadvantages (downsides). Explicating possible downsides is important, because they have to be checked in the evaluation. For example, a contrastive explanation function for a task reallocation advice might be expected to enhance the trust of a team-leader in the advising agent, with only a minor increase of workload. Claims have to be testable and, if appropriate, refer to stakeholders' values.

Third, to generalize over functions and claims, so-called "*design patterns*" are used to capture the "generic" design rationale by explicating common solutions for classes of problems (e.g., as formulated in a use case [12]). As far as possible, proven models and solutions have to be looked for and maintained as an accessible library in the foundation component. Section 4 will elaborate on the identification and construction of design patterns.

Fourth, the design rationales with their claims are evaluated via simulation and prototyping, resulting in approval, rejection or refinement of the concerning models. In TRADR, the ontology, team and interaction design patterns have been developed and tested in several cycles, applying simulations and realistic field tests with a complete team, including unmanned ground vehicles and unmanned aerial vehicles.

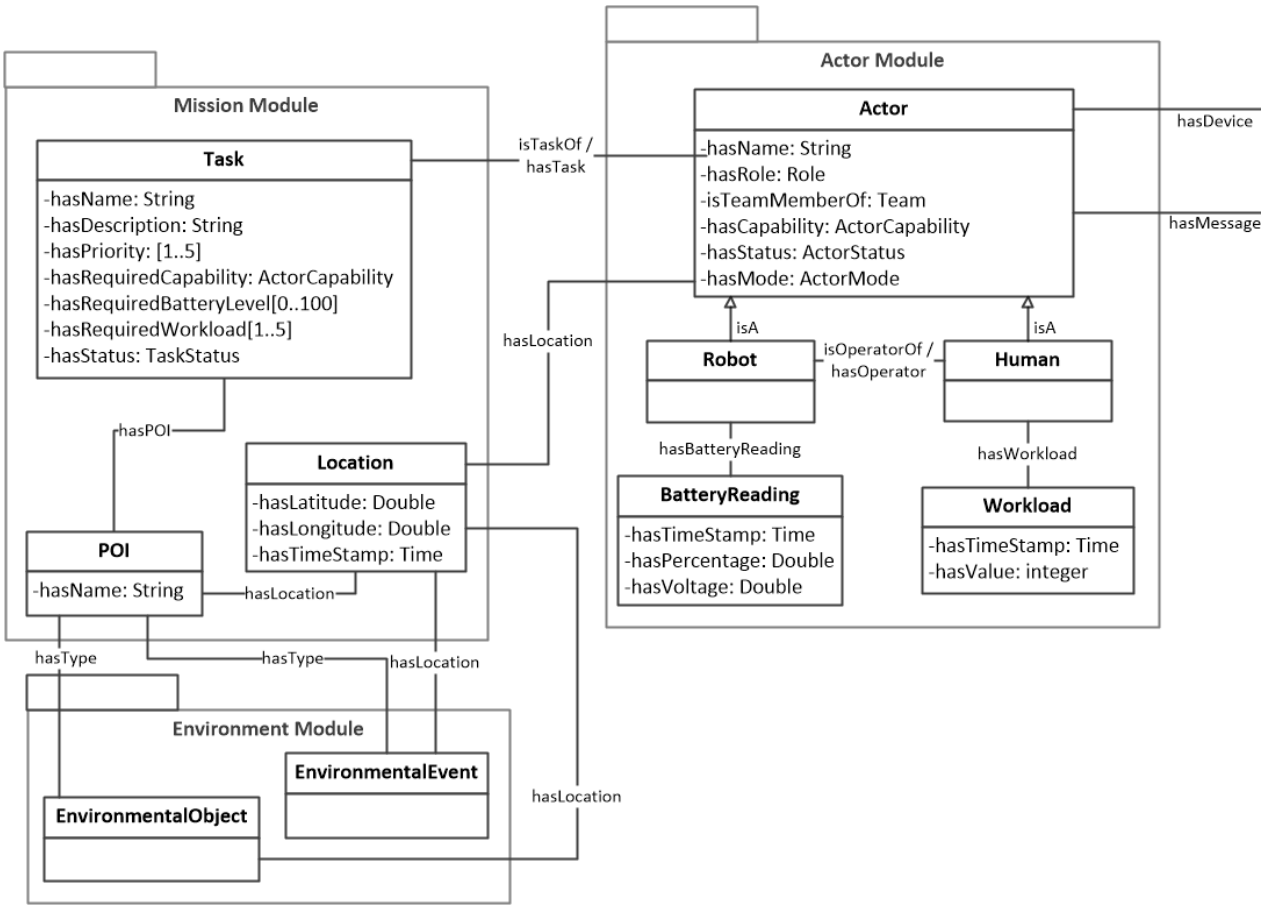


Figure 4: Part of the TRADR ontology (domain model of sCE in Fig 3; more information on the TRADR ontology, see [11]).

3.0 PERCEPTUAL AND COGNITIVE E-PARTNER FRAMEWORK

As a team-member, ePartners need to acquire and show understanding of the meaning, weighting and implementation method of the joint goals they are committed to. This type of understanding and commitment is being built and communicated via three partnership interactions in Human-Agent Teams. Currently, a generic perceptual-cognitive (PeCo) framework-suite is being developed for this purpose ([15], [16], see Fig. 5). This PeCo-framework distinguishes three components: An Objective, Work Agreement and Explanation component.

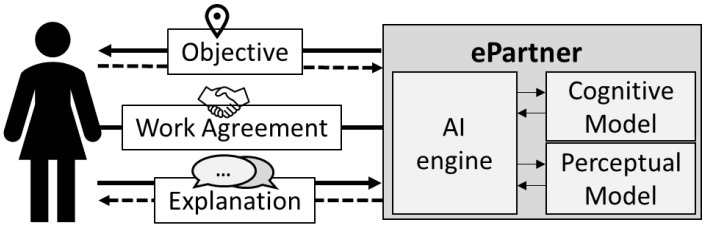


Figure 5: Three human-agent partnership interactions: objective sharing, harmonizing activities via work agreements and explaining task outcomes.

First, active partnerships require the setting, progress monitoring, refinement and adjustment of joint *objectives*, based on a shared understanding, awareness and communication of the objectives and their situated urgencies. Stakeholders' values, strategic intentions and momentary context conditions determine the urgency and feasibility of the objectives to pursue. The choices of the objectives can require re-assessment and adjustment to deal with changing circumstances, possibly just in advance or even during the work processes. Objective ontology design patterns [5] [17] [18] and utility elicitation methods [19] are being developed to support the setting of joint objectives. Adaptivity is an important capability of human-machine teams, which may involve creativity to find a new work process (possibly with an adjusted objective) when the current process is expected to fail. Humans perceive, associate and reason differently than machines; consequently, the perspective-taking and solution space for new (possibly wicked) problems is larger when machines and humans share their perceptions, associations and reasoning.

Second, the proposed human-agent partnerships involve explicit, univocal and possibly adjustable, Work Agreements (WA) on how the tasks should be allocated and performed [16]. Similar to the modeling of the objectives, an ontology design pattern has been developed for work agreements, defining the knowledge required for a team member to set, reason about and adjust work agreements. The WA-ontology defines core concepts with their relations encompassing the knowledge to specify, activate, monitor, and reason about work agreements: <***creditor, debtor, antecedent, consequent, lifespan, acceptance***>. This ontology was implemented and tested for human-robot partnerships in disaster response teams (i.e., in the TRADR-project). Such a team needs to dynamically adapt the task allocation and coordination to the momentary context. A disaster response field test showed that the fire-brigade officers could specify work agreements as required, which, subsequently, brought forward the desired adaptive team behavior of the concerning robot.

Third, we developed explanation models for ePartners that should invoke adequate trust calibration (i.e., to mitigate over- or under-reliance) and improve the learning of the team-members. Currently, a generic perceptual-cognitive explanation (PeCo) framework is being developed for this purpose [15]. The perceptual level provides an Intuitive Confidence Measure [18] and identifies the "foil" that can be used in a contrastive explanation [19]. Such an explanation fits with human explanation dialogues, centering on the question "Why this output (the fact) instead of that output (the foil)?" It reduces the factors in the explanation to the ones that are of interest to the human, so that it can be better interpreted by the human. The cognitive level provides the beliefs, goals and emotions for explanations [15]. Initial evaluations showed that humans

can understand the explanations that PeCo will provide, but further personalization and contextualization is needed.

4.0 DESIGN PATTERNS FOR HUMAN-E-PARTNER COLLABORATION

As mentioned in section 2, design patterns explicate the rationale of the chosen design solution as a common solution for a class of design problems. The following procedure aims at a common pattern library for human-agent collaboration:

1. *Identify key design problems for specific functionality*
2. *Search for available design patterns*
3. *If no pattern can be found,*
   and if it is a general, recurrent design problem:
   - *Start with a Proto Pattern, a pattern "in construction", i.e., a design problem and solution documented in a pattern form (yet lacking empirical grounding)*
4. *Provide different instantiations (examples)*
5. *Test, refine and validate these examples*
6. *If successful:*
   - *Make the Design Pattern accessible in library (of best practices)*

In our approach, the pattern language comprises a structure, template and ontology. Patterns can be constructed at different abstraction levels with different types of relationships, forming the building blocks of the research and development of team agents. We distinguish three pattern levels: Organization (e.g., work and rest schedules), team (e.g., constructive vs. destructive) and interaction (e.g., direct feedback).

Organizations often maintain norms for the joint activities of the workers. These norms can be explicated and formalized as policies: Explicit enforceable constraints on human and agents performance in a given situation, as authorization ("permissions") and obligation [23]. This way, policies govern the behavior of specific agent roles within the organization to enforce normative behaviors (i.e., they apply to a set of actors). The normative model allows to reason about (and to verify) obligations, permissions, and prohibitions, based on deontic logic [25]. Example policies for human-robot collaborations come from the space [23] and automotive domain [24]. In the last domain, the policies specified when an autonomous car should provide the driver with personalized supportive information, and when to hand over control to the driver. Both policies were linked to the driver's workload ([24], cf. the workload ontology of section 2). In these examples and in the TRADR project, organization design patterns have not yet been formulated. However, we have identified some "Proto Patterns" (see above) that need to refined and validated, such as the rest and work schedule pattern mentioned in the introduction of this paper, and design patterns for dealing with hazardous substances in populated areas. Following such patterns, the fire fighters and robots would operate within the safety and health norms of the organization. Van Diggelen et al. [13] provide a design pattern language for teamwork, which seems to be a suitable starting point to describe the normative joint activities at the level of the organization.

Team design patterns specify generic reusable behaviors of actors for supporting effective and resilient teamwork [13]. These patterns can be used to describe the dynamics of teamwork over longer periods of time. They can be constructive (i.e. leading to a more coherent team) or destructive (i.e. leading to a less coherent team), which can be observed in the behavior of an individual or group of actors. During a field test, for example, the TRADR team showed an individual, destructive micro-management pattern of the team-leader. By formalizing and implementing this pattern in the PeCO-framework, the agent can mitigate this behavior in future situations.

Interaction design patterns describe the shaping of the (multimodal) human-agent interactions, such as the dialogue acts for work agreements in TRADR [16]. Figure 6 presents part of the Interaction Design patterns for Human-Agent Collaboration (HAC) that have been specified so far. It distinguishes patterns for a Tailored Situation View (SV), Adjustable Work Agreements (WA) and Harmonized Interactive Notifications (IN) (see [13] for more details on these patterns).

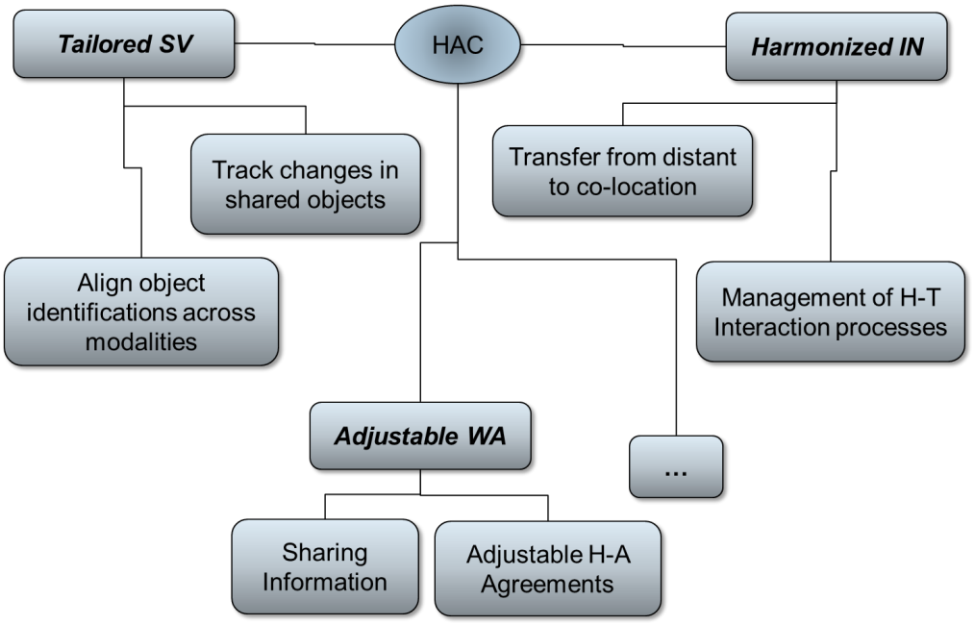


Figure 6: Part of the Interaction Design patterns for Human-Agent Collaboration (HAC), distinguishing patterns for a Tailored Situation View (SV), Adjustable Work Agreements (WA) and Harmonized Interactive Notifications (IN) (see [13] for more details on the patterns).

5.0 CONCLUSIONS

For the integration of robots into the operations of rescue, reconnaissance, security and defense teams, there is a need for theoretical and empirical grounded human-agent collaboration models. This paper proposes the development of organization, team and interaction design patterns with corresponding ontologies, as re-usable building blocks of human-agent collaboration. A situated Cognitive Engineering (sCE) methodology supports the incremental development and implementation of these building blocks (i.e., ePartner abilities and behaviors) into a perceptual-cognitive agent framework. Core cognition concepts are formalized as ontology design patterns (e.g., "workload" and "situation awareness") and related to the concerning domain ontology (e.g., disaster response with concepts like "point of interest" and "urgency"). In parallel, relevant design patterns are identified or constructed with the corresponding ontologies, at the organization (e.g., norms), team (e.g., work agreements, and interaction level (e.g. explanation). The ontologies are implemented into the prototype and tested on their expected outcomes (i.e. the claims).

A key proposal of this paper is to build on past and current agent systems via the specification and sharing of *design rationales*. This rationale includes the explication of testable claims ***and*** the derivation of design patterns. The design patterns are formulated at three abstraction levels as building blocks of the research and development of team agents or ePartners: Organization, team and interaction. A second key proposal is to use *ontologies* that formalize human factors concepts and relate them to the domain concepts. These ontologies underpin the knowledge base ("beliefs") and reasoning of the TRADR agents, for example for proposing, adjusting, accepting, fulfilling or rejecting a work agreement. By formalizing and implementing

constructive and destructive team patterns in the knowledge base, the agent can detect and support constructive behaviors and mitigate destructive behaviors.

Following the proposed methodology, robots can evolve as partners in disaster response by instantiating design patterns for the (a) sharing objectives, beliefs and experiences, (b) committing to norms and work agreements, and (c) uptake and learning of explanations and feedback (Fig 5).

ACKNOWLEDGEMENTS

This work was partly funded by the EU FP7 TRADR project (grant no. 60963) and the Dutch Ministry of Defence's research program (project V1719). We thank the fire fighters from the "Gezamenlijke Brandweer Rotterdam" in the Netherlands for participating in this study.

REFERENCES

[1] Neerincx, M. A. (2011). Situated cognitive engineering for crew support in space. *Personal and Ubiquitous Computing, 15*(5), 445-456

[2] Bosse, T., Breebaart, L., Diggelen, J. V., Neerincx, M. A., Rosa, J., & Smets, N. J. (2017). Developing ePartners for human-robot teams in space based on ontologies and formal abstraction hierarchies. *International Journal of Agent-Oriented Software Engineering, 5*(4), 366-398.

[3] Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V., & Johnson, W. W. (2017, July). Why Human-Autonomy Teaming?. In *International Conference on Applied Human Factors and Ergonomics* (pp. 3-11). Springer, Cham, Germany.

[4] Kruijff-Korbayová, I., Colas, F., Gianni, M., Pirri, F., de Greeff, J., Hindriks, K., Neerincx, M.A., Ögren, P., Svoboda, T. & Worst, R. (2015). TRADR project: Long-term human-robot teaming for robot assisted disaster response. *KI-Künstliche Intelligenz, 29*(2), 193-201.

[5] Gangemi, A. and Presutti, V. (2009). Ontology design patterns. In *Handbook on ontologies*, 221-243. Springer, Berlin, Heidelberg, Germany.

[6] Smets, N. J., Neerincx, M. A., Jonker, C. M., & Båberg, F. (2017, March). Ontology-based situation awareness support for shared control. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 289-290). ACM.

[7] Harbers, M., & Neerincx, M. A. (2017). Value sensitive design of a virtual assistant for workload harmonization in teams. *Cognition, Technology & Work, 19*(2-3), 329-343.

[8] Colin, T. R., Mioch, T., Smets, N. J., & Neerincx, M. A. (2012). Estimating an operator's cognitive state in real time: A user modeling approach. In *RO-MAN* (pp. 627-633).

[9] Cohen, I., den Braber, N., Smets, N. J., van Diggelen, J., Brinkman, W. P., & Neerincx, M. A. (2016). Work content influences on cognitive task load, emotional state and performance during a simulated 520-days' Mars mission. *Computers in Human Behavior*, *55*, 642-652.

[10] Neerincx, M. A., Kennedie, S., Grootjen, M., & Grootjen, F. (2009). Modeling the cognitive task load and performance of naval operators. In *International Conference on Foundations of Augmented Cognition* (pp. 260-269). Springer, Berlin, Heidelberg.

[11] Looije, R., Neerincx, M. A., & Hindriks, K. V. (2017). Specifying and testing the design rationale of social robots for behavior change in children. *Cognitive Systems Research, 43*, 250-265.

[12] Schulte, A., Donath, D., & Lange, D. S. (2016). Design patterns for human-cognitive agent teaming. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 231-243). Springer, Cham.

[13] Neerincx, M. A., van Diggelen, J., & van Breda, L. (2016). Interaction design patterns for adaptive human-agent-robot teamwork in high-risk domains. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 211-220). Springer, Cham, Germany.

[14] Mioch, T., Peeters, M.M.M. and Neerincx, M.A. (2018). Improving Adaptive Human-Robot Cooperation through Work Agreements. *RO-MAN2018 27th IEEE International Symposium on Robot and Human Interactive Communication.*

[15] Saad, E., Hindriks, K. V., & Neerincx, M. A. Ontology Design for Task Allocation and Management in Urban Search and Rescue Missions. *ICAART 2018.*

[16] Van Diggelen, Neerincx, M.A., Peeters, M.M.M., and Schraagen, J.M.S. (in press). Developing effective and resilient human-agent teamwork using team design patterns. *IEEE Intelligent Systems*.

[17] Mioch, T., Ledegang, W., Paulissen, R., Neerincx, M. A., & van Diggelen, J. (2014). Interaction design patterns for coherent and re-usable shape specifications of human-robot collaboration. In *Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems* (pp. 75-83). Association for Computing machinery (ACM).

[18] Neerincx, M.A. , Van der Waa, J. , Kaptein, F. and Van Diggelen, J. (2018). Using perceptual and cognitive explanations for enhanced human-agent team performance. *EPCE 2018, LNAI 10906*, pp. 1–11, Springer.

[19] Peeters, M.M.M. and de Visser, E.J. (2018). "Longitudinal trust development in human-agent-robot teams," in *Workshop on Longitudinal Human-Robot Teaming, held at the Human-Robot Interaction Conference*, February 2018, Chicago, IL, USA.

[20] González-Ortega, J., Radovic, V., and Insua, D.R. (2018). Utility elicitation. In *Elicitation*, pp. 241-264. Springer, Cham

[21] Van der Waa, J., Diggelen, J. and Neerincx, M.A. (2018). The design and validation of an intuitive certainty measure. *Proceedings of IUI 2018 Workshop on Explainable Smart Systems*.

[22] Van der Waa, J., Robeer, M..J., Van Diggelen, J., Brinkhuis, M.J.S. and Neerincx, M.A. (2018). Contrastive Explanations with Local Foil Trees. *IJCAI 2018 Workshop on Human Interpretability in Machine Learning (WHI).*

[23] Van Diggelen, J. M. Bradshaw, J.M., Johnson, M. Uszok, A. and Feltovich, P.J. (2010). Implementing collective obligations in human-agent teams using KAOS policies. In: *Coordination, Organizations, Institutions and Norms in Agent Systems V*. Springer, pp. 36–52.

[24] Mioch, T., Kroon, L., & Neerincx, M. A. (2017). Driver readiness model for regulating the transfer from automation to human control. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (pp. 205-213). ACM.

[25] Wieringa, R.J. and Meyer, J.-J. C. (1993). Applications of deontic logic in computer science: A concise overview. *Deontic logic in computer science*, pp. 17–40, 1993.

Measuring Effectiveness of Human Autonomy Teaming

Helen Lashley, Aaron Thorpe, Robert Taylor and Antony Grabham
Defence Science and Technology Laboratory
Dstl Portsdown West, Fareham, PO17 6AD
UNITED KINGDOM

hjkeirl@dstl.gov.uk

ABSTRACT

The aim of this work is to test and evaluate the effectiveness of HAT, specifically the interactions and dialogue between a human operator and autonomy system in joint, collaborative manned-unmanned operations.

Throughout the comprehensive UK MoD Dstl programme of R&D on Autonomy and Mission Systems, a number of different metrics have been developed predominantly focused on exploiting the critical decision methodology to address the different constituent parts of the complete HAT System.

The Human task work and teamwork elements of the HAT system, through CAPTEAM which is designed to estimate mission efficiency by the metrication of Reward and Effort associated with critical mission events and decision processes. The Autonomy Task Work component of the HAT system is addressed through a broken down multi-dimensional Trustworthiness scale. The entirety of the HAT system can be assessed through a combination of the HAT Capability Maturity Model and the Risk Assessment. The new component of the UK HAT System assessment methodologies is the REMEDE assessment protocol. The protocol itself is focused on the Human Autonomy Teamwork component of the entire system, and has been developed from a variety of information processing, communication and team work models.

At this stage, the REMEDE protocol has been selected for use with the STRATUS project under the UK MoD Dstl Autonomy Research programme, with a specific instantiation developed for the specifics of the trial. This trial will serve as the initial opportunity to undertake verification and validation testing of the REMEDE protocol, the outcomes of which will allow a greater understanding of the discrimination and sensitivity of the data captured.

1.0 INTRODUCTION

Human-Autonomy Teaming (HAT) - the creation of a mutually supportive, co-operative, collaborative partnership between the human user and advanced “intelligent” automation technology - has been recognised as the necessary strategic, human-system integration design objective for Unmanned Systems (UxS). HAT is aimed at delivering efficiently and effectively the required agile, adaptive, context sensitive, multi-mission capability, whilst enabling meaningful and effective Command and Control (C2).

Prior to the advent of Artificial Intelligence (AI) and before the introduction of decision making technology, the human commander/operator retained sole responsibility for providing cognitive decision making capability. The commander/operator uses sensing, thinking, values, reasoning, learning, knowledge and memory, coupled with judgement, vision and imagination, shared and developed through social communication and dialogue. Cognitive capability enables human judgement to be reasonable, dependable and reliable in dealing with uncertainties and ambiguity. But it also enables humans to be innovative and creative, thereby anticipating options, contingencies, risks, opportunities and threats. This creative cognitive

capability provides the essential context sensitivity, flexibility, agility and adaptability needed for decision superiority in the complex, dynamic military environment.

Concurrently, the provision of effective test and evaluation methodology for proving HAT efficacy, with verification and validation of human-system performance and effectiveness, has been identified as presenting a significant technical challenge as described by R Taylor [1], [2].

The aim of this work is to describe the test and evaluation methodologies used in determining the effectiveness of HAT, specifically the interactions and dialogue between a human operator and autonomy system in joint, collaborative manned-unmanned operations.

2.0 MILITARY CONTEXT (RATIONALE)

Remote and automated systems employing unmanned and remotely operated platforms, and systems employing embedded automation and self-governing autonomous functionality, are being increasingly employed as enablers of military capability. However, it is widely recognised that in the complex, dynamic and uncertain environment of military operations, human effectiveness and HAT is needed, as an axiomatic design imperative, for autonomous systems to be assuredly safe and effective.

Test methods are needed that are designed to achieve the levels of proof required to support evidence-based decision making in UK MoD Defence Equipment procurement. Selection of test methods involves balancing requirements for innovation and demonstration (e.g. use case testing), compared with the need for robust performance-based trade-space analysis, valuing discrimination power and optimisation (e.g. multi-variate performance testing). Innovation test methods can be less rigorous experimental design with changes to technology, adaptations to Tactics, Techniques, Plans and Procedures (TTPs) and learning encouraged throughout the duration of a trial. The data and outcomes are used as indicators for future research and lessons identified rather than statistically significant results. Discrimination test methods are more robust and rigorous with dependent and independent variables. The trials outputs are more defined and lessons and conclusions can be drawn directly from the data. Figure 1 illustrates the distinction between innovation and discrimination test methods, (R Taylor 2015 [1]). A balanced approach using both test methods has been employed by UK MoD for HAT research, resulting in quantitative and qualitative data as well as validated assessment methodologies for testing HAT effectiveness, as described in this paper.

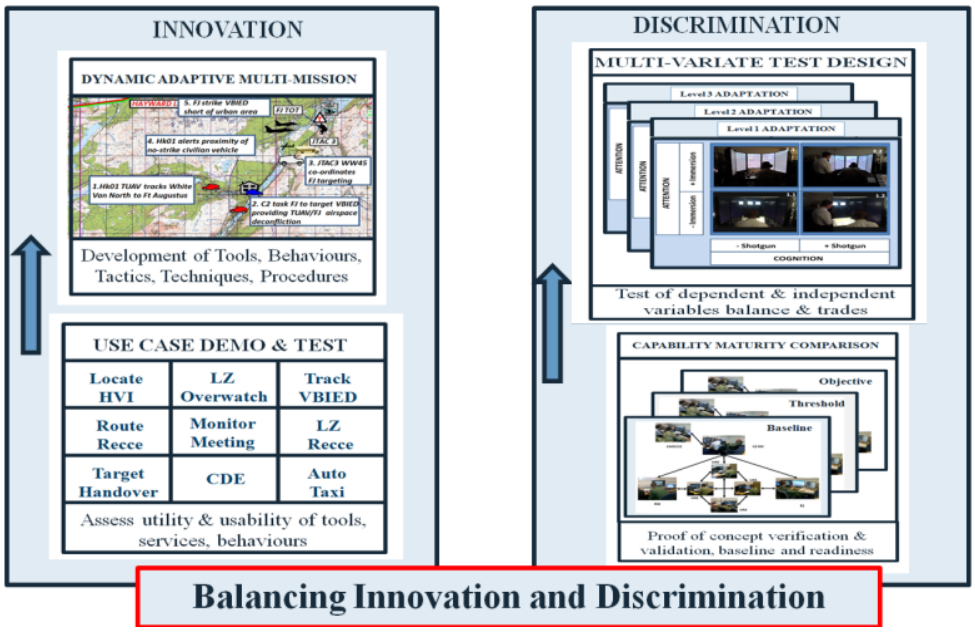


Figure 1: Applicable Test Methods

2.1 Critical Decision Methodology

Underpinning the test methodology, and providing the required human-effectiveness centric focus, Dstl has developed a cognitive capability testing approach, based on dynamic Mission Critical Decision Making (MCDM) methodology. MCDM focuses the locus of testing on Course of Action (CoA) adjustment responses to dynamic mission critical events. Uncertainties in the military combat environment affect understanding of risks and opportunities associated with delivering the command intent and in the performance of associated tasks, with dynamic targets and threats, resulting in adjustments and changes to CoA tactics, techniques, plans, procedures, all affecting the delivery of effects. (Taylor and Grabham, 2012 [3])

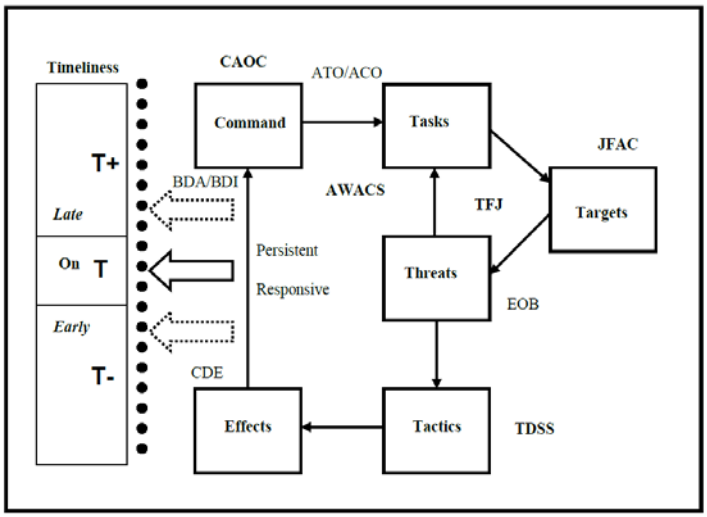


Figure 2: Model of Dynamic Air Mission Management

MCDM employs objective, task analysis and effects-based Measure of Effectiveness and Measures of Performance (MoE/MoP), coupled with subjective decision-based metrics for estimation of human and mission effectiveness (Collaborative Adaptability Proficiency Test Evaluation Assessment Methodology (CAPTEAM)). Essentially, the aim is to capture the agility (speed) and adaptability (quality) of mission commander/operator decision-making in dynamic mission timeline context as illustrated in Figure 3.

Subsequently, under the UK Autonomy and Mission Systems Research Programme, the MCDM methodology was further developed to measure collaborative adaptability proficiency with autonomous systems concepts and technologies. This has included investigating C2-MM decision making by varying and controlling demands for levels of conflict resolution, and by observing collaborative mitigation/adaptation responses, considered in terms of coordination (de-conflicting acts), cooperation (de-conflicting means) and collaboration (de-conflicting goals).

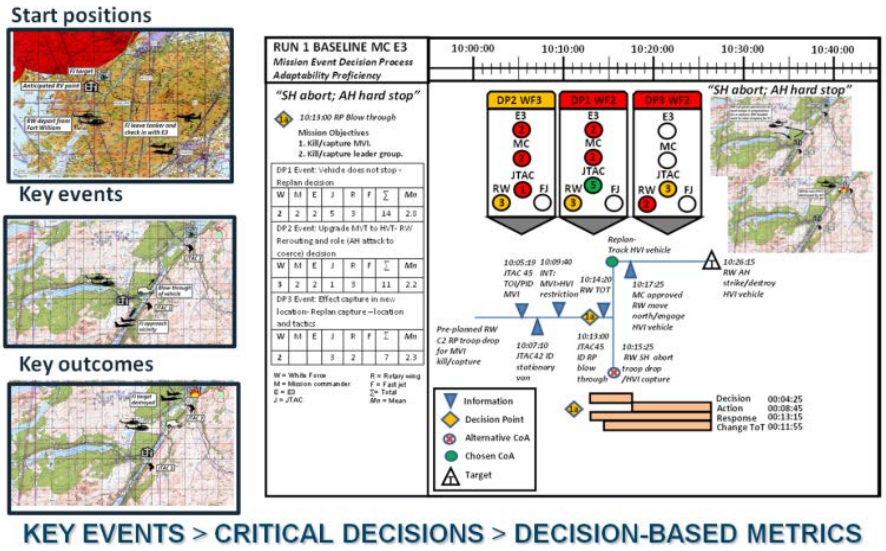


Figure 3: Dstl MCDM Methodology

When designing experiments to measure the effectiveness of HAT in a military context it is important to have militarily representative scenarios which require military operators or autonomous systems to make realistic decisions. Key events must be designed into the scenario to ensure operators are making critical decisions, with effects, in a layered C2 environment, to support our decision based assessment approach.

3.0 HAT METRICS

3.1 Introduction

The technical challenge of proving HAT efficacy, with verification and validation of human-system performance and effectiveness can be broken down in to a number constituent components, the first being the task work performed by a human operator, and the interactions that a human has with other human members of the team. To further add to the HAT system, it is necessary to consider the task work that is undertaken by the autonomy, and the teaming that occurs between autonomous components of the wider team. The final component is the human autonomy teamwork. The combination of the components is illustrated in Figure 4.

Throughout the comprehensive UK MoD Dstl programme of R&D on Autonomy and Mission Systems, a

number of different metrics have been developed predominantly focused on exploiting the critical decision methodology to address the different constituent parts of the complete HAT System. These components are described in the following sections, with the human autonomy teamwork metrics being described in detail.

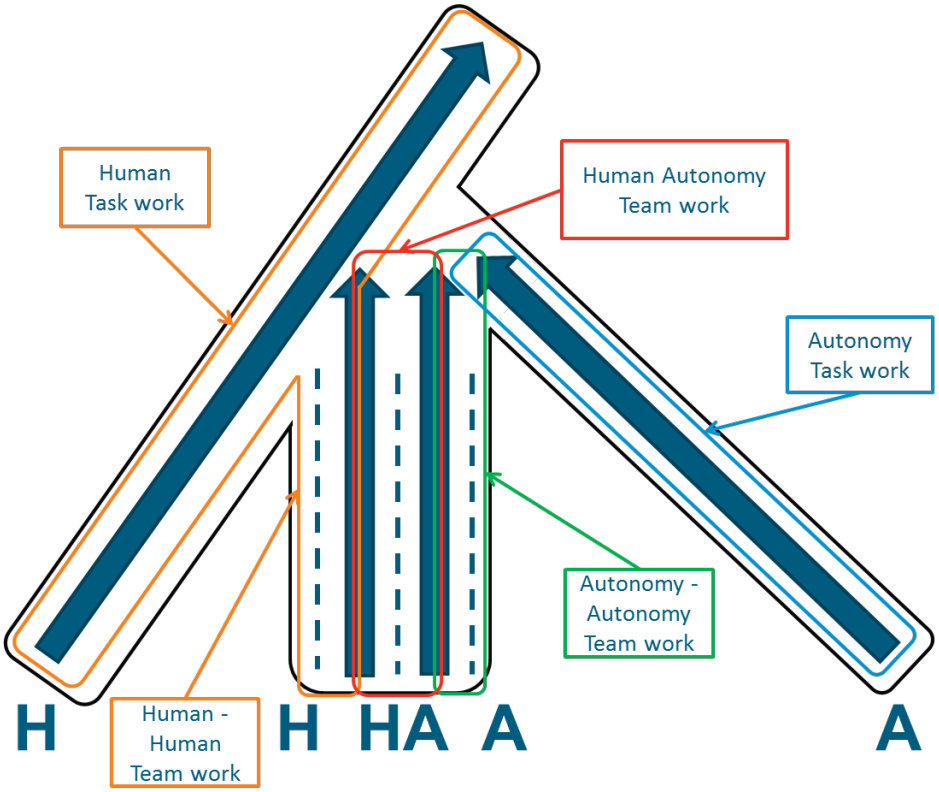


Figure 4: Dstl HAT Component Breakdown

3.2 Human System Teaming

Collaborative Adaptability Proficiency Test Evaluation Assessment Methodology (CAPTEAM) is an independent assessment provided by Dstl, which is designed to estimate mission efficiency by the metrication of Reward and Effort associated with critical mission events and decision processes [3].

The metrics factors measured in CAPTEAM are Workload, Re-plan Task Load, Situational Awareness, Decision Quality, Teamwork, Performance, Tools and Technical System. The metrics are split into Human Task work and Human-Human Teamwork, Task work is assessed in the Participant Protocol and Teamwork in the Subject Matter Expert Observer Protocol using Likert 7-point, Low-High, anchored rating scales.

LOW ANCHORS	DIMENSION Low 1	2	3	4	5	6	High 7	HIGH ANCHORS
WORKLOAD								
Slow/Leisurely/No time pressure	**WL Time Pressure**							*Rushed/Rapid/Frantic*
Very little/Extremely easy	**WL Mental Effort**							*Extremely hard/Stretched*
Relaxed/Placid/Tranquil/Calm	**WL Stress**							*Anxious/Worried/Up tight/Harassed*
REPLAN TASK LOAD								
Forgiving/Simple/Easy	**Replan Decision** Recognise–Evaluate-Mitigate							*Difficult/Demanding/Complex/Exacting*
Forgiving/Simple/Easy	**Replan Action** Disseminate–Acknowledge–Execute-Report							*Difficult/Demanding/Complex/Exacting*
SITUATION AWARENESS								
Stable/Simple/Fixed	**SA Demand on Attentional Resources**							*Unstable/Complex/Variable*
Relaxed/Inattentive/Spare Capacity/Singular	**SA Supply of Attentional Resources**							*Concentrated/Full capacity/Divided*
Uninformative/Meaningless/Unfamiliar	**SA Understanding**							*Informative/Meaningful/Familiar*
DECISION QUALITY								
Indecisive/Doubtful/Guess	**DQ Confidence**							*Decisive /Confident /Obvious*
Dangerous/Vulnerable/Risky	**DQ Survivability**							*Safe/Unthreatening/Secure*
Ineffective/Un-useful/Unproductive	**DQ Effectiveness**							*Effective/Capable/Useful*
Behind/Late/Pressurised	**DQ Timeliness**							*Ahead/On-time/In control*
PERFORMANCE								
Unsatisfactory/Unacceptable/Failure	**Task Performance**							*Perfect/Successful/Satisfied*
Useless/Complicated/Difficult	**Tools Utility**							*Easy/Intuitive/Effective*
No backup plans/ Insensitive/Unresponsive	**Adaptability Proficiency**							*Sensitive/Responsive/Timely effective backup plans*
Fails to meet any mission objectives/0%	**Probability of Mission Success**							*100%/Successfully achieved all mission objectives*

LOW ANCHORS	DIMENSION Low 1	2	3	4	5	6	High 7	HIGH ANCHORS
DECISION QUALITY								
Dangerous/Vulnerable/Risky	**DQ Survivability**							*Safe/Unthreatening/Secure*
Ineffective/Useless/Unproductive	**DQ Effectiveness**							*Effective/Capable/Useful*
Behind/Late/Pressurised	**DQ Timeliness**							*Ahead/On-time/In control*
TEAMWORK								
Silent/Unclear/Confused/Slow/Uninformative	**Communication**							*Clear/Timely/Coherent/Concise/Rapid/ Informative*
Uninformative/Meaningless/Unfamiliar	**Shared SA**							*Informative/Meaningful/Familiar*
Indecisive/Unassertive/Confused/ Uncommunicative	**Leadership**							*Commanding/Directing/Guiding/Initiating management*
Ignoring/Hindering/Blocking/Criticising/ Conflicting	**Support**							*Monitoring/Advising/Correcting/Assisting*
Light/Quiet/Unpressured	**Team Workload**							*Heavy/Busy/Pressured*
PERFORMANCE								
Unsatisfactory/Unacceptable/Failure	**Task Performance**							*Perfect/Successful/Satisfied*
Separate/Independent/Divided/Conflicting	**Collaboration**							*Joint/Shared/Coordinated/Cooperating*
Powerless/Weak/Irrelevant/Not Involved	**Influence Power**							*Influential/Powerful/Forceful/Decisive/Critical*
No backup plans/ Insensitive/Unresponsive	**Adaptability Proficiency**							*Sensitive/Responsive/Timely effective backup plans*
Fails to achieve any mission objective/0%	**Probability of Mission Success**							*100%/ All mission objectives successfully achieved*

Figure 5: Dstl 2011 CAPTEAM Participant Taskwork and Observer Teamwork Assessment Protocols

This assessment methodology relies on military operators to make critical decisions in response to dynamic mission events, with measurable effects, within a realistic military scenario in either a live or synthetic environment. Note that this early 2011 Dstl CAPTEAM protocol was designed for assessing human performance in UK MoD research on Dynamic Air Mission Management (DAMM). Here, the maturity of the automation and autonomy of the technical system was not yet ready to be demonstrated or evaluated as an explicit, discrete capability. In Dstl DAMM research, the focus was on evaluating human effectiveness with distributed networked, C2, at operational and tactical mission levels, with exploitation of human teaming collaboration capability supported by networked mission enabling technologies. The DAMM technical system was treated as being designed to provide tools for human use, in the sense of user-centred mission enabling technologies and operator decision support, with assessment focused on measures of tool usability and utility. Tools usability and utility assessment techniques contrast with the challenges of assessing the efficacy for human use of automatic or intelligent agent technical systems. Such systems are potentially capable of non-deterministic, fully autonomous functioning of operations (sensing, decision making, behaviour), completely independent of human control, with properties such as agency, self-

determination and autonomy associated with advanced Artificial Intelligence (AI), Big Data processing and Machine Learning (ML) technologies.

The Dstl CAPTEAM assessment protocol was used for operator assessments in the 3rd US-UK STRIKE WARRIOR III Joint SE Trial, July 2011 (Cottrell 2011 [4]). Statistical analysis by Dstl of data obtained from application of CAPTEAM indicated the extent to which CAPTEAM component metrics associated with Effort, or Reward are statistically correlated with Probability of Mission Success (PMS) (Taylor and Grabham 2012 [3]). The results showed that relatively beneficial Reward factors, such as improved SA and Decision Quality, were significantly highly correlated with PMS. In comparison, Effort factors and sub-components, such as Workload Stress and Mental Effort a exhibited relatively low, non-significant correlations with PMS. The inter-correlations of the Reward and Effort factors for both Task Work and Team Work in relation to PMS are shown in Figure 6. This illustration uses a structural equation model technique (Castor 2009 [5]). This technique seeks to provide a meaningful, evidence-based, visual data representation of the relationships between the assessment components. The structure of the CAPTEAM components is based on analysis of the data principal components and inter-correlations with PMS The positioning of the sub-components is arranged to be indicative of the progressive strength of the relationships with the intended high level effects. In this graphical representation, improved Adaptability Proficiency and raised PMS are shown to be the outputs, products, or positive effects, arising from the pluri-potential benefits of successful Collaboration in Task Work with Team Work.

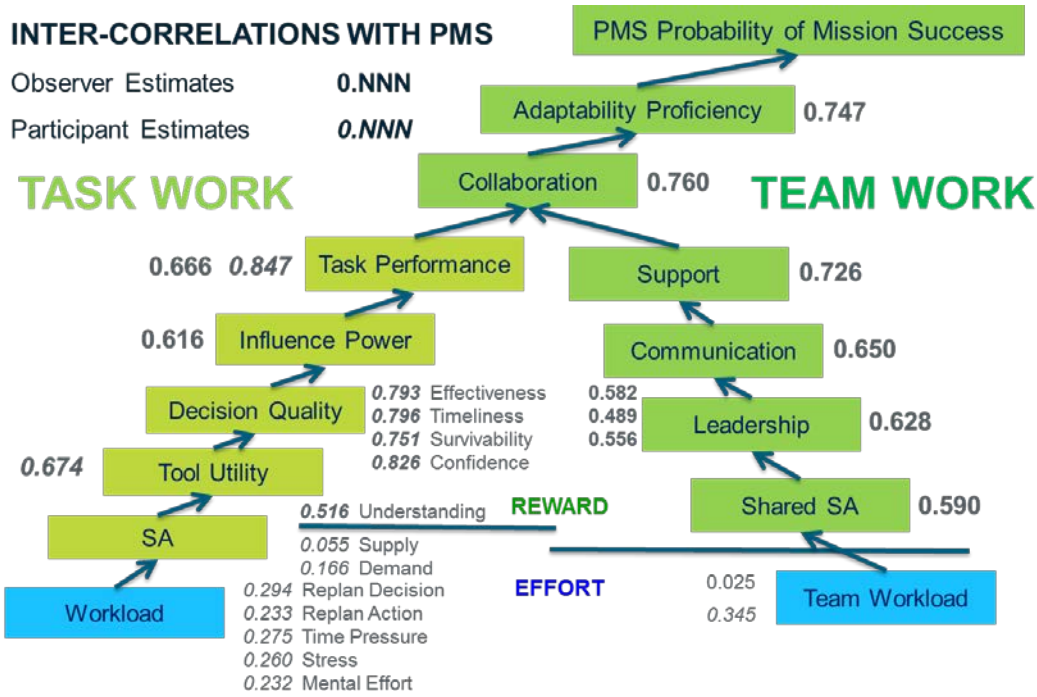


Figure 6: Inter-correlation of the Task work and Team work, Reward and Effort CAPTEAM metrics

An additional benefit from re-using MCDM/CAPTEAM methodology for Autonomous Systems Research T&E consistently over time has been to enable a set of recognised benchmarks to be established, and identified for comparative re-use, in particular for AP and PMS. These benchmarks help frame the assessment context, They act as validated human-system performance standards for comparison, gauging the relative strength of collaborative adaptability proficiency in development prototypes, and providing estimates of human, technical and autonomous system readiness (Taylor 2015 [1]).

3.3 Autonomy Task work

3.3.1 Trustworthiness

Selection of assessment components of autonomy task work seems particularly problematic since much of the detailed functioning of algorithms lacks observability at most commander/operator interfaces. This is reflected in common need for explanation and increased transparency at autonomous systems user interfaces. However, recent work under UK MoD Autonomous Systems Underpinning Research programme, SATLAOP[1] (Searle 2007 [6]) has investigated improving understanding of the requirements for trustworthiness of unmanned systems, based on Publicly Available Specification 754 Software Certification [7]. The results of this work helps guide selection of components of trustworthiness to include safety, reliability, availability, resilience and security components. With the addition of a component for dependability based on human factors literature on trust (Yagoda 2011 [8]), and using the notion of software operation predictability, defined in the sense of adherence to planning in the execution of operations, Dstl have developed and tested both a single and a multi-dimensional subjective rating scale protocol for estimating autonomy task work trustworthiness. The more mature multi-dimensional DSTL HAT Trustworthiness rating scale, with dimension definitions and anchors, is shown below in Figure 7.

HAT TRUSTWORTHINESS ASSESSMENT PROTOCOL								
Low Anchors	DIMENSION							High Anchors
	Low 1	2	3	4	5	6	High 7	
Unreliable; Uncertain; Inconsistent	**Reliability** The ability of the system to operate on missions, perform tasks and deliver effects as specified							Reliable; Certain; Consistent
Undependable; Fickle; Defective	**Dependability** The ability of the system to be relied upon to operate on missions, perform tasks and deliver effects as specified							Dependable; steady; loyal; faithful; responsible; unfailing
Unpredictable; Unlikely; Unexpected	**Predictability** The ability of the system to respond to events and to operate, perform and deliver effects consistently and reliably as planned and anticipated							Predictable; Likely; Expected
Unavailable; Unattainable; Unready	**Availability** The ability of the system to operate on missions, perform tasks, and deliver effects when requested							Available; Attainable; Unready
Brittle; Weak; Slow	**Resilience** The ability of the system to transform, renew and recover in timely response to events							Resilient; Robust; Agile
Unsafe; Unprotected; Unstable	**Safety** The ability of the system to operate without harmful states							Safe; Protected; Stable
Unsecure; Vulnerable; Risky	**Security** The ability of the system to remain protected against accidental or deliberate attacks							Secure; Protected; Assured

Figure 7: Dstl HAT Trustworthiness Protocol

The trustworthiness protocol has been used by the Autonomy Programme in various synthetic environment trials with front line military operators. Data from those trials has shown that operators and software developers are able to complete the protocol understanding the dimensions without difficulty (Taylor, Keirl, Thorpe and Grabham, 2018 [2]). The results indicate that the seven trustworthiness components potentially have useful sensitivity, discriminatory and diagnostic power in assessing HAT. Figure 8 shows the Trustworthiness Ratings obtained on the DAY17 C2 Autonomy Trail from individual software developer SMEs assessing the five individual software component capabilities contributing to the operator UxV GCS system, with comparison WITS Trial Min-Max benchmarks. Data from those trials also indicate that the trustworthiness assessments are not solely dependent on training on a system but on the usability and TTPs associated with the system use.

[1] Self-Aware Trustworthiness Levels and Assurance with Operational Policy

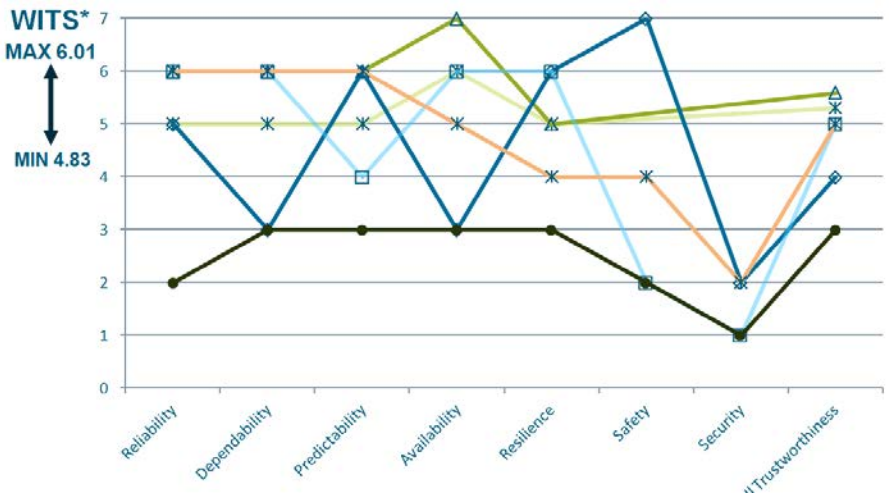


Figure 8: Results of Dstl Trustworthiness Ratings for 5 GCS Software Tools Components from DAY17 C2 Autonomy SE Trial with WITS Trial Comparison Benchmarks

3.4 Human Autonomy Teaming System (Whole System)

3.4.1 Capability Maturity Model

To help guide UK MoD trials in the assessments of autonomy teaming dimensions (human-autonomy and autonomy-autonomy), Dstl have proposed, developed and tested a model with metrics for assessing HAT capability maturity (Taylor, Keirl, Thorpe and Grabham, 2018 [2]). The Dstl HAT Capability Maturity Model (HAT CMM) approach is based on the 1980s software CMM work, originating from Carnegie Mellon University(Carnegie Mellon, 2002 [9]), with consideration of properties of later CMM derivatives, including more people-based dimensions (Systems Engineering, Usability, Organisation, People, Team Process). The underpinning CMMs for Dstl HAT CMM are derived from the basic five software CMM levels (initial through to optimising) with two additional extending, relevant levels, conveniently providing compatibility for integration with the seven level CAPTEAM metrics. The levels progress from Level 0-Not Performed to Level 6-Adaptive, passing through Level 1-Initial, Level 2-Recognised, Level 3-Defined, Level 4-Managed, and Level 5-Optimising. The seven levels of Dstl HAT CMM are each discretely calibrated and defined, with differentiation in terms of goals, processes, and behaviours.

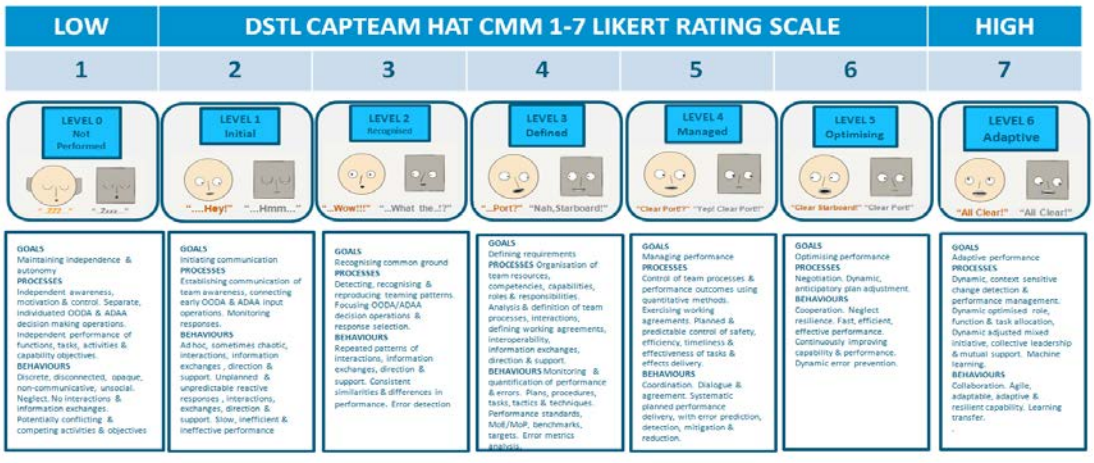


Figure 9: Dstl CAPTEAM HAT Capability Maturity Model Rating Scale

The Dstl HAT CMM Protocol has been applied successfully in recent Dstl Autonomy Research Programme SE and LVC trials. The protocol was used to assess the maturity of UxV GCS technology concepts and systems for assured C2 of Autonomy with military operators. It was found that the maturity assessment concept, including the dimensionality, calibration and scoring method, are novel, complex and needing familiarity and training to understand. Experience and training in relevant and representative military operations and in the proposed technical system use are essential to achieved stability and reliability in HAT maturity identification and classification. Notwithstanding, the capability maturity approach has been found to have both usability and utility for HAT C2 assessment purposes with military operator participant testing in both the DAY 17 SE and CB17Air LVC C2 of Autonomy Trials, The data obtained has indicated that Dstl HAT CMM has the potential to provide both sensitivity and discrimination power, with diagnostic utility, in relation to multiple dimensions of systems readiness and maturity, The scope of potential applications ranges across a diversity of relevant trial and system quality subcomponents, including the maturity of the military operations tested and the maturity of the system mission enabling autonomy technologies.

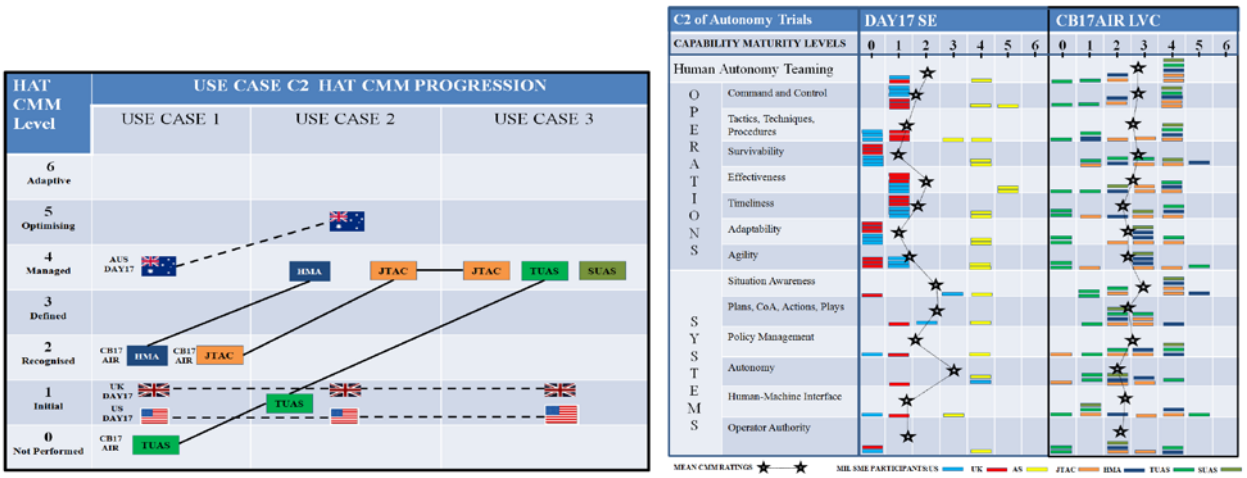


Figure 10: Results of Dstl HAT CMM from DAY17 SE and CB17Air UAS LVC C2 of Autonomy Trials.

3.4.2 Risk Assessment

In Synthetic Environment (SE) and Live, Virtual Constructive (LVC) Trials of Mission Enabling Technologies, assessment methods and protocols typically focus on operator and system performance and effectiveness, necessarily set within the conditions, assumptions, constraints and affordances of the SE/LVC operational trials design environment. To understand the Human Autonomy Teaming system from a broader perspective and identify wider potential requirements and issues, including Defence Procurement Lines of Development (DLODs) that might otherwise be missed, a Dstl Risk Assessment protocol was developed for the WITS[2] trial in March 2017 [10]. The Dstl WITS Risk Assessment protocol was an adaptation of the risk matrix approach used widely in hazard and safety risk assessment. The risk matrix provides identification of calibrated levels of risk probability and impact severity. The Dstl WITS Risk Assessment protocol modification sought to use this framework to address teaming risks for missions. The specific aim of this Dstl Mission Teaming Risk Assessment protocol was to identify the risks of teamwork and the implemented enabling technologies with the focus on the effects on mission success. It is necessary to note, that the capturing of mission teaming risks is focused away from specific platform risks which are typically used in risk based safety assessments of aircraft systems.

[2] Wildcat ISTAR Teaming for Strike, trial in March 2017 was designed to assess the benefits of teaming with potential future platform upgrades

Teaming Risk Assessment Protocol

TEAMING RISK					
NAME..... ROLE/POSITION.....					
DATE..... RUN NUMBER.....					
Mission Critical Decision Teaming Event:					
Description of Risk:					
Mitigation:					
PROBABILITY/ SEVERITY	Extremely Improbable	Improbable	Remote	Occasional	Frequent
Catastrophic					
Severe					
Major					
Minor					
Negligible					

Teaming Risk – Probability and Severity

RISK PROBABILITY OF OCCURRENCE					
Classification	Frequent	Occasional	Remote	Improbable	Extremely Improbable
Quantitative Definition	Likely to occur many times during mission	Likely to occur sometime during mission	Unlikely but possible to occur during mission	Very unlikely to occur during mission	Almost inconceivable to occur during mission

RISK CONSEQUENCE SEVERITY					
Classification	Catastrophic	Severe	Major	Minor	Negligible
Definition	Total failure to achieve mission objectives	Severe reduction in achievement of mission objectives	Major reduction in achievement of mission objectives	Minor reduction in achievement of mission objectives	No effect on mission

Teaming Risk Acceptability Matrix

RISK ACCEPTABILITY					
PROBABILITY/ SEVERITY	Extremely Improbable	Improbable	Remote	Occasional	Frequent
Catastrophic	Review	Unacceptable	Unacceptable	Unacceptable	Unacceptable
Severe	Review	Review	Unacceptable	Unacceptable	Unacceptable
Major	Acceptable	Review	Review	Review	Unacceptable
Minor	Acceptable	Acceptable	Acceptable	Acceptable	Review
Negligible	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable

Figure 11: Dstl Mission Teaming Risk Assessment – Protocol, Definitions and Acceptability Matrix

Risk is defined as a combination of the probability of occurrence and the severity of that occurrence if it should happen. A qualitative probability of occurrence between frequent and extremely improbably and a severity between Negligible and Catastrophic are assigned to each identified risk. The combination of probability and severity results in a matrix of combinations, with catastrophic/frequent the worst outcome and negligible/extremely improbable the best outcome.

The military SMEs (both participants and observers) are required to identify possible risks either that occurred during the scenario or that they could identify as possibly happening as a result of the implemented concepts during the trial. The implemented concepts could either be team orientated, or system/ technology orientated. As a part of the Protocol, the SMEs are able to suggest potential mitigations that may reduce the risk.

Experience with the application of the Dstl Mission Teaming Risk Assessment protocol initially on the WITS trial, and in its development and application subsequently for other Dstl Autonomy SE Research trials, indicate that the protocol purpose is familiar and popular with military operators, it provides a relevant, focusing, useable and useful facility, and a most welcome addition to the Dstl HAT assessment toolkit.

4.0 HUMAN AUTONOMY TEAM WORK - REMEDE

4.1 Introduction

Early Human information processing and team work models were developed in the 1940's, which were expanded upon in the proceeding decades. In the late 1980's MM Taylor developed communication layered protocols which introduced the basis protocol loop. Control abstraction layers were developed in the late 1980's and introduced to support systems for effective C2. All of these approaches model different elements

of human autonomy team work. It is necessary to develop a hybrid model that takes elements from each of the different models to create an assessment methodology that accurately covers Human Autonomy Team work. Each of the different methods will be described in detail, before detailing the hybrid approach that has been developed.

4.1.1 Human Information Processing Phases

In literature on cognitive systems engineering, the functioning of automation is commonly decomposed into a series of four sequential phases of machine information processing, namely Information Aquisition>Information Analysis>Decision Selection >Action Implementation, or AADA (Parasuraman et al, 2000 [11]). This widely used AADA framework mirrors the structure of the simple and popular human decision making OODA loop model (Observe>Orient>Decide>Act) based on understanding of pilot decision making in air combat (Boyd, 1986 [12]; Boyd 2001[13]). The OODA framework, and by implication AADA, has been criticised as a critically deficient over-simplification for conceptualisation of C2, when judged in terms of understanding of control systems principles used in cybernetics (Brehmer, 2005 [14]). This is because the OODA/AADA frameworks fail to identify the functions necessary for effective C2 dynamic decision making, A more complete and efficacious dynamic decision loop, or DOODA loop, includes functions identifying what needs to be accomplished (e.g. information collection, sense-making, and planning), and in particular the provision of feedback on products. Feedback on products is essential for dynamic control and effective C2. The key role in C2 of feedback on performance of critical functions - the products of effective functioning - necessitates representation of the Effects of actions, or Results, in models of C2 dynamic decision making e.g. Information> Decision>Action>Results/Effects.

In CAPTEAM, the Task Work model (Workload>SA>Decision Quality>Task Performance) substantially reflects the OODA Loop structure, with raised AP and PMS as high level Results/Effects, augmented by Teamwork Collaboration. PMS can be further usefully decomposed into attributes of Offensive and Defensive Performance (Castor 2009 [5]). In MOD Air Systems applied research on Mission Enabling Technologies, improved Survivability, Effectiveness and Timeliness (SET) has been long regarded as key performance trade attributes and evaluation assessment criteria for Decision Support Systems (DSS) for combat decision making, It is noteworthy that SET closely mirrors the Cost/Quality/Delivery tradespace commonly used elsewhere in the design of business services. In CAPTEAM, SET are identified as key performance assessment attributes of the Decision Making Task Work function.

4.1.2 Communication Dialogue Protocols

The CAPTEAM Team Work model, based on core attributes of Crew Resource Management (CRM) training, identifies Communication as a key factor in effective Team Work, enabling Shared SA and Leadership Support, and critically underpinning good Coordination, Cooperation and Collaboration. Cybernetics research indicates that communication in natural language, and for effective C2, involves multi-modal dialogue loops between actors with layered protocols for primary messaging and feedback (Taylor, M.M [15]). This dialogue protocol model, with feedback on functioning as a core attribute, applies equally to the design and conduct of effective human-computer or human-autonomy interaction in effective C2 operations.

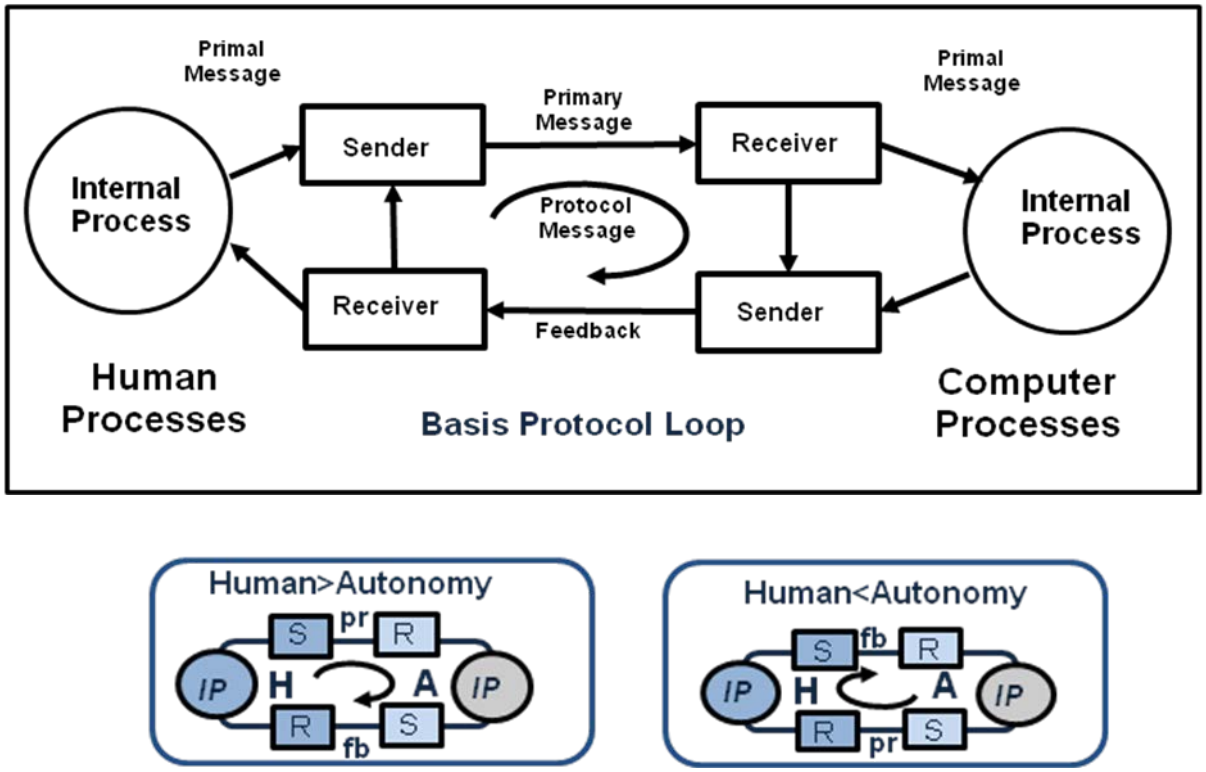


Figure 12: Communication Dialogue Protocols

4.1.3 Control Abstraction Layers

Cognitive systems engineering has shown that systems for effective C2 involve control abstraction layers, modelled largely on the Skills, Rules, Knowledge (SRK), goals-means-acts, hierarchical framework for human behaviour control (Rassmussen [16][17])

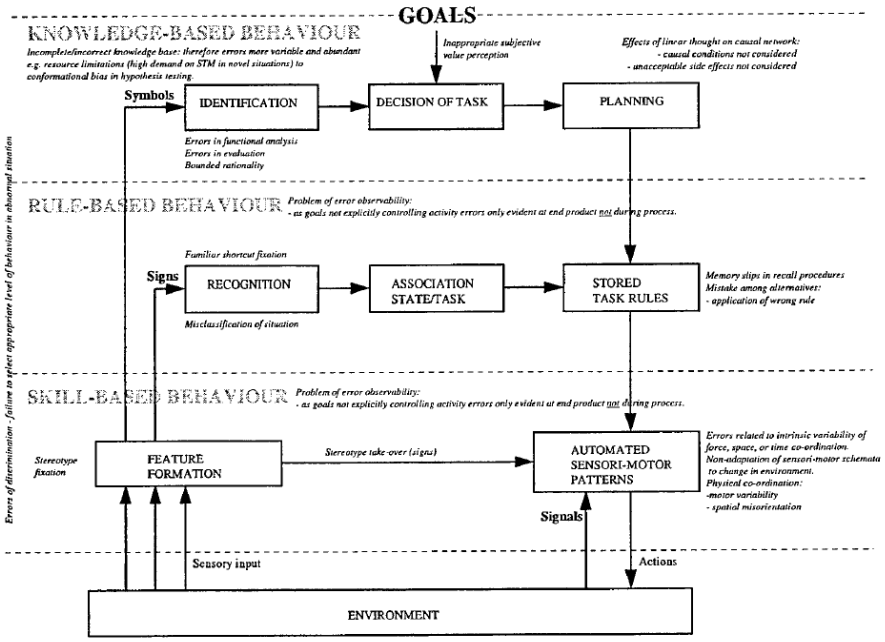


Figure 13: Hierarchical framework for Human behavior control

Subsequent research on Cognitive Work Analysis led to the proposal of a Decision Ladder model with layered components spanning Situation Analysis, Value Judgement, and Planning and Execution: (Observation>Identification>Evaluation > Decision >Planning >Execution) (Vincente, 1999 [18]). The Decision Ladder approach has subsequently been used to represent the collaborative interactions between pairs of actors/agents (Sanderson [19]). This has included control task analysis of user interactions involving PACT levels of automation/autonomy, performed under the DERA Cognitive Cockpit research programme (Taylor, [20], [21], and [22])

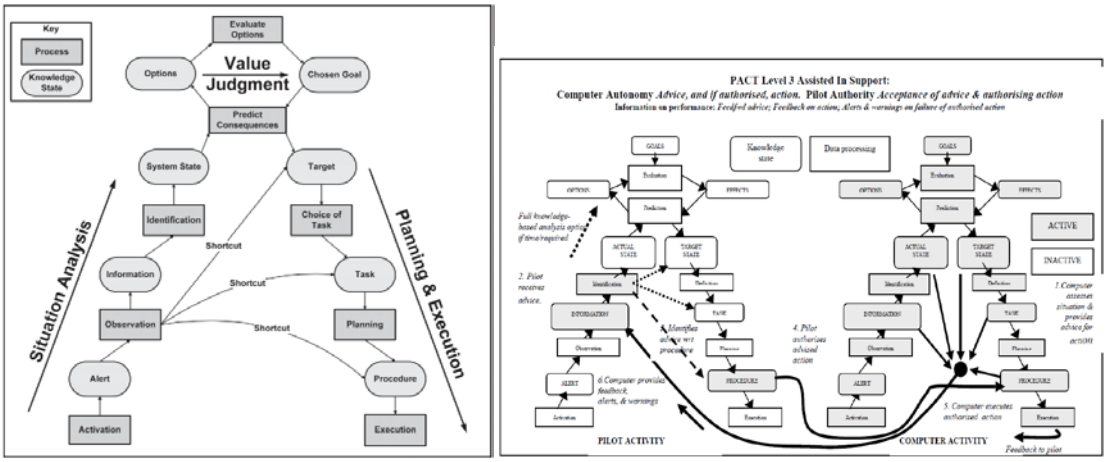


Figure 14: Decision Ladder model and Control Task Analysis of User Interactions

Research on Joint Cognitive Systems (Hollnagel and Woods, 1983 [23], Hollnagel, 1993 [24]) is particularly pertinent to Human-Autonomy Teaming (Taylor 2002 [25]). A multi-layered cognitive control model has been proposed for representation of the control of multiple UxVs (Hollnagel, 2007 [26]). Synchronised,

interacting layers of control identified include Targeting (Governing) > Monitoring (Directing) > Regulating> Controlling (Operating/Tracking).

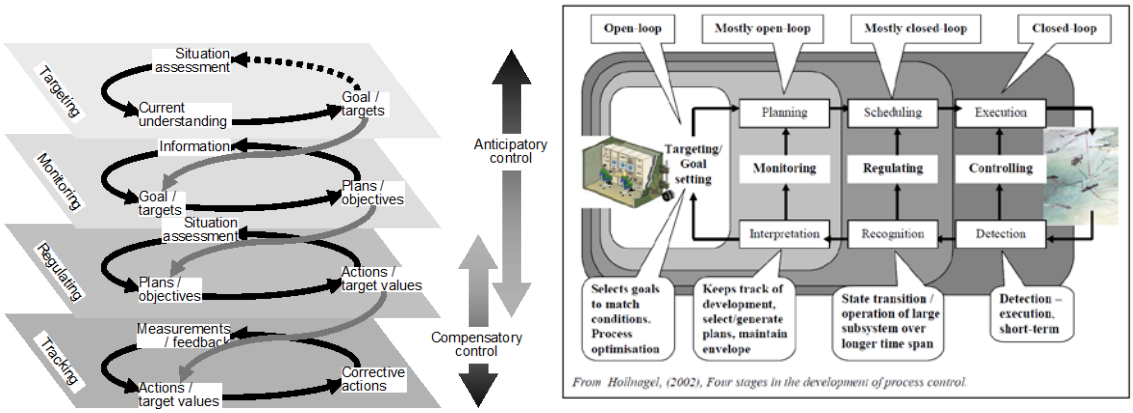


Figure 15: Multi-layered cognitive control model

Informed by this work on cognitive systems engineering and on the concepts of joint cognitive systems, cognitive work analysis and decision ladder control task analysis, work under the Dstl DAMM project identified the applicability of the REMDAER model (Recognise > Evaluate > Mitigate > Disseminate > Acknowledge > Decide > Execute > Report) for representing C2 decision making in a distributed, highly networked military operating environment (Taylor and Grabham 2012 [3]). This framework provided the components for multi-player, distributed or team, decision making cycle, within the operational and tactical C2 architecture C2 OODA or "COODA" layered control system and is illustrated in Figure 16.

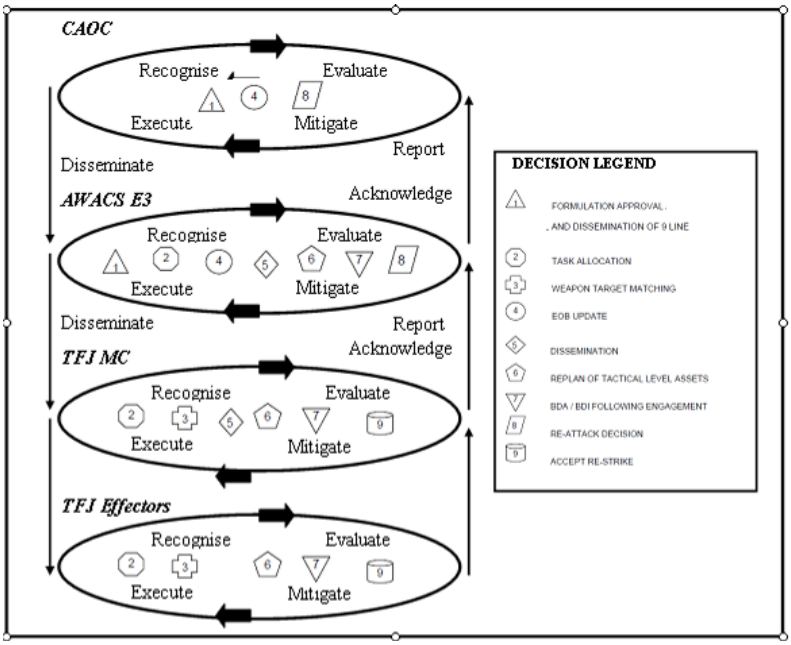


Figure 16: REMDAER Decision Model

4.2 Hybrid Approach

The REMDAER model for the layered control system for airborne mission management was used as an initial step to develop an assessment Protocol for HAT effectiveness. By acknowledging that HAT effectiveness is critically dependent on team member inter-communication, in a similar way to the COODA layered control system, the REMDAER elements were broken down into those that the human and autonomy can complete and those that are dependent on communication between the two. Both humans and autonomy can Recognise, Evaluate, Mitigate and Execute to cause Effects, with Disseminate, Acknowledge and Reporting dependent on the communication Dialogue between the team members.

For assessment purposes, significantly contributing components need to be characteristic and observable; measurable, sensitive and discriminating; and logical, meaningful and evidential. The focus has been on interaction essential information exchange, (and) where the key interactions happen during decision processes, and on the effectiveness of the action taken by the human or the autonomy. In particular the effectiveness of Recognising, Evaluating, Mitigating and Executing (REME) as a team within a mission, and effectiveness of the Communication Dialogue between the human and the autonomy (disseminating, acknowledging and reporting), and the consequences of HAT in terms of specific intended dynamic Effects. Specific measures have been identified to evaluate the effectiveness of the teaming framing around REMEDE.

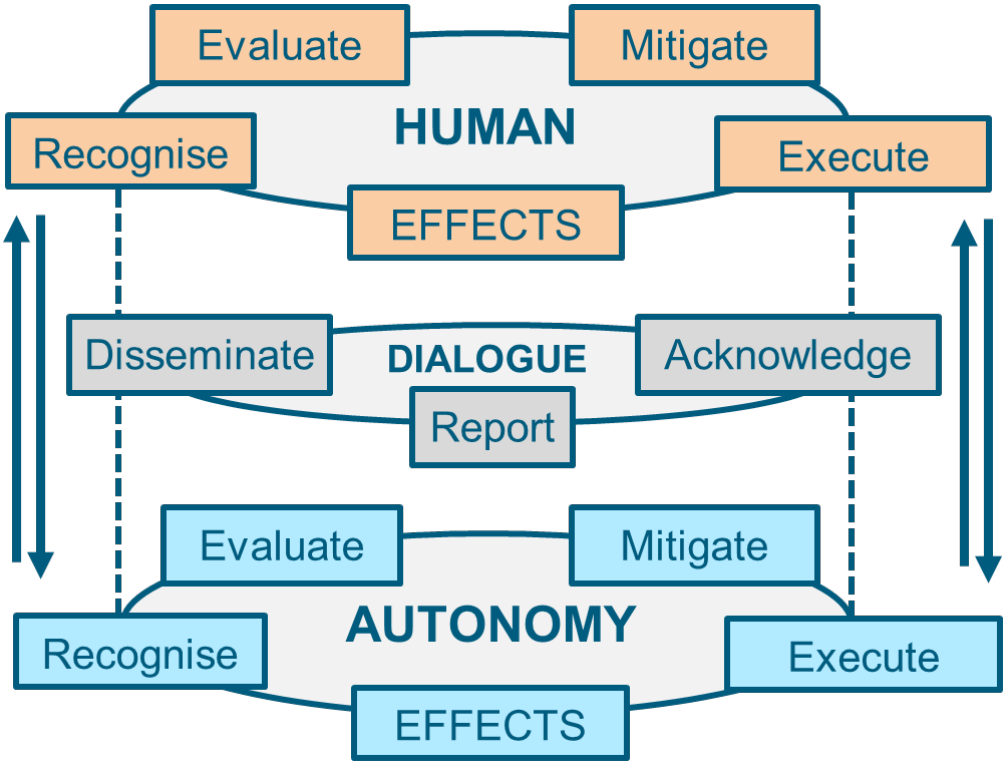


Figure 17: REMEDE Co-Dependent C2 COODA-Model

4.3 REMEDE Assessment Protocols

Seven individual assessment protocols have been proposed initially for assessing the effectiveness of HAT, for validation testing. Six of these protocols are focused on the individual components of the REMEDE model, Recognition, Evaluation, Mitigation, Execution, Dialogue and Effects, with the final protocol being a composite of the REMEDE components. This structure is similar to the Trustworthiness protocol, where the

development and availability of individual protocols and a composite protocol allows for the tailoring of the assessment to the type of experimentation that is being undertaken. Currently neither the composite nor the individual protocols has undergone validation and verification testing, but is scheduled to be utilised in the autumn 2018.

With the exception of the Dialogue protocol, which is unique amongst the different protocols, the remaining six protocols utilise a Likert rating, which ensures compatibility amongst HAT metrics including CAPTEAM, to allow SMEs (operators and/or observers) to score the human functioning component, the system autonomy component and the joint shared communication of each of the REMEDE components. Joint shared communication comprises both the human to autonomy and autonomy to human communication. The Dialogue protocol offers additional granularity for each of the composite REMEDE components in relation to human to autonomy communication and autonomy to human communication, with an individual rating available for each.

EFFECTS	Dstl REMEDE Human Autonomy Teaming Protocol - Effects																				
Impact of actual effects on achievement of Mission and Task Objectives.	Human Functioning Contribution							Joint Shared Communication *H>A & A>H*							System Autonomy Contribution						
	Impact							Impact							Impact						
	Low 1	2	3	4	5	6	High 7	Low 1	2	3	4	5	6	High 7	Low 1	2	3	4	5	6	High 7
Survivability																					
Effectiveness																					
Timeliness																					
Agility																					
Adaptability																					
Offensive Performance																					
Defensive Performance																					
Probability of Mission Success																					
EFFECTS COMPOSITE																					
EFFECTS COMMENTS																					

Figure 18: Dstl HAT REMEDE Effects Component Protocol

5.0 REMEDE VALIDATION AND VERIFICATION TESTING

The REMEDE HAT Assessment Protocol has not yet undergone validation and verification testing However, REMEDE has been adapted for, and will be used in the upcoming STRATUS[3] trials in September 2018 assessing human and system contribution and communication as well as an additional feature asking operators to assess the layered ISTAR contribution towards the REMEDE components. An example assessment protocol from the Stratus evaluation is shown below in Figure 19.

[3] The STRATUS project is assessing the benefits of a layered Intelligence, Surveillance, Target Acquisition, and Reconnaissance (ISTAR) approach in an urban environment with novel autonomy and decision making tools and technologies.

EFFECTS	Dstl REMEDE Human Autonomy Teaming Protocol - Effects																												
Impact of actual effects on achievement of Mission and Task Objectives.	Human Functioning Contribution							Joint Shared Communication							System Autonomy Contribution							Layered C2 ISTAR Collaboration						
	Impact							Impact							Impact							Impact						
	Low 1	2	3	4	5	6	High 7	Low 1	2	3	4	5	6	High 7	Low 1	2	3	4	5	6	High 7	Low 1	2	3	4	5	6	High 7
Raised Understanding																												
Reach Area Depth																												
SA for Specific Effects																												
Survivability																												
Effectiveness																												
Timeliness																												
Agility																												
Adaptability																												
Offensive Performance																												
Defensive Performance																												
Probability of Mission Success																												
EFFECTS COMPOSITE																												
Comments																												

Figure 19: Dstl HAT REMEDE Effects Stratus Specific Component Protocol

6.0 CONCLUSIONS

The assessment of HAT efficacy, while technically challenging can be broken down into a number of individual components of the Human Autonomy Teaming System to simplify the challenge. The understanding of the various components, Human Task work and Teamwork, Autonomy Task work and Teamwork and Human Autonomy Teaming, of the HAT system has been built up over more than 18 years of UK MoD Dstl R&D on Autonomy and Mission Systems. While it is possible, and in some cases easier to understand the HAT system in its entirety, it is often not possible to identify the locus of success or failure at such a high level depending on the system that is being tested.

The Human task work and teamwork elements of the HAT system, through CAPTEAM which is designed to estimate mission efficiency by the metrication of Reward and Effort associated with critical mission events and decision processes. CAPTEAM has been extensively utilised for trials within the UK research programme, with statistical analysis indicating the inter-correlation of the assessment components with relatively beneficial Reward factors, such as improved SA and Decision Quality, were significantly highly correlated with PMS. In comparison, Effort factors and sub-components, such as Workload Stress and Mental Effort a exhibited relatively low, non-significant correlations with PMS.

The Autonomy Task Work component of the HAT system is addressed through a broken down multi-dimensional Trustworthiness scale. While only used in a limited number of synthetic trials, the results indicate that the seven trustworthiness components potentially have useful sensitivity, discriminatory and diagnostic power in assessing HAT, with data from the trials also indicating that trustworthiness is not solely dependent in training on a system but on the usability and TTPs that are associated with the system in use.

The entirety of the HAT system can be assessed through a combination of the HAT Capability Maturity Model, which was derived and developed from software CMMs developed by Carnegie Mellon University and the Risk Assessment. The HAT CMM has been used successfully in recent Dstl Research Programme SE and LVC trials, assessing UxV technology concepts and systems for assured C2 of Autonomy with military operators. It was found that the concept of the maturity assessment, are novel, complex and needing familiarity and training to understand. With experience and training in relevant and representative military

operations and in the proposed technical system use are essential to achieve stability and reliability in HAT maturity identification and classification. Notwithstanding, the approach to capability maturity has been found to have both usability and utility for HAT C2 assessment purposes, with the obtained data indicating that the Dstl HAT CMM has the potential to provide both sensitivity and discrimination power. The Risk Assessment protocol has been used in both synthetic and LVC trials recently, and provides a broader perspective of the HAT system and allows for the potential identification of wider requirements and issues, including those associated with the Defence Procurement Lines of Development that might otherwise be missed. The focus of risks is away from the traditional platform safety risks that are typically addressed in risk based safety assessments of aircraft systems, but instead focused on the risks associated with teamwork and the implemented enabling technologies with respect to mission success.

The new component of the UK HAT System assessment methodologies is the REMEDE assessment protocol. The protocol itself is focused on the Human Autonomy Teamwork component of the entire system, and has been developed from a variety of information processing, communication and team work models. Previously these had been combined to form the REMDAER decision model, as part of previously reported work for the DAMM project, which was focused on a multi-player, distributed or team, decisions making cycle within the operational and tactical C2 architecture "COODA" layered control system. With the critical dependency of team member inter-communication being similar across the HAT system and the REMDAER layered control system it is possible to break down the REMDAER elements into what the Autonomy can complete and what the Human can complete and what dialogue is required between them to be effective. This results in the components of Recognise, Evaluate, Mitigate, Execute to cause effects being applicable to the Autonomy and the Human, with Dissemination and Reporting and Acknowledging forming the dialogue between them. The application of this to the assessment of HAT has been proposed in two different ways, from a high level composite protocol that only looks at high level of the REMEDE components, to an individual protocol for each of REMEDE components. This breakdown of the individual components allows for an increased granularity of the underlying aspects of the different components. At this stage, the REMEDE protocol has been selected for use with the STRATUS project under the UK MoD Dstl Autonomy Research programme, with a specific instantiation developed for the specifics of the trial. This trial will serve as the initial opportunity to undertake verification and validation testing of the REMEDE protocol, the outcomes of which will allow a greater understanding of the discrimination and sensitivity of the data captured.

Autonomy Teamwork has not been addressed within the UK research programme at this time and remains an area of interest in the context of Human Autonomy Teaming and the contribution that Autonomy Teaming has to the understanding the efficacy of the entirety of the HAT System.

7.0 REFERENCES

[1] Taylor R.M. 2015. Human Autonomy Teaming: Verification, Benchmarking and Readiness. In, RPAS Achievements and Challenges. Proceedings of the 2015 Presidents Conference, Royal Aeronautical Society, London' 7-8 October 2015. SG UAS Event Code 772. ID P2PP2R-2015-09-28T 10:34:36. DSTL/TR91332 967705572. ISBN 1 8576831X.

[2] Taylor, R.M., Keirl, H., Thorpe E. and Grabham A. 2018. UK-1: HAT-CV&V Human Autonomy Teaming – Collaboration Verification and Validation. Final Report NATO HFM247. November 2017.

[3] Taylor R.M. and Grabham A. 2012. UK-1: Dynamic Airborne Mission Management. In, Supervisory Control of Multiple Uninhabited Systems – Methodologies and Enabling Human-Robot Interface Technologies. Chapter 13, pp 1-27. Lessons Learnt. Annex A, pp 1-14. HFM-TG-170. AC/323(HFM-170)TP/451. December 2012

[4] Cottrell, R.J. 2011. RE314 July 2011 Strike Warrior III Joint Experimental Trial Report. QinetiQ/AEG/TI/CR1102929/V0.1. 14 Dec 2011. Cody Technology Park, QinetiQ Ltd., Farnborough, UK.

[5] Castor, M. 2009. The use of structural equation modelling to describe the effect of operator functional state on air-to-air engagement outcomes. Linkoping Studies in Science and Technology. Dissertation No. 1251. Linkoping University Institute of Technology. Linkoping, Sweden. ISBN 978-91-7393-657-6 ISSN 0345-7524.

[6] Searle T. 2017. ASUR Self-Aware Trustworthiness Levels and Assurance with Operation Policy 1014_C3_Ph2_060/073/110, 49048-45595R Issue 2, Frazer-Nash Consultancy Ltd.

[7] British Standards Institution, PAS 754:2014, 2014.

[8] Yagoda, R.E. 2011. WHAT! You want me to trust a ROBOT? The development of a human robot interaction (HRI) trust scale. MSc. Psychology Graduate Thesis. North Carolina State University, Raleigh, North Carolina 2011.

[9] Carnegie Mellon. 2002. Capability Maturity Model® Integration (CMMISM), Version 1.1. CMMISM for Software Engineering (CMMI-SW, V1.1). Continuous Representation. CMU/SEI-2002-TR-028. ESC-TR-2002-028. CMMI Product Team, August 2002. Carnegie Mellon Software Engineering Institute, Pittsburgh, PA 15213-3890

[10] Keirl, H, Hennessy, DKR, Thorpe, A, Taylor, R, Halpin, D. 2017. Wildcat ISTAR Teaming For Strike – Phase 1 Experimentation. DSTL/CR101870 Issue 1. March 2017. Defence Science and Technology Laboratory, UK MoD.

[11] Parasuraman, R., Sheridan, T. and Wickens, C. 2000. A Model for Types and Levels of Human Interaction with Automation, IEEE Transactions on Systems, Man, and Cybernetics. – Part A: Systems. and Humans, Vol. 30, pp. 286-297.

[12] Boyd, J.R. 1986. Organic Design for Command and Control. Briefing slides, October 1986

[13] Boyd, J.R. 2001. “An essay on winning and losing”. http://defence and the national interest. d-n-i-net, 2001

[14] Brehmer, B. 2005. One Loop to Rule Them All. In Proceedings of 11TH ICCRTS Coalition Command and Control in the Networked Area: C2 Analysis, C2 Modelling and Simulation, Cognitive Domain Issues. Department of War Studies, Swedish National Defence College, Stockholm, Sweden.

[15] Taylor M.M. 1989. Response Timing in Layered Protocols: A Cybernetic View of Natural Language. In, The Structure of Multi-Modal Dialogue. M.M. Taylor, F. Neel, and D.G.Bouwhuis (Eds). Chapter 11, pp 159-172 Amsterdam: Elselvier Sciences B.V. North Holland.

[16] Rasmussen J. 1986. Information Processing and Human- Machine Interaction: An Approach to Cognitive Engineering. New York: North Holland.

[17] Rasmussen, J., Pejtersen, A.M., and Goodstein, L.P. 1994. Cognitive Engineering: Concepts and Applications. New York: Wiley.

[18] Vicente, K.J. 1999. Cognitive Work Analysis: Towards Safe, Productive, and Healthy Computer based

Work. New Jersey: Lawrence Erlbaum.

[19] Sanderson, P., Naikar, N., Lintern, G., and Goss, S. 1999. Use of cognitive work analysis across the system life cycle: From requirements to decommissioning. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, Houston, TX. Santa Monica: HFES

[20] Taylor, R.M. 2001. Cognitive Cockpit Control Task Analysis. DERA Memo, DERA/CHS3/6.3/14/7, 07 March 2001.

[21] Taylor R.M 2002. Technologies for Supporting Human Cognitive Control. In Human Factors in the 21st Century. NATO MP-077. 18:1-14. RTO Human Factors and Medicine Panel (HFM) Specialists' Meeting, Paris, France, 11-13 June 2001.

[22] Taylor, R.M 2007. Human Automation Integration with Contractual Autonomy. In, Uninhabited Military Vehicles (UMVs): Human Factors in Augmenting the Force. Final Report of RTO Human Factors and Medicine (HFM) Task Group HFM-078/TG017. Chapter 7, Human Automation Integration, pp 3-12 to 29. AC/323(HFM-078)TP69. ISBN 978-92-837-0060-9 July 2007.

[23] Hollnagel, E. and Woods, D. D. 1983. Cognitive Systems Engineering: New wine in new bottles, International Journal of Man-Machine Studies, 18, pps

[24] Hollnagel E. 1993. Human Reliability Analysis: Context and Control. London: Academic Press. 583-600.

[25] Taylor, R.M. 2002. Capability, Cognition and Autonomy. RTO HFM Symposium on "The Role of Humans in Intelligent and Automated Systems", Warsaw, Poland, 7-9 October 2002, NATO RTO-MP-088. KN3, 1 to 26.

[26] Hollnagel, E. 2007. Modelling Multi Layered Control: Application of the Extended Control Model to the Analysis of UAV Scenarios. In, Uninhabited Military Vehicles (UMVs): Human Factors in Augmenting the Force. Final Report of RTO Human Factors and Medicine (HFM) Task Group HFM-078/TG017. Chapter 7.7, Human Automation Integration, pp 73 to 96. AC/323(HFM-078)TP69. ISBN 978-92-837-0060-9. July 2007.

Understanding Human-Autonomy Teaming Through Interdependence Analysis

Matthew Johnson, Micael Vignati and Daniel Duran
40 S. Alcaniz Street
Pensacola, FL 32502
USA

mjohnson@ihmc.us, mvignati@ihmc.us, dduran@ihmc.us

ABSTRACT

As any craftsman knows, having the right tool for the job makes all the difference. Sadly, adequate formative tools are lacking for the design of human-autonomy teaming. Most existing tools focus on the taskwork, in particular physical activity. More advanced efforts extend this with consideration for cognitive activity. While these traditional task-based approaches are important, they do not sufficiently capture human-machine teaming requirements. The key to understanding human-autonomy teaming, and in fact teaming in general, is understanding interdependence. This paper proposes several design principles to help designers reorient their minds and view the problem space through the lens of interdependence. The paper then discusses a design and analysis tool called the Interdependence Analysis tool. This tool's purpose is to understand how people and automation can effectively team by providing insight into the interdependence relationships used to support one another throughout an activity. This tool helps recognize both the human factors and the technological factors that enhance or inhibit effective teaming. As such, it can be used to derive both algorithmic and interface implementation requirements. The goal of interdependence analysis is improving the guidance provided to designers building human–machine systems in order to achieve more effective teaming. As we build more intelligent machines to tackle more sophisticated challenges, designers need formative tools for creating effective teaming. The Interdependence Analysis tool meets this need, explicitly modelling the machine, the human, the work and the interplay of all three.

1.0 MOTIVATION

As any craftsman knows, having the right tool for the job makes all the difference. Sadly, adequate formative tools are lacking for the design of human-autonomy teaming. The vast majority of formative design tools are technology-centric in which the human is not even considered part of the system (e.g. MATLAB, LabVIEW). Human factors research has highlighted the need for consideration of the human, but this is often accomplished using summative assessments to evaluate existing systems at the end of development (e.g. NASA TLX). User-centered design has pushed for such evaluations to happen earlier in the development process. Apple, for example, has a reputation for early user evaluations translating into successful products. However, this remains the exception rather than the norm for most system development efforts. Moreover, while user evaluations are undoubtedly valuable, designers need to be able to account for such considerations long before there is a system to evaluate. To assist designers in accomplishing this, they need formative tools available in the design process that include the human as part of the system from the beginning.

Another major impediment to designing human-autonomy teaming is that most existing tools focus on the taskwork, in particular the physical activity. Taskwork is only a small piece of the teaming puzzle. More advanced efforts from cognitive systems engineers extend this with consideration for cognitive activity (e.g. [1]). This is a valuable contribution, but still tends to be taskwork focused. While these traditional task-based approaches are important, they do not sufficiently capture human-machine teaming requirements. Moreover, the language, concepts, and products of those who focus on cognitive aspects are often far removed from those who design and implement working systems and do not translate "into a language that helps a product

be built or coded" [2].

The key to understanding human-autonomy teaming, and in fact teaming in general, is understanding interdependence. Teaming, both human-human and human-machine, comes in many forms with varying characteristics and properties. The one concept that is consistent in every case is the importance of interdependence. This truth is invariant across all domains. This paper begins with an explanation of interdependence. To help designers reorient their minds and view the problem space through the lens of interdependence, we propose several design principles. These principles are intended to help designers shift their perspective and reframe the design challenges to focus on the core elements of teaming, specifically the interdependence needed to support it. Lastly, we discuss a formative tool called the Interdependence Analysis tool.

2.0 UNDERSTANDING INTERDEPENDENCE

In order to analyze a human-machine system's interdependence, it is necessary to understand the concept of interdependence. It is often simply equated to dependence, where one entity relies on another because it lacks some capability provided by the other. However, this definition of the concept is too simplistic to capture the nuances observed in interdependence relationships among teams engaged in joint activity. Our definition of interdependence is as follows:

> *"'Interdependence' describes the set of complementary relationships that two or more parties rely on to manage required (hard) or opportunistic (soft) dependencies in joint activity"* [3].

Interdependence is about relationships, not taskwork. It is true that the taskwork influences the potential relationships, however, designing human-machine teaming is about designing the interdependence relationships not the taskwork functions. These relationships are used to manage the interdependence in the joint activity. Relationships can be required, but a significant portion of teaming is about exploiting opportunistic relationships.

To better intuit the concept of interdependence, consider an example of playing the same sheet of music as a solo versus a duet. Although the music is the same, the processes involved are very different [4]. The difference is that the process of a duet requires ways to support the interdependence among the players. Understanding the nature of the interdependencies among team members provides insight into the kinds of coordination or teaming that will be required, such as the management of timing, tempo, and volume. Supporting these relationships through agreed signaling and exploiting these interdependence relationships at run-time is what teaming is all about. Success in a duet requires not only execution of the musical score (i.e. individual taskwork competency), but also the extra work of coordinating with others via interdependence relationships to produce effective performance. Flawless execution of the musical score individually is not enough for the duet to be successful, nor is exceptional individual competence sufficient for effective teaming.

3.0 INTERDEPENDENCE DESIGN PRINCIPLES

While having the right tools is valuable, having the right mindset or perspective on the problem is also essential to effectively leveraging the tools. This turns out to be particularly challenging for human-autonomy teaming largely because of the pervasiveness of today's function allocation-based paradigms [5]. In general, such approaches date back to Sheridan and Verplank's work on supervisory control [6]. Follow up research on dynamic and adaptive function allocation has led to numerous proposals for dynamic adjustment of autonomy. Such approaches have been variously called adjustable autonomy, dynamic task allocation, sliding autonomy, flexible autonomy, and adaptive automation. In each case, the system must decide at runtime which functions to automate and to what level of autonomy [7], thus we refer to them as

function allocation-based approaches. A restrictive outcome of such approaches is that automation choices dictate interaction possibilities.

Our approach inverts this paradigm, suggesting that desired teaming interactions should shape the automation design [8]. Thus, our focus is on the interdependence relationships that enable interaction. Focusing on relationships instead of functions is challenging for most people raised under the traditional task decomposition paradigm. Task decomposition is the staple approach to tackling hard problems by breaking them down into smaller ones. Understanding interdependence requires an additional step of considering how those pieces will fit back together again when distributed between people and machines. As Peter Drucker noted long ago, "when it comes to the job itself, however, the problem is not to dissect it into parts or motions but to put together an integrated whole" [9]. While his statement was about business management, it summarizes the true challenge of human-machine teaming nicely.

To help designers reorient their minds and view the problem space through the lens of interdependence, we propose several design principles. The first principle is about the value of teaming. While there is certainly a cost to teaming, and that cost must be controlled [10], there is a value to teaming that must be balanced against the cost. Historically, traditional approaches appear almost surprised by interdependence. For example, an early task decomposition approach to planning noted that "the expansion of each node produces child nodes. Each child node contains a more detailed model of the action it represents... The individual subplan for each node will be correct, but there is as yet no guarantee that the new plan, taken as a whole, will be correct. There may be interactions between the new, detailed steps that render the overall plan invalid" [11]. Similarly, another approach noted "because the relevant constraints have been shared during the planning process, the expectation is that few, if any, conflicts will appear during plan merging. However, because of the complexity of planning dependencies, conflicts can arise" [12]. These references from early planning systems unknowingly highlighted the need to handle interdependence. Their solution was to try to supplement their task decomposition approach with capabilities such as critics used to resolve conflicts and backtracking techniques to handle cases that could not be resolved. Today's function allocation approaches are no different from these early planning approaches because they focus on what to automate and who to assign it to [7]. More recent work still notes that "'conventional wisdom' is often an over-simplification, and will be modified and sometimes reversed by a host of contextual factors" [13]. The main problem with traditional approaches was noted by Stefik long ago who observed "Subproblems interact. This observation is central to problem solving", but concluded "A key step in design is to minimize the interactions between separate subsystems" [14]. While traditional approaches view interdependence as an unexpected or unfortunate problem to be resolved, avoided or minimized, we view it as the core design element. It is something to be leveraged opportunistically. The main reason is because the value of teaming comes not from avoiding, ignoring, limiting or minimizing teaming, but from exploiting it. Thus, our first principle:

> *Principle 1: The value of teaming comes from exploiting it not avoiding it.*

In order to exploit teaming, it is important to understand teaming. Historically, engineers have designed work to be done by an individual. Robotic algorithms generate actions for an individual robot. Intelligent planners, in general, generate a sequence of actions for an individual agent or distributed to several agents to be executed individually. In contrast, we propose that all behaviors should be designed from the beginning to be joint work [8]. This means that as any behavior or algorithm is created, designers should consider the potential teaming associated with the activity. If this is done, the single agent case is just a degenerate case and is achieved for free. However, the converse is not true. Thus, our second principle is:

> *Principle 2: All work should be designed as joint activity (coactive), with independent work being the degenerate case.*

The implication of this principle is that designers should be designing collaborative algorithms for distributed and decentralized systems. This is important because human-machine systems are inherently

distributed and decentralized.

In order to achieve the second principle, we need to be able to describe joint activity. An essential theoretical understanding that is absent in existing approaches are principles needed to determine how work can be performed jointly. Currently, engineers decide where to divide work into sub-tasks or actions. This is done with limited understanding of the consequences or impact of these choices, resulting in well-known issues like the substitution myth [15]. It is not just the automation, but the design of the automation that affects performance. Because of this, there is a need for understanding the work itself and the interdependence created by decomposition and distribution. Some examples of this can be found in research on group processes and productivity which provides a taxonomy of group tasks [16]. It includes knowing if a task is divisible or not, whether the task requires maximizing or optimizing, whether the task is additive, disjunctive or conjunctive just to name a few. This leads us to our third principle:

Principle 3: Any work has inherent potential and limitations on joint-ness.

This principle is about uncovering and identifying the potential interdependencies in joint work. One particularly human capability is our amazing ability to team. People are thrown into teams without training and sometimes even without domain knowledge and yet are capable of at least rudimentary teaming. Our life experiences have enabled us to understand what aspects of work require synchronization, when acknowledgement is useful, what information is relevant to share and how to request assistance. While people do make mistakes and some people are more naturally skilled than others, this generalizability across domains - regularly exhibited by people - provides the intuition that there are common patterns in joint work which support our third principle.

In order to model joint work, we must first address what counts as work. Traditional engineering and planning target physical actions that affect the world, but can ignore cognitive aspects such as sensing, perception, memory, reasoning and understanding. This oversight is demonstrated by the development of an entirely separate field known as Cognitive Engineering [17] to address issues ignored by traditional engineering. While teamwork does involve coordinating physical actions, a significant role of teamwork is coordinating cognitive activities as well [18]. Designs that do not include these aspects will not produce effective team players [10]. Thus, our fourth principle:

Principle 4: Teaming occurs not just on the physical level, but also on the cognitive level.

As we get into more specific discussions on modelling and representation, we can draw upon good design principles that cross domain boundaries. Given the complexity of teaming, our fifth principle, that is just good design practice in general and will be applied in several ways is:

Principle 5: Separate the "what" from the "how".

As an example, consider the goal of making sure a teammate does not fall into a hole while working. This could be accomplished by specifying a route for the teammate that avoids the hole, informing the teammate about the hole and letting them avoid it, positioning oneself or an object by the hole to force the teammate to avoid both the obstacle and the hole. All of these achieve the same purpose (the what) but through different means (the how).

There are many factors that go into effective teaming. These include the work itself, the team members involved in the work and the environmental factors in which the work is performed. The challenge is that all of these factors interact. In accordance with principle five, we propose our sixth principle:

Principle 6: Joint work can and should be modelled in an agent agnostic manner.

By this we mean that the description of the work should not include considerations for specific team compositions or team member capabilities. This does not mean team composition will not be accounted for, just that the initial work description is not the appropriate place to do this. This is indeed a foreign concept in robotics, as most developers strive to customize their solutions to the specific targeted hardware, though recent efforts like ROS (Robot Operating System [19]) has shown the value of abstraction in the robotics domain. This concept is less foreign in domains such as cross platform application development. In such domains, layers of abstraction are used to enable generalization. Similarly, our principle is trying to emphasize that you may not know the characteristics of the teammates ahead of time, so designing your joint work on top of a reasonable abstraction of the work itself will allow for broader applicability and reuse.

If work should be modelled as joint work, how is this different than regular work? We propose it is the inclusion of interdependence in the model that captures the potential joint-ness. Malone and Crowston define coordination as "managing dependencies between activities" [20]. Studying human teams and team effectiveness, researchers have identified team member interdependence as a critical feature defining the essence of a team [21]. From human-machine research, Feltovich et. al. propose interdependence is the essence of joint activity [22]. Interdependence focuses on how the decomposed and potentially distributed work remains interdependent. This provides exactly what is needed to understand joint work, to understand the implications of different decomposition choices, and to specify requirements based on teaming [3]. Thus, our seventh principle:

Principle 7: Interdependence provides the basis for understanding potential joint-ness.

Some interdependencies are obvious. For example, resource constraints on a shared resource or sequencing constraints on dependent tasks. These are hard requirements that must be coordinated through teaming. As hard requirements, they are unavoidable and therefore obvious. However, a significant amount of normal human teaming involves opportunistic (i.e., soft) interdependence relationships. Soft interdependence does not stem from a hard constraint or a lack of capability. It arises from recognizing opportunities to be more effective, more efficient, and/or more robust by working jointly. Soft interdependence is less obvious because it is optional and opportunistic rather than strictly required. It includes a wide range of helpful things that a participant may do to enhance team performance. Examples include progress appraisals ("I'm running late"), warnings ("Watch your step"), helpful adjuncts ("Do you want me to pick up your prescription when I go by the drug store?"), and observations about relevant unexpected events ("It has started to rain"). Many aspects of teamwork are best described as soft interdependencies. Our observations to date suggest that good teams can often be distinguished from great ones by how well they manage soft interdependencies. Thus, our eighth principle:

Principle 8: Teaming involves both required (hard) and opportunistic (soft) interdependencies.

So how does a designer identify interdependencies, particularly the less obvious soft interdependencies? Coactive Design proposed three essential interdependence relations; observability, predictability and directability [3]. Observability means making pertinent aspects of one's status, as well as one's knowledge of the team, task, and environment observable to others. Predictability means one's actions should be predictable enough that others can reasonably rely on them when considering their own actions. Directability means one's ability to influence the behavior of others and complementarily be influenced by others. There are certainly additional types of interdependence relationships, such as explainability and trust, but we view these three as foundational to the others. This leads to principle nine:

Principle 9: Observability, Predictability and Directability are compulsory interdependencies in teamwork.

One characteristic of teaming that provides compelling value is the ability for teams to be robust to individual failures. This does not happen by accident or without any effort. Teams monitor and assess the

state of each other in order to achieve this advantage. Underlying this skill is the understanding that failure is always an option. Human failings and limitations are well known and often the motivation for more automation. Yet automation has its own failings and limitations. This is not a problem as long as the team is attentive to the potential for failure. More directly, that the human-machine team is designed to address it. Thus, principle ten:

> *Principle 10: Failure is always an option and teaming should be designed to help the team be robust to failure of both people and machines.*

This means that designers should not only be considering if performing some task is possible, but should be considering the risks and frailties of a task with respect to all team members. This provides insight into the importance and potential necessity of different teaming options in order to prevent any single team member from being a critical point of failure. Supporting this principle involves designing appropriate observability and predictability relationships to support monitoring, backup behaviors and other teaming competencies.

Our last principle is another based on principle five — separating the what from the how. It addresses an aspect of teaming that is important, namely teaming strategy. Teaming strategy is the means by which a team chooses to exploit available interdependence options. In traditional systems that tackle multi-agent teaming, the teaming strategy is often tightly coupled to the overall system implementation, making it difficult if not impossible to change teaming strategies (e.g. [23]). It also confounds the scientific analysis of teamwork by not separating the teaming capability within the work from the teaming strategy employed by the team members. Learning to understand the teaming capability of the work separate from the teaming strategy of the team members helps designers comprehend the influence of each other on overall system performance. Thus principle eleven is:

> *Principle 11: The joint-ness of work and the teaming strategy can and should be considered separately, though they need to be designed to work together.*

Teaming strategy is not addressed by the IA tool, but it is included in the discussion because strategy is an important part of teaming. The IA tool provides insight into the tools and options available to teaming strategies.

The purpose of these principles is to help reshape a designer's mindset in order to effectively employ the Interdependence Analysis tool. Table 1 summarizes the interdependence design principles.

Table 1. Summary of Interdependence Design Principles

Interdependence Design Principles	
1	The value of teaming comes from exploiting it not avoiding it.
2	All work should be designed as joint activity (coactive), with independent work being the degenerate case.
3	Any work has inherent potential and limitations on joint-ness.
4	Teaming occurs not just on the physical level, but also on the cognitive level.
5	Separate the "what" from the "how."
6	Joint work can and should be modelled in an agent agnostic manner.

7	Interdependence provides the basis for understanding potential joint-ness.
8	Teaming involves both required (hard) and opportunistic (soft) interdependencies.
9	Observability, Predictability and Directability are compulsory interdependencies in teamwork.
10	Failure is always an option and teaming should be designed to help the team be robust to failure of both people and machines.
11	The joint-ness of work and the teaming strategy can and should be considered separately, though they need to be designed to work together.

4.0 INTERDEPENDENCE ANALYSIS TOOL

The purpose of Interdependence Analysis (IA) is understanding how people and automation can effectively team by identifying and providing insight into the potential interdependence relationships used to support one another throughout an activity. The Interdependence Analysis tool was developed to assist with IA. The tool can be used to analyze existing systems, but one of the tool's strengths is that it can also be used formatively to guide the initial design process. The need for formative tools is consistent with Kirlik et al., who emphasize "the importance of understanding why cognitive demands are present, prior to determining a strategy for aiding the operator in meeting these demands" [24](p. 950). Understanding and designing for interdependence can provide this type of guidance. IA provides insight into when situation awareness is adequately supported and when it is not. It can inform the designer of what is and is not needed, what is critical, and what is optional. Most importantly, it can indicate how changes in capabilities affect relationships. As systems develop and improve, understanding the impact of how these changes impact human-autonomy teaming is critical to ensuring acceptance and utility of new technology.

The IA tool is in the form of a table, as shown in Figure 1, with three main sections: joint activity modelling, assessment of potential interdependence, and analysis of potential workflows.

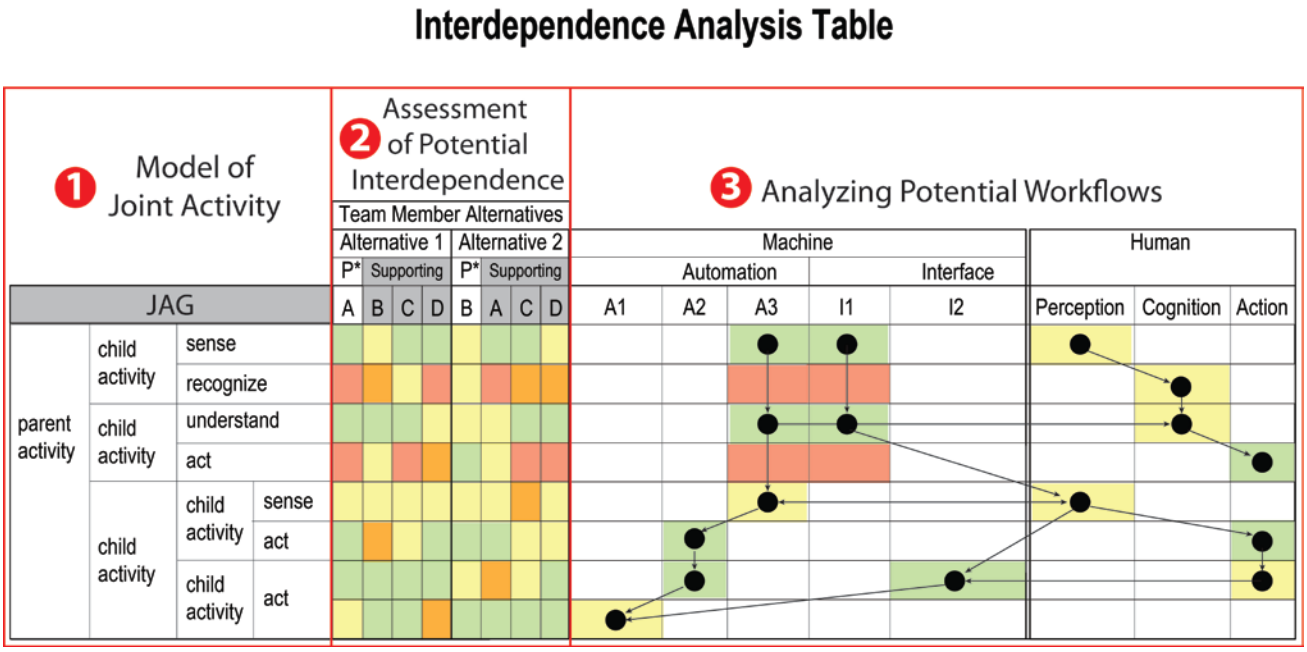


Figure 1. Generic Interdependence Analysis Table with three main section labelled. Section 1 helps designers model the joint activity, section 2 helps them identify potential interdependencies in the activity, and section 3 helps analyze the potential workflows to better understand the flexibility and risk in the human-machine system.

4.1 Modelling of Joint Activity

The first section of the IA tool focuses on modelling the joint activity. In accordance with principle 2, this modelling should model all work as joint activity. In order to accomplish this, it is important to understand what is unique and important about joint activity and how these aspects can be modelled.

4.1.1 What is Joint Activity?

So, what does it mean to model some task or function as joint activity? Our view of joint activity comes from work on joint activity theory [10], [22], a generalization of Herbert Clark's work in linguistics [4]. Joint activity has important characteristics with respect to structure, process and the potential for interaction [25].

The first distinction is in the overall structure of joint activity. Joint activity is embedded sets of actions. When viewed as a task, traditional engineering practice suggest making the given work an atomic isolated module that performs the specified function when called. The internals of the standalone module are typically hidden through encapsulation. While this is good programming practice for software development and integration, it turns out to be problematic for teaming. This is because tasks or functions are actually part of embedded sets of actions, in other words an activity. The function may have several embedded sets of actions within it, or it may itself be embedded in a set of actions or both. Though team members may be working on functions that can be represented individually, it is important, from a teaming perspective, to understand the overall team activity context provided by the activity structure.

The second distinction is that joint activity is a process, one that extends in space and time. When viewed as a task or function, this is often overlooked, as if no time transpires and the world stands still. This has many implications from a teaming perspective. The first is that events of the past, current status within an activity and intentions for the future are all potentially relevant for effective teaming. Additionally, having a process implies that there is additional work necessary to compose the hierarchical structure. This additional work,

often referred to as coordination, is also important for effective teaming.

The last distinction important to this discussion is that joint activity has the potential for interaction. While considering work as embedded sets of actions that are part of a process shifts thinking from individual tasks to activities, it is the potential for interaction that enables the activity to be considered joint. If there is not substantive interaction, then the work is parallel — not joint [25]. This need to support interaction drives what it means to model joint work. It means to decompose joint activity into a sets of embedded actions and to instrument those actions to properly support interactions needed for distributed team members to recompose the final solution. This is where designing for interdependence opposes traditional information-hiding practices. Functional encapsulation and opaque automation are often at odds with effective teaming. Understanding the nature of joint activity also plays a role in a more nuanced understanding of interaction. Since activity involves nested sets of activities, interaction can happen across a variety of levels of abstraction. Since it is a process, interaction can involve the past, the future, and ongoing progress during an activity. Defining a system as a human-machine team is defining it as joint activity, but defining it alone is not enough to make it a reality; the system must be designed and built.

4.1.2 How to Model Joint Activity – The Joint Activity Graph

While there are conceivably numerous ways to model joint activity, we propose one solution called the Joint Activity Graph (JAG). The purpose of the JAG is to capture the key elements of joint activity: structure, process and potential interactions.

The data structure underlying a JAG, is a simple graph. It is a tree of unit height — a single activity with zero or more children. More complex activity can be constructed by assembling multiple JAGs into a larger structure that is itself a JAG. Any high level JAG can be understood by recursively expanding all the children to provide a tree-like view on the overall activity. This recursive data structure has intrinsic benefits for composition and reuse. The JAG provides a description of all the embedded activities that are necessary to the proper execution of the activity it describes. As a hierarchical structure, it is similar to hierarchical task networks (HTNs) [26], [27]. As with HTNs, JAGs can be directly executable or conceptual like a goal, sometimes referred to as primitive and non-primitive tasks respectively. However, JAGs do not require explicit declaration as such. Whereas planners are built on top of existing systems with known capabilities, human-machine teaming is constantly evolving and the potential variation in team composition means rigid assumptions on the type of actions that are executable by one agent will likely be invalid for some other agent. Thus, JAGs were designed to postpone the decision on what is directly executable allowing teams flexibility to resolve this at runtime.

If structure alone was all that was needed to understand teaming, then JAGs would be unnecessary as a hierarchical tree would suffice. However, the assembly of a JAG into a more complex activity also requires a process description. JAGs contain a process model whose purpose is modelling how to assemble its immediate descendants into the parent. As with structure, flexibility is an important aspect of the teaming process. The goal of modelling the process with a JAG is not to dictate the “right way” to complete an activity, but to model the permissible ways to do so. The JAG aims to depict the roadmap of options not just one viable road. In other words, it is not meant to define the one solution, but a solution space. The re-composition process can be individually defined for each unit JAG and is fully customizable. While these can be designed from scratch to fit specific needs, we natively support a set of common useful alternatives. These alternatives are combinations of two broad concepts: sequencing and logic. Sequencing provides for sequential or parallel execution. Logic provides the operators “And” and “Or” for specifying conjunctive and disjunctive tasks. Together these provide four useful combinations:

1. Sequential-And: All activities need to be completed successfully in sequence.

2. Sequential-Or: Activities executed in sequence until one is completed successfully.

3. Parallel-And: All activities can be executed in parallel and need to be completed successfully.

4. Parallel-Or: All activities can be executed in parallel but are racing each other until one completes successfully.

These four processes provide sufficient mechanisms to model many different activities. However, because it is unlikely that we can cover all types of processes imaginable, we left the process definition open for future extensions.

Structure and process dictate how an activity can be decomposed and distributed and how it can be recomposed to achieve the goals of the activity, however neither defines the potential for interactions that determine the joint-ness. This is accomplished by identifying the interdependence relationships that underlie and define the needed interactions which in turn enable teamwork. Some interdependencies come from the structure. For example, if a JAG's children are distributed across team members, to know that the parent JAG is successful may require information sharing. Interdependence can also stem from the process. For example, if sequential activities are distributed across team members then they will need to coordinate their sequencing. However, many of the important teamwork dimensions such as adaptability, situation awareness, performance monitoring and feedback, and decision making [28] involve interdependencies beyond those associated with structure or process. The purpose of the IA tool is to help designers identify potential interdependencies. Specifically, observability, predictability and directability (OPD) requirements explain what kind of interactions are potentially valuable to achieve effective teaming (principle 9). More importantly they depict precisely how a given activity must be instrumented to expose and ingest appropriate information.

4.2 Assessing Potential Interdependence

After modelling the joint activity, the next step is to assess the potential interdependence. This involves enumerating the viable team role alternatives, assessing the capacity to perform the work, assessing the capacity to support another team member as they perform the work, identifying potential interdependencies, and then determining the requirements to support the interdependence relationships of interest.

4.2.1 Enumerating Viable Team Role Alternatives

While the joint activity is modelled in an agent-agnostic manner, the assessment of interdependence is in general not. This is because interdependence is about relationships. Therefore, any team alternative needs at a minimum two team members, though larger teams are permitted. Having specific individual team members is not required, though it does permit greater specificity. The team alternatives section of the tool captures the fact that team composition has an impact on teaming. It also permits analysis and comparison of how changes to team composition impact teaming.

To aid interdependence analysis, the IA tool makes a distinction between a performer and supporting team members. The performer is the individual primarily doing the given aspect of the work. The supporting team members are then viewed from the perspective of assisting the performer. It is the supporting team member columns that are key to identifying interdependencies. For a given alternative, only one entity is assigned as the performer, labelled P* in Figure 1. This is not to say others cannot do the work, but is simply a mechanism to aid the designer in considering a certain perspective.

In general, IA tables will have a minimum of two alternatives. The first with a given performer and supporter and the second with the roles reversed. This allows the designer to consider all of the joint work from both perspectives. This permutation is another key element to identifying all potential interdependencies, which in

turn guides the design of effective teaming. In addition to considering any party performing any part of the work, this also forces consideration for any party assisting in any part of the work. If a team consisted of two identical team members, then having two team alternatives would be redundant since they would be identical. The main reason for a minimum of two alternatives is that people and machines are inherently different. This asymmetry needs to be understood and accounted for when designing human-autonomy teaming.

The columns in each alternative can represent specific individuals (existing or planned). If the team has more than two members then additional columns can be used, as represented by columns A, B, C and D in Figure 1. For larger teams this can become unwieldy, but categories and roles can be used to keep it manageable. For example, consider a single operator managing four unmanned aerial vehicles (UAVs) and two unmanned ground vehicles (UGVs). Assuming the vehicles are of the same type then categories can be used. The team alternative would be three columns: one human operator, one for UAVs, and one for UGVs. Multiple people are also permitted and can be simplified with roles. Consider extending the previous example to be two such units being managed by a commander. These fifteen entities can be captured by four columns: commander, operator, UAV, UGV. There is a limit to the feasibility of extending the table to large teams, but such teams are more like organizations than teams.

In general, a permutation of team members in which each is assigned the performer is how team alternatives are created. While permutations across a team seems a bit daunting and suggest visions of exponential explosion, in practice we have found that humans generally team in small numbers and larger teams use categories and roles to manage scale. Some may question whether such consideration for the team composition is necessary for effective design. Consider a simple example of providing some driver instructions on how to get to your house with some road construction blocking the obvious route. If the driver you are teaming with is someone familiar with your neighborhood, versus someone unfamiliar you probably need different coordination mechanisms. If the driver is a teenager learning to drive, or a young adult, or a senior citizen you might also imagine different teaming mechanisms. There is no “one size fits all” when designing teaming, especially when considering team composition. This is not a limitation of our approach, but merely a reflection of reality. There is a tradeoff to be made with respect to how much detail to include, but having no representation of team composition is unlikely to be effective.

4.2.2 Assessing Capacity to Perform and Capacity to Support

After the team alternatives are determined, the next step is an assessment of each column in a team alternative to each row in the joint activity model. For formative tasks, the assessment is necessarily subjective, but when assessing existing systems, it is possible to use empirical data to inform the assessment [29]. The assessment process uses a color coding scheme, as shown in Figure 2. The color scheme is dependent on the type of column being assessed.

Color Key for Team Member Role Alternative Capability Assessment	
Performer	**Supporting Team Members**
I can do it all	My assistance could improve efficiency
I can do it all but my reliability is < 100%	My assistance could improve reliability
I can contribute but need assistance	My assistance is required
I cannot do it	I cannot provide assistance
Not applicable / Not significant	Not applicable / Not significant

Figure 2. Color key for team member role alternative capability assessment.

Under the "performer" columns, the colors are used to assess the individual's capacity to perform the activity specified by the row. It complies with principle 10 — considering the potential for failure. The color green in the "performer" column indicates that the performer can do the task. For example, a robot may have the capacity to navigate around an office without any assistance. Yellow indicates less than perfect reliability. For example, a robot may not be able to reliably recognize a coffee mug all the time. Orange indicates some capacity, but not enough for the task. For example, a robot may have a 50 pound lifting capacity, but would need assistance lifting anything over 50 pounds. Another use of orange is to indicate hard-coded assumptions that limit the performance to very specific contexts. For example, a robot may be able to pick up a coffee mug from a table, but it may not be able to do so from the floor or cluttered cabinet. The color red indicates no capacity, for example, a robot may have no means to open a door.

Under the "supporting team member" columns, the colors are an assessment of that team member's potential to support the performer for the activity specified by the row. The color red indicates no potential for interdependence, thus independent operation is the only viable option for the task. Orange indicates a hard constraint, such as providing supplemental lifting capacity when objects are too heavy. Another example of orange is when a machine needs human authorization to perform the activity. Yellow is used to represent improvements to reliability. For example, a human could provide recognition assistance to a robot and increase the reliability in identifying coffee mugs. Green is used to indicate assistance that may improve efficiency. For example, a robot may be able to determine the shortest route much faster than a human or could assist in cleaning up a room.

One last note on color coding. Some relationships are not significant and so either the performer or the supporting team members' assessment may not be applicable. In these cases, use of grey is suggested to minimize the attention drawn by such cases [30].

4.2.3 Identifying Potential Interdependence Issues

Once the assessment process is finished, the color pattern can be analyzed. The color pattern characterizes the nature of the interdependence within a team for the given joint activity.

Colors other than green in the "performer" column indicate some limitation of the performer, such as potential brittleness due to reliability (yellow) or hard constraints due to lack of capacity (orange). The hard constraints in the performer column indicate a need to team to accomplish the work and are usually fairly obvious. The more interesting situation is the potential brittleness which is often less obvious. Teaming is not required in this circumstance, but doing so can make the team more resilient.

Colors other than red in the "supporting team member" columns indicate potential required (orange) interdependence relationships between team members. Again, it is the opportunistic cases (yellow and green) that tend to be the most interesting, less obvious and contribute to resilience.

4.2.4 Determining System Requirements

With the assessment complete, the IA tool can now help designers extract clear-cut design requirements needed to enable and support specific interdependence relationships. For each relationship of interest, the design considers the compulsory interdependencies of observability, predictability and directability (OPD) in accordance with principle 9. In other words, identifying who needs to observe what from whom, who needs to be able to predict what, and how members need to be able to direct each other for a given aspect of the work. As an example, we have created a small IA table based on Fong's Collaborative Control work [18], as shown in Figure 3. In Fong's example there was one teaming alternative: the human assisting the robot. The robot was capable of performing obstacle avoidance; however, it was less than 100 percent reliable (yellow) in interpreting if an obstacle is passable. The human was capable of providing assistance, thus increasing the reliability (yellow) of the robot in this task. The yellow coloring of the human column indicates soft

interdependence. Requirements can be derived from analyzing the IA table in Figure 3. The robot must be predictable in notifying the human when unsure about an obstacle because the human is not constantly watching (red). The human's ability to interpret depends on being able to sense the obstacle, so there is an observability requirement. Once the human has interpreted if the obstacle is passable, this information must have a way to alter the robot's behavior, so there is a directability requirement. These requirements define what is needed by both the algorithm and the human interface to support this interaction. Note that the IA tool's purpose is to identify what the requirements are, not how to meet them (principle 5). That is an implementation choice. These particular OPD requirements are based on the desire to support a particular interdependence relationship: the human assisting in interpretation of whether an obstacle is passable. This example demonstrates how OPD requirements derive from the role alternatives the designer chooses to support, their associated interdependence relationships, and the required capacities. This example does not include the reciprocal teaming alternative with the human assisted by the robot because this was outside the scope of the original work.

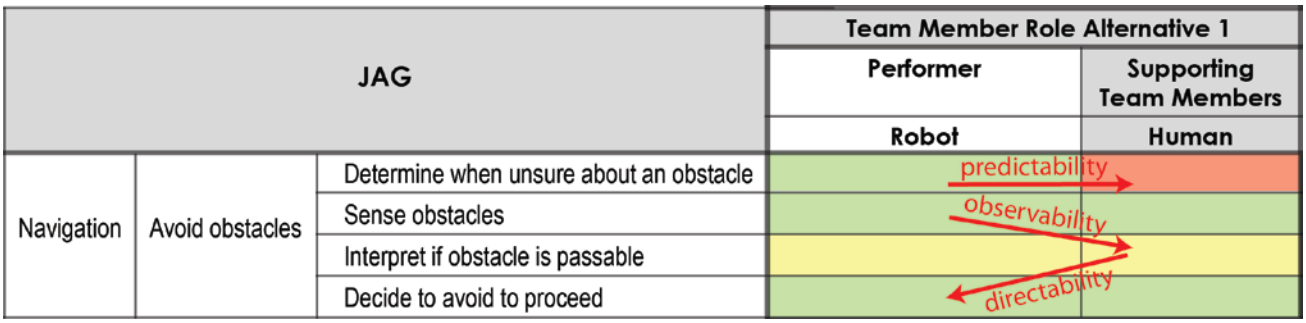


Figure 3. Interdependence analysis example from Fong's (2001) Collaborative Control work, showing observability, predictability and directability requirements based on choosing to allow the human to provide interpretation assistance to the robot during navigation.

4.3 Analyzing Potential Workflows

The original IA tool was introduced as a way to help designers understand and design for the interdependence in human-machine teams [3]. The assessment of potential interdependence provided this, but demanded a fair bit of imagination to envision the implementation alternatives. As the tool was applied in the development process within a team of engineers, it became clear that it would be valuable to provide an improved visualization of the teaming alternatives to aid the team in developing a unified understanding [29]. Specifically, we needed a way to connect the theoretically understanding of interdependence to the physical instantiation of what had been developed thus far (i.e., the implementation) and what was planned to be developed.

4.3.1 Establishing Workflows

To achieve this, we associated leaf JAGs to particular team member's capability of supporting them. We also expanded the team members to allow more detailed mapping to specific algorithms, interface elements, or human abilities. Across the top of section 3 in Figure 1, there are column headings for each algorithm, interface element or human capability used to accomplish the task. Below each heading is a black dot to indicate where in the activity that particular component has a role. We then connect the dots with arrows to indicate potential workflows to accomplish the goal. The resulting graph structure is a visual description of all existing and potential workflows. Therefore, it is a depiction of the flexibility in the system [31]. The graph can be thought of as the adjustment options in adjustable autonomy or the initiative options in mixed-initiative interaction [32]. However, what they really are is an enumeration of the possible ways a team can

complete the joint activity. This is how the IA tool provides not a solution, but a solution space.

Additionally, the graph makes it clear that discrete function allocation is not what is happening because the human is informed by automation and display elements, and automation can be assisted by the human as indicated by the numerous horizontal and diagonal lines shown in section 3 in Figure 1. Most importantly, it shows where teaming is supported and where it is not. This allows grounded debate on the value of additional teaming support versus the cost of implementation.

4.3.2 Assessing Workflows

The color coding in the workflow section of the IA tool, section 3 of Figure 1, is indicative of risk for a given pathway. The color for a given cell is determined by the performer column of the assessment section. Moving vertically the risk is compounded across different functions. However, horizontal pathways indicate potential mitigations through teaming.

The potential workflows section of the IA tool is valuable for several reasons. First it ensures the joint activity model grounds out in actionable terms and does not fall prey to wishful mnemonics [33] or suitcase words [34]. Delineating the roadmap of possibilities also helps to guide a design team in their development. They can easily see what pathways are supported and which are not, but more importantly they can be guided to potentially opportunistic pathways they may not have found otherwise. The IA tool structure is also well suited to collection of empirical data to help inform design choices throughout the design iteration process (e.g. [29]). Information about reliability, performance time and frequency of use associated with specific pathways can help shape design decisions and provide credence to implementation directions.

5.0 INTERDEPENDENCE ANALYSIS EXAMPLE

As an example of how to use the IA tool we will use it to analyze manned aircraft collision avoidance. Figure 4 depicts an abbreviated IA of this challenge. The JAG, shown in the left side of Figure 4, depicts the agent agnostic joint activity (principle 3, 5 and 6). The work includes the cognitive aspects (principle 4) to sense and interpret traffic, decide when there is a conflict and what to do about it, as well as the physical task to take the necessary actions. We assume the work is jointly shared (principle 2) by some automation working together with a human pilot. Our hypothetical system includes automation, such as a Traffic Collision Avoidance System (TCAS) and a Detect and Avoid System (DAA), shown as headings in the workflow section of Figure 4. The figure depicts two TCAS options for discussion purposes. In teaming alternative one (TA1), assessing the automation's ability to perform, we see that TCAS can reliably sense and interpret traffic, but only cooperative traffic. TCAS has no ability to detect non-cooperative traffic (i.e. not transponder equipped). While TCAS-I only warns about traffic, TCAS-II can decide what to do and will provide vocalized instructions, but this is based on the limited sensing capability of TCAS-I. While DAA systems are not fully operationalized yet, here we assume we are designing a system to have such capability. Our envisioned DAA system leverages TCAS decisions to determine control inputs. We have assumed perfect execution (green) but left open the possibility that the automated system might misdiagnose on rare occasion (yellow) accounting for potential failure (principle 10). By considering that the human pilot has capabilities that could assist the automation (yellow) such as recognizing uncooperative traffic and making the avoidance decision we have identified potential soft interdependencies (principle 7 and 8). TA2 shows the human is capable of everything except directly executing the controls with errors most likely occurring in sensing due to attention or decision making (slow reaction time). Automation can assist the pilot in sensing cooperative traffic and making the avoidance decision (yellow), and generating control commands (green).

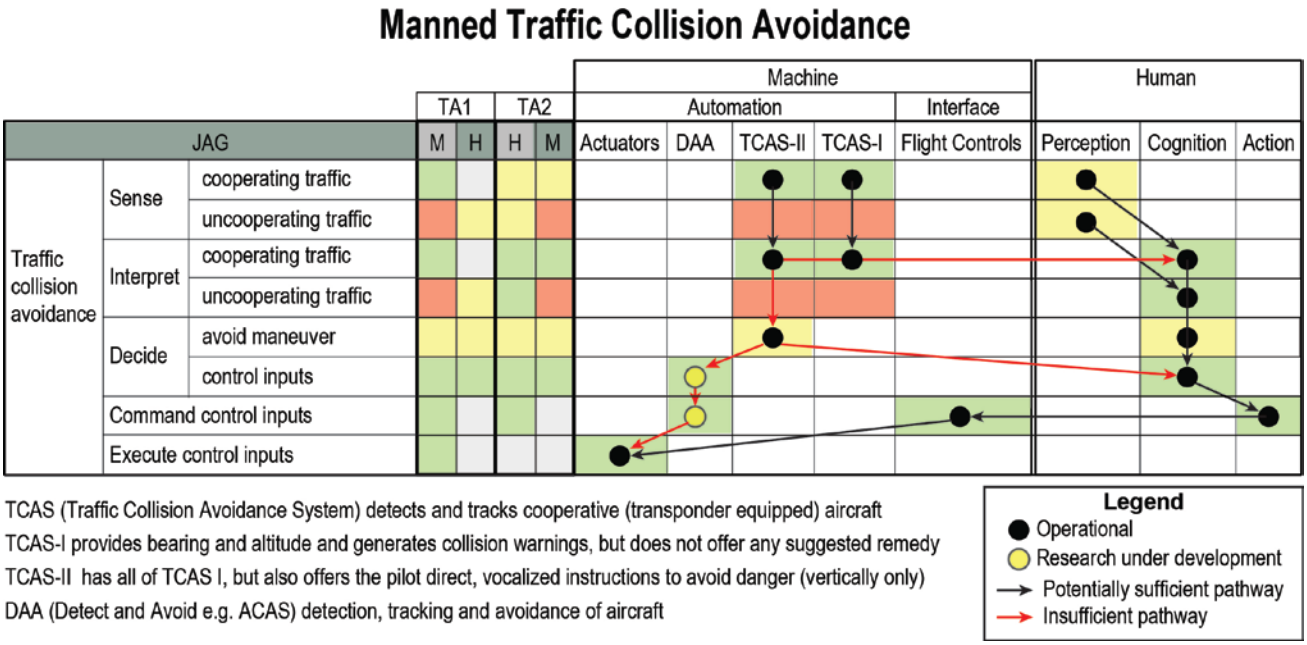


Figure 4. Interdependence Analysis of Manned Traffic Collision Avoidance.

The potential workflows are depicted in the right half of Figure 4. One can see the potential automated solution with TCAS identifying traffic and DAA commanding avoidance maneuvers. Though the final execution of commands is highly reliable by today's automation (green), it is dependent on sensing that lacks some context (red) and decision making that may not be perfect (yellow). This provides designers with an overall risk assessment for the automated solution path. The manual solution is also possible with the pilot detecting traffic, deciding what to do and taking action. There is risk in this solution, but it is different. Pilots have limited attention and may miss traffic and they may misjudge a decision. These are just two solutions, neither of which exploits teaming. By considering the supporting team member columns in each teaming alternative, designers are guided toward teamwork options throughout the activity.

TA2 considers how the automation can help the pilot. Each option can be represented by horizontal lines that cross any boundary between teammates. Each relationship will have its own associated OPD requirements (principle 9). For example, TCAS can sense and interpret cooperating traffic for the pilot. To support this teaming pattern, the system must make potential traffic conflicts it detects observable to the pilot, in this case through a warning display. This activity can be done in parallel and thus both the pilot and the automation can contribute. The workflow includes both sense and interpret pathways because the automation is not replacing, but supplementing the pilot's ability, as most flight manuals will warn. It also means the risks of each individual pathway, missing context for automation and lack of attention for pilots, can potentially be mitigated by the other. TCAS-II can provide a suggested course of action. This requires a mechanism to direct pilot action, in this case voice instructions. Note that although the DAA has the potential to determine control inputs, it currently does not support providing such assistance directly to the pilot, only through the fully automated solution.

TA1 considers how the human can help the automation. Although the assessment column for the human in TA1 indicates the potential to help recognize uncooperative traffic or decide on a course of action, no support is provided, as shown in the workflow. If the pilot performs these activities, they are done outside of any teaming and are not sharable with the automation in the current design.

This example demonstrates the process of interdependence analysis and how it helps designers understand human-machine teaming. We connected the design principles to the analysis story to show how they relate to

teaming and how the IA table guides the designer through these considerations. It should be clear that the potential for teaming and the requirements for supporting it will change as design changes are considered, such as different types of automation (e.g. TCAS-I vs. TCAS-II) or having the pilot be remote. This is how the IA tool aids designers in diagnosing the teaming implications of design choices.

5.1 Summary of IA Tool Capabilities

The IA tool is unique in its capability to aid designers in assessing interdependence in human-machine teams. Currently, it is the only design tool that specifically shows interdependence. It is also novel in that it does not try to determine which team member, the human or the machine, should be allocated to what aspects of work. Instead it focuses on capitalizing on the opportunities for synergy in the team. It enables designers to more effectively find teaming opportunities and helps them understand the requirements needed to exploit those opportunities (principle 1).

The IA tool is also novel in how it includes both people and machines in the design of a system. Many approaches do not include people as part of the system, while others assess human performance without the context of the automation and interfaces. The novel approach to modelling all work as joint work in an agent agnostic way allows for complete flexibility of teaming. The support for consideration of teaming alternatives helps designers to see the problem from multiple perspectives. Providing workflow visualization that includes people, algorithms and interfaces grounds the theory into actionable practice.

The IA tool is also unique in how it captures soft interdependencies. These types of capabilities are often overlooked, but their importance to teaming should not be. Support for such opportunistic teaming options can lead to improved flexibility and better overall system resilience [29]. All too often the purpose of human-autonomy teaming is viewed as a constraint due to automation not being able to do something. IA is about capitalizing on opportunities for synergy, not just filling required slots. It is about mutual enhancement, not replacement or substitution.

The IA tool is also unique in providing not just a single solution, but a roadmap of design alternatives. This type of view is critical for development of advanced technologies which involve highly iterative development. Supporting multiple pathways on the roadmap provides flexibility and contributes to system resilience.

A critique of the IA process is that it is potentially time consuming to develop joint activity models and consider the different teaming alternatives. This is a valid concern, but building systems that fail or struggle to meet people's needs is also costly. While interdependence analysis will definitely take more time than simple task decomposition, it is unlikely to take more time than the first iterative development of any complex system. Our experience is that its value throughout the life of the project greatly exceeds the initial time investment.

Another concern is that IA requires a high level of expertise. In some sense this is true. The main issue is shifting away from the traditional paradigm. Shifting from a function allocation mind set is not easy, just as shifting to object oriented programming or functional programming is not easy for someone trained in procedural programming. It also requires two skill sets not often found in combination: technology expertise and human expertise. However, ignoring either technology or the human because dealing with both is difficult is not a viable answer. Future educational tracks that emphasize both could alleviate this problem. We have proposed several guiding principles to help designers reorient their perspective, but as in any field, there is no substitute for practice and experience.

6.0 CONCLUSION

The purpose of Interdependence Analysis (IA) is understanding how people and automation can effectively team by identifying and providing insight into the potential interdependence relationships used to support one another throughout an activity. These interdependence relationships enumerate the potential human-machine interactions that comprise teaming. In this way, it provides a roadmap of the opportunities afforded by a given system design. Interdependence relationships help to understand both the human factors and the technological factors that enhance or inhibit effective teaming. IA can be used to derive specific design requirements, both algorithmic and interface, that determine human-machine interactions for each alternative. The analysis also provides a risk assessment associated with each teaming option. Having alternatives provides flexibility, while understanding the risk associated with each option assists operational decision making.

As we build more intelligent machines to tackle more sophisticated challenges, designers need formative tools for creating effective human-autonomy teaming. Interdependence Analysis meets this need by explicitly modelling the machine, the human, the work and the interplay of all three. This enables developers to architect effective teaming by designing support for interdependence. Interdependence Analysis should be viewed as an approach to determining how and what automated capabilities are built such that intelligent systems are imbued with teaming competence.

7.0 REFERENCES

[1] B. Crandall and G. Klein, Working Minds: A Practitioner's Guide to Cognitive Task Analysis. Bradford Books, 2006.

[2] R. R. Hoffman and S. V. Deal, "Influencing versus Informing Design, Part 1: A Gap Analysis," IEEE Intell. Syst., vol. 23, no. 5, pp. 78–81, 2008.

[3] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, B. M. van Riemsdijk, and M. Sierhuis, "Coactive Design: Designing Support for Interdependence in Joint Activity," J. Human-Robot Interact., vol. 3, no. 1, pp. 43–69, 2014.

[4] H. H. Clark, Using language. Cambridge [England] ; New York: Cambridge University Press, 1996.

[5] M. Johnson et al., "Beyond cooperative robotics: The central role of interdependence in coactive design," IEEE Intell. Syst., vol. 26, no. 3, 2011.

[6] T. B. Sheridan and W. Verplank, "Human and Computer Control of Undersea Teleoperators." Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT, Cambridge, MA, 1978.

[7] R. Parasuraman, T. Sheridan, and C. Wickens, "A model for types and levels of human interaction with automation," Syst. Man Cybern. Part A, IEEE Trans., vol. 30, no. 3, pp. 286–297, 2000.

[8] M. Johnson, J. Bradshaw, P. Feltovich, C. Jonker, B. van Riemsdijk, and M. Sierhuis, "The Fundamental Principle of Coactive Design: Interdependence Must Shape Autonomy," in Coordination, Organizations, Institutions, and Norms in Agent Systems VI, vol. 6541, M. De Vos, N. Fornara, J. Pitt, and G. Vouros, Eds. Springer Berlin / Heidelberg, 2011, pp. 172–191.

[9] P. F. Drucker, The Practice of Management. Harper Business, 1954.

[10] G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich, "Ten Challenges for Making Automation a 'Team Player' in Joint Human-Agent Activity," IEEE Intell. Syst., vol. 19, no. 6, pp. 91–95, 2004.

[11] E. D. Sacerdoti, "The Nonlinear Nature of Plans," Proc. 4th Int. Jt. Conf. Artif. Intell., vol. 1, pp. 206–214, 1975.

[12] M. desJardins and M. Wolverton, "Coordinating a Distributed Planning System," AI Mag., vol. 20, no. 4, p. 45, 1999.

[13] C. D. Wickens, H. Li, A. Santamaria, A. Sebok, and N. B. Sarter, "Stages and Levels of Automation: An Integrated Meta-analysis," Proc. Hum. Factors Ergon. Soc. Annu. Meet., vol. 54, no. 4, pp. 389–393, Sep. 2010.

[14] M. Stefik, "Planning with constraints (MOLGEN: Part 1)," Artif. Intell., vol. 16, no. 2, pp. 111–139, 1981.

[15] K. Christoffersen and D. D. Woods, "How to Make Automated Systems Team Players," Adv. Hum. Perform. Cogn. Eng. Res., vol. 2, pp. 1–12, 2002.

[16] I. D. Steiner, Group process and productivity. Academic Press, 1972.

[17] D. a Norman, “Cognitive Engineering,” User-centered Syst. Des., pp. 31–61, 1986.

[18] T. W. Fong, “Collaborative Control: A Robot-Centric Model for Vehicle Teleoperation,” no. CMU-RI-TR-01-34. Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 2001.

[19] M. Quigley et al., “ROS: an open-source Robot Operating System,” Icra, vol. 3, no. Figure 1, p. 5, 2009.

[20] T. W. Malone and K. Crowston, “The interdisciplinary study of coordination,” ACM Comput. Surv., vol. 26, no. 1, pp. 87–119, 1994.

[21] E. Salas, M. A. Rosen, C. S. Burke, and G. F. Goodwin, “The wisdom of collectives in organizations: An update of the teamwork competencies,” in Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches, New York: Routledge/Taylor & Francis Group, 2009, pp. 39–79.

[22] P. J. Feltovich, J. M. Bradshaw, W. J. Clancey, and M. Johnson, “Toward an Ontology of Regulation: Socially-Based Support for Coordination in Human and Machine Joint Activity,” in Engineering Societies in the Agents World VII, vol. Lecture No, G. O’Hare, M. O’Grady, A. Ricci, and O. Dikenelli, Eds. Heidelberg, Germany: Springer, 2007, pp. 175–192.

[23] M. Dias, R. Zlot, N. Kalra, and A. Stentz, “Market-Based Multirobot Coordination: A Survey and Analysis,” in Proceedings of the IEEE, 2006, pp. 1257–1270.

[24] A. Kirlik, R. a Miller, and R. J. Jagacinski, “Supervisory control in a dynamic and uncertain environment II: A process model of skilled human environment interaction,” IEEE Trans. Syst. Man. Cybern., vol. 23, pp. 929–952, 1993.

[25] J. M. Bradshaw, P. J. Feltovich, and M. Johnson, Human-agent interaction. 2011.

[26] K. Erol, J. Hendler, and D. S. Nau, “HTN planning: complexity and expresivity,” AAAI, pp. 1123–1128, 1994.

[27] I. Georgievski and M. Aiello, “An Overview of Hierarchical Task Network Planning,” 2014.

[28] J. E. Driskell, G. F. Goodwin, E. Salas, and P. G. O’Shea, “What makes a good team player? Personality and team effectiveness.,” Gr. Dyn. Theory, Res. Pract., vol. 10, no. 4, pp. 249–271, 2006.

[29] M. Johnson et al., “Team IHMC’s Lessons Learned from the DARPA Robotics Challenge: Finding Data in the Rubble,” J. F. Robot., vol. 34, no. 2, pp. 241–261, Mar. 2017.

[30] C. Beierl and D. Tschirley, “Unmanned tactical autonomous control and collaboration situation awareness,” 2017.

[31] M. Johnson et al., “Team IHMC’s Lessons Learned from the DARPA Robotics Challenge Trials,” J. F. Robot., vol. 32, no. 2, pp. 192–208, Mar. 2015.

[32] J. E. Allen, C. I. Guinn, and E. Horvitz, “Mixed-Initiative Interaction,” IEEE Intell. Syst., vol. 14, no. 5, pp. 14–23, 1999.

[33] D. McDermott, “Artificial intelligence meets natural stupidity,” ACM SIGART Bull., 1976.

[34] R. Brooks, “The Seven Deadly Sins of AI Predictions,” MIT Technology Review, 2017. [Online].

Available: https://www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/?utm_campaign=add_this&utm_source=email&utm_medium=post. [Accessed: 17-Nov-2017].

Creating a Well-Situated Human-Autonomy Team:

The Effects of Team Structure

Elizabeth Marie Frost
Wright State Research Institute[1]
UNITED STATES OF AMERICA

Elizabeth.Frost.6.ctr@us.af.mil

Jessica Bartik **Gloria Calhoun** **Sarah Spriggs**
Air Force Research Laboratory
UNITED STATES OF AMERICA

Jessica.Bartik.1@us.af.mil Gloria.Calhoun@us.af.mil Sarah.Spriggs@us.af.mil

Heath Ruff **Kyle Behymer**
Infoscitex[1]
UNITED STATES OF AMERICA

Heath.Ruff.ctr@us.af.mil Kyle.Behymer.1.ctr@us.af.mil

ABSTRACT

Intelligent agent technologies are increasing the potential capacity for systems to behave more autonomously and are enabling more advanced human-autonomy teaming. The objective of this research was to gain a better understanding of future applications of human-autonomy teaming for the command and control of semi-autonomous systems. Specifically, the task division and relationships/roles between a human operator and an autonomous teammate were empirically examined. A control station that supports single operator management of multiple simulated unmanned vehicles performing a base defence mission was employed to study the implications of team structure and mission complexity on team performance. Operator-driven and role-driven team structures were evaluated across two levels of mission complexity by both human-human teams and human-autonomy teams.

1.0 INTRODUCTION

Effective teamwork is essential for achieving reliable operations in environments that are hypercomplex, dynamic, time compressed, and rely upon synchronized outcomes, such as military command and control [1]-[2]. Thus, the application of advanced intelligent agent capabilities to more complex and dynamic task environments, has led to an interest in research looking at how to support teaming between human operators and autonomous systems (e.g., [3]-[5]). With human-autonomy teaming being a relatively new concept, the majority of published teaming literature is focused on human-human teams. However, understanding what makes these

[1] This research was sponsored by the Air Force Research Laboratory (AFRL) under Contract FA8650-14-D-6500. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of AFRL or the US Government.

teams effective may provide insight for human-autonomy teams [6]. In the large body of research looking at different aspects of effective human teams in a wide range of domains, three specific factors have been shown to have significant influence on team performance: team cognition, communication/coordination, and team structure [7]-[9].

Team cognition, an emergent collection of cognitive processes, has been shown to be a driver of effective teams [10]-[11]. Team cognition spans various constructs such as shared mental models, transactive memory, and shared cognition, generally addressing both task knowledge and team knowledge. Task knowledge focuses on features of the team's tasks such as understanding the tools available, the task objective, and constraints. Team knowledge addresses the interaction between team members as well as the members of the team themselves, including their knowledge, skills, status, and intentions. Teammates with shared task and team mental models have an effect on overall team performance by positively influencing team processes [12]. For example, team members with a shared mental model are more likely to work towards a common objective and that eases coordination of their actions, as well as collaboration and communication [12].

Team cognition research has largely focused on human-human teams but it has been argued that the concept of shared mental models is also relevant for human-autonomy teams [13]-[14]. In this case it is important for the human operator to have appropriate system and teammate understanding in order to coordinate effectively with the autonomy [15]. This understanding can be supported by providing a human operator with sufficient training as well as experience working with the system in order to develop accurate task and team mental models. The system's interfaces will also need to support information sharing important to the development of team cognition. Additionally, the system will need to enable communication and coordination among the teammates, which is necessary for team cognition to have a positive influence on performance [12], [16].

The importance of effective communication and coordination to achieve successful teamwork is well-researched (e.g., [8], [17]-[18]). These two aspects naturally go hand in hand in that the ability to effectively communicate is vital for team coordination. This goes beyond just human-human teams, and will also need to be addressed in human-autonomy teams if they are to succeed [3], [5]. However, human-autonomy teams face different challenges when it comes to communication and coordination, such as the current lack of language technologies that support the types of communication that are regularly used in human-human teams when performing task management [19]. This means that communication that occurs naturally in human-human teams, may now present a design challenge for human-autonomy teams, making it difficult to support team performance without causing information overload and degrading information retrieval [20]-[21].

Though an interface can be designed to help support human-autonomy coordination and communication, the team structure for sharing task completion also needs to be considered. Team structure is commonly referred to as "the division of the team task into component pieces of information and capabilities and the assignment of these elements to individuals in the team" [22, p. 302]. Various human-human teams have been compared in terms of division of information and capabilities, with the results indicating that some team structures yield more superior performance than others [9]. In particular, human-human teaming research has shown that a non-hierarchical team structure may result in better team performance than a hierarchical team structure [23]. Current supervisory control approaches usually involve task delegation using a hierarchical structure, similar to a supervisor delegating tasks to a subordinate. Alternatively, a functional team structure, that involves teammates having defined roles while still working together collaboratively to achieve certain goals, is representative of a non-hierarchical structure. However, it is important to note that the task environment can influence the effectiveness of various team structures [24]. For example, in a review of air traffic control systems it was observed that having a hierarchical structure during normal operation times (with low to moderate demand) allowed for a controlled environment that still provided adequate responses. In high stressed situations, though, a

shift towards a flatter, decentralized structure allowed for the adaptability and flexibility needed to respond successfully.

2.0 OBJECTIVE

The present research specifically examined two team structures, operator-driven (supervisory control) structure compared to a role-driven (functional) one. These two team structures were examined in both human-human teams and human-autonomy teams, with each type of team completing the same mission tasks. In the operator-driven team structure, the human operator received all the tasking and then decided which tasks to delegate to an autonomous or human teammate and which tasks to personally complete. Alternatively, in the role-driven structure, each team member fulfilled roles based on their expertise and skills (with the human teammate's skills mimicking that of the autonomous teammate). The role-driven team structure was less centralized, such that the teammate automatically began completing tasks without waiting for tasks to be delegated. Experimental participants worked with their respective teammate (either human or autonomous) to complete base defence mission tasks using multiple simulated unmanned vehicles.

3.0 METHOD

3.1 Participants

Participant recruitment and data collection are underway. Plans are to recruit twenty-four participants (12 males, 12 females) from a U.S. Midwestern university and the U.S. general public. All participants will have normal or corrected-to-normal vision and hearing. Participants are being randomly assigned such that there will be 12 participants for each type of team composition (human-human and human-autonomy). Experimental participants serve as the "operator" in all missions, either delegating tasks to a teammate in the operator-driven team structure or managing preassigned tasks in the role-driven team structure.

3.2 Simulation Testbed

The experimental trials were conducted at the U.S. Air Force Research Laboratory (AFRL) utilizing the IMPACT (Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies) testbed [25] developed by a tri-service U.S. team under the leadership of AFRL. IMPACT (Figure 1) combines several autonomy advancements into a single control station to support the command and control of multiple unmanned vehicles (UVs). These autonomous capabilities include cooperative control algorithms for rapid vehicle route planning [26], intelligent agent reasoning that compares possible courses of actions and makes vehicle allocation recommendation [27], and autonomic monitoring technologies [28].

Figure 1: IMPACT Control Station Simulation Testbed (early version).

Figure 2 illustrates the four monitors used in the participant and teammate control stations. The Tactical Situation Display (TSD) (top monitor) provided a map and current UV information such as pertinent mission locations, UVs and their associated routes, as well as a vehicle panel showing a summary of each UV's status. The bottom monitor provided a 'sandbox' display, which mirrored the TSD but also allowed the operator to create 'what-if' scenarios by generating and comparing possible UV plans before implementing them. The sandbox was also where most of the interfaces that supported operator UV management were located. The right monitor provided a detailed dashboard for each of the twelve UVs that included status information and that UV's sensor feed. Three Dell Precision R7610 workstations running Microsoft Windows 10 located in a different room provided the sensor feeds for the twelve UVs (four feeds per workstation). SubrScene, an in-house simulation visualization toolkit, was used to produce the imagery in the sensor feeds. The left monitor contained a help window and Task Log (to be described later).

Figure 2: IMPACT Testbed Display Layout Employed.

IMPACT was designed to support operator management of multiple heterogeneous UVs (air, ground and sea) in support of base defence missions. During "normal" operations (i.e., no immediate threat identified), there were two types of tasks. One type termed "RAMs" (random anti-terror measures) involved positioning one or more UVs at certain locations at either a specified time or within a time frame. For example, one of the four types of RAMs involved having an air UV circle a particular building every 10 minutes. "Normal base defence events" (NBDEs) were the other task type during normal operations and included checking on unidentified vehicle/building alarms, providing escort (ground/sea UV) or overwatch (air UV) to friendly forces, and getting a UV's "eyes on" to obtain a specified sensor view.

The simulation also supported tasks performed when an intruder was detected until an "all clear" signal was received. Each task type in response to an intruder had specified subtasks to be performed, depending on the event (e.g., perimeter breach, gate runner, mortar fire, crowd forming, and an improvised explosive device detonation). During intruder events, execution of RAMs was discontinued.

A fourth type of task involved responding to a UV failure (e.g., engine or sensor) or an environmental event (e.g., smoke or fog at a location). Finally, the chat interface in IMPACT supported the posing of query tasks, some addressing vehicle state (e.g., speed, altitude, and which vehicle was closest to a certain location) and others designed to assess the participant's situation awareness related to tasks (e.g., "What was the last task completed?"). The chat interface was also utilized for participants to rate their current workload level every 5 minutes of the scenario without pausing the simulation. The participants responded to prompts in the chat window using the Bedford Workload Scale [30].

IMPACT's flexible delegation interfaces supported high level task delegation similar to the interactions that occur in human teams. Base defence tasks were completed using the control station's interfaces that employed a playbook-based approach [31]. One or more UVs could be quickly tasked by selecting the appropriate pictorial coded icon representing a high-level play. Play selection then initiated a series of automatic actions to complete the task (e.g., call a "point inspect" play and specify the location to task a UV to automatically position an on-board sensor at that location). Participants primarily called plays from the Task Manager (see Figure 6), but plays could also be called from a play calling tile or directly from a location or vehicle on the map via a radial menu (these other interfaces are described in [32]-[34]). Once a play was called, IMPACT's intelligent agents [35] determined all the other parameters to complete the task (i.e., which vehicle(s) and route(s)). If desired, the operator could instead define one or more of these parameters as well as other details via the Play Workbook and associated pages (Figures 3 and 4). For example, by selecting the "cloud" icon to the right of the "sun" icon, the participant communicated to the intelligent agent that there was an environmental change that could influence the recommended UV/sensor type. Parameters that could be specified included environmental conditions, optimizations (e.g., time to get to target, tracking capabilities, etc.; Figure 3), a start time or end time for the play (Figure 4a), and UVs to consider for allocation (Figure 4b). Once the participant's inputs were received and the intelligent agent's plan was accepted by the participant, the play icon and details were added to an Active Play Manager (Figure 5a), and the vehicle(s) began executing the selected play.

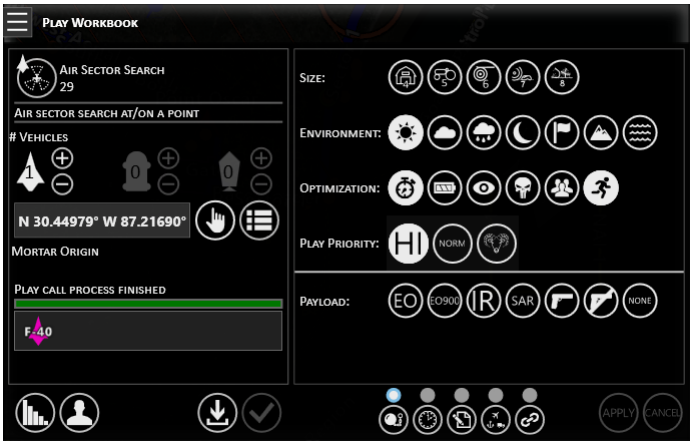
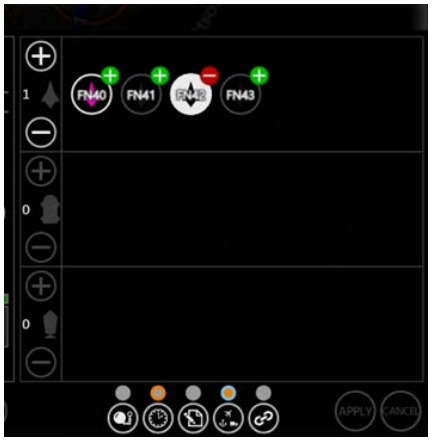


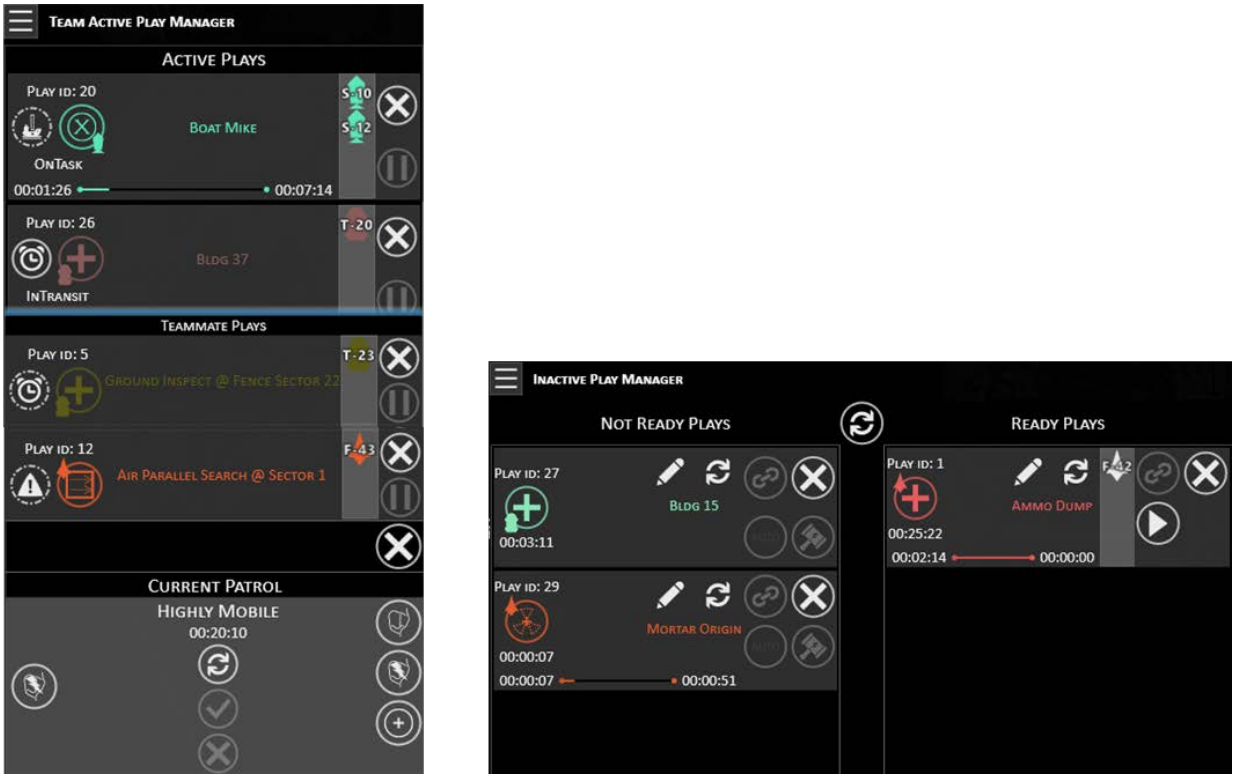
Figure 3: Play Workbook.

a) Scheduling Page



b) Vehicle Selection Page

Figure 4: Play Workbook Pages Accessible via Buttons on the Bottom of the Workbook's Right Pane.



a) Active Play Manager **b) Inactive Play Manager**

Figure 5: Play Manager Tiles.

3.3 Simulation Testbed Enhancement to Support Teaming

Although the original IMPACT simulation [25] can be viewed as representing the state-of-the art integration of several autonomous capabilities to support an operator's task delegation (via plays), it was still limited with respect to the goal of human-autonomy peer-to-peer type communication and coordination. First, with the delegation scheme, the human operator initiated each play in response to a mission event (the autonomy did not initiate plays in response to a tasking until commanded by the operator). There was no give-and-take between the operator and autonomy in terms of trading tasks or subtasks based on the teammate's capabilities or current situational demands. Even though the autonomy was quite capable for some subtasks, IMPACT's original control and display interfaces failed to provide adequate mechanisms to support operator-autonomy task coordination. Thus, before experimentation began, the IMPACT simulation was enhanced to better support this coordination and enable the evaluation of alternative team structures as a function of mission complexity.

Refinement of IMPACT's interfaces focused on improving the method by which the participant received and responded to tasks, including the ability to coordinate with a teammate. When a mission related event occurred, the participant was notified not only through the chat messaging system, but also with a pictorial icon corresponding to the task that appeared in the left column of a refined Task Manager (Figure 6). Tasks were organized in rows based on task type, with the top row being tasks generally considered highest priority (intruder events) and the bottom row being lowest priority (queries). The line coding used for the circle around the task icon designated if the task was assigned to be completed by the participant (solid), by the teammate (dashed), or

involved subtasks with both the participant and teammate sharing the workload (dashed-dotted line). The Task Manager's right pane listed the subtask or steps for completing the task that was selected on the left pane. For example, in Figure 6 the fence alarm task is highlighted in the left pane and its associated subtasks ("send UGV to fence location" and "UAV expanding square search at fence location") are shown in the right pane. The majority of tasks could be completed by a play call (or series of play calls) or, in the case of queries, by sending a chat message. In the human-autonomy teams, the "autonomous teammate" completed tasks/subtasks assigned to it (either because of the initial assignment in the role-driven team structure or by the participant selecting the lightning bolt icon on either the task or a subtask). The same task assignment procedure was used in the human-human teams.

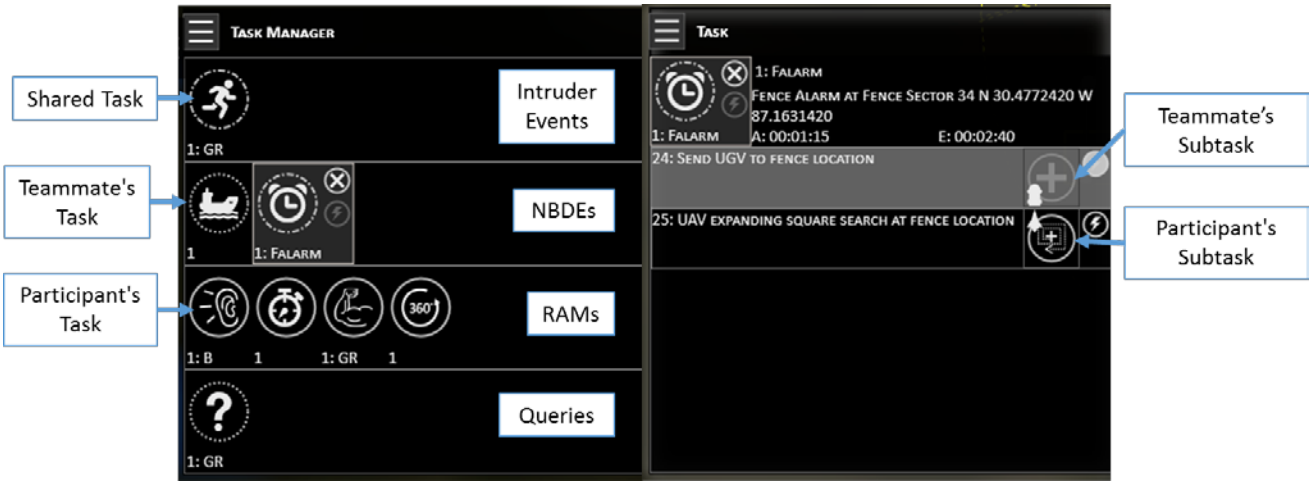


Figure 6: Task Manager.

The participant could remove a task from the Task Manager at any time. After being deleted from the Task Manager, the task was added to a new interface called the Task Log (Figure 7) that maintained a record of all the tasks received as well as details about each task (e.g., when the task was created in the Task Manager and if the operator ("owner"), teammate ("viewer"), or both ("mixed") completed the task. A task could also be sent back to the Task Manager from the Event Log.

Figure 7: Task Log.

The Active Play Manager (Figure 5a) was also revised to include all ongoing plays, grouping the plays called by the participant versus the teammate. Ongoing plays, including those called by the teammate, were able to be cancelled from the participant's Active Play Manager. Any plays that were created but not yet activated (e.g., scheduled to start at a later time or vehicles currently not available) were stored in an Inactive Play Manager (Figure 5b).

These revised interfaces complemented the other IMPACT interfaces to better support human-autonomy teaming and enable the evaluation of operator-driven versus role-driven team structures. The interfaces informed team cognition and promoted communication and coordination between the participant and teammate. This was accomplished by providing more status information about what tasks had been completed and what tasks needed to be completed, as well as providing information about intentions (who was responsible for completing a task, how they would complete it and the constraints to be taken into consideration). The participants also learned about the capabilities of the teammate during training (see 3.4.2 and 3.5.1).

3.4 Experimental Design

The mixed experimental design is illustrated in Figure 8. The team composition (human-human or human-autonomy team) was a between-subjects variable, with each of the 24 participants randomly assigned to a composition resulting in 12 human-human teams and 12 human-autonomy teams. The teammate in both conditions, either human or autonomy, had the same abilities in the experiment. Team structure and mission complexity were varied within-subjects. Each participant completed two half-hour trials, one low complexity trial and then one high complexity trial, with each of the team structures (operator-driven and role-driven), for a total of four trials with the assigned team composition (human or autonomy teammate). Team structure trials were counterbalanced to control for order effects.

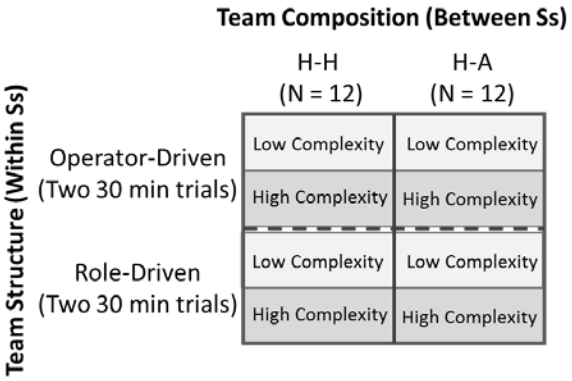


Figure 8: Experimental Design.

Table 1 shows how the levels of mission complexity differed in the number of each task type that needed to be completed. The low and high complexity missions also differed in terms of the interrelatedness and simultaneousness of tasks (and time between tasks). Besides the base defence tasks, participants were also tasked to respond to Commander queries and complete imagery tasks by confirming with a sensor operator through chat that a target was in sight. (The sensor operator role was performed by an experimenter sitting at the Test Operator Console, to be described later.)

Two experimental scenarios were developed for each complexity level (four scenarios in total). For each complexity level, the number and types of tasks (e.g., RAMs, intruder events, vehicle failures, etc.) and the task coordination demands (e.g., number of events that could be completed independently versus jointly) were the same in both scenarios. The scenarios were counterbalanced across the team structure condition, such that the low complexity scenario experienced with one team structure differed from the low complexity scenario in the other team structure; the same procedure was used for the high complexity scenarios.

Table 1: Number of Each Task Type for Low and High Complexity Missions.

	Task Response to Event Involved a:						
	Play(s)			Response	Query		
	RAMs: Random …Measures	**Normal Base Defence**	**Intruder Detected**	**UV or Environ Problem**	**UV Info**	**Task SA**	**Workload Rating**
TOTAL	**8**	**7**	**5**	**5**	**10**	**10**	**12**
Low Complexity Mission	3	2	2	2	5	5	6
High Complexity Mission	5	5	3	3	5	5	6

In order to better evaluate human-autonomy teaming, the mission scenarios were designed to provide situations that required teammate interaction for optimal performance. These included:

- Task coordination – Mission tasks varied on level of coordination, with some events providing opportunities for the participant and teammate to work together to complete a single task and other tasks capable of being completely accomplished by the participant or the teammate.
- Asymmetrical information – Either the participant or teammate obtained information that the other was not aware of (e.g. vehicle failures, environmental changes). The changes in the situation required coordination across the team for successful task completion.
- Demands exceed teammate's capability – Due to certain constraints, the teammate could not complete a task that it normally would be able to. An example was a teammate not being able to respond to an event because there were no available vehicles capable of completing the task and the teammate was not authorized to cancel on going plays to free up vehicles capable of completing the task.

3.4.1 Team Composition

For participants assigned to a human-autonomy team, the additional intelligent agent capabilities were referred to as the "autonomous teammate". The autonomous teammate role was filled by a human confederate manning a duplicate IMPACT control station. (Participants were not informed that a confederate was fulfilling this role.) The autonomous teammate had the capability to complete certain tasks without the need for participant input, such as responding to queries and calling plays in response to an intruder event (see Table 2). The autonomous teammate's capabilities were consistent across the two team structures with the only difference being whether the task assignment was operator-driven or role-driven. The operator and autonomous teammate were responsible for the same mission area and controlled the same number and types of assets (i.e., two air, two ground, and two sea UVs each). However, the teammates could not task each other's assets or trade control of assets. Asset allocation by the autonomous teammate (i.e., play calls) was reflected on the experimental participant's control station display and vice-versa, however the human operator was not able to visually view the displays of the other station. There were limited messages that participants could send through chat to the autonomous teammate. If a chat was sent that was not on that list or not in the proper format, a chat response of "I cannot understand" was sent back to the participant.

The human-human team composition consisted of the experimental participant (operator) and a human teammate that worked together to complete mission goals. The human teammate was played by a confederate whose system actions were constrained such that they matched that of the autonomous teammate's capabilities (see Table 2; participants were told their human teammate was "not authorized" to complete certain tasks rather than "not capable"). The participant and human teammate each used an individual control station to complete shared mission goals. Consistent with the human-autonomy team composition, each teammate was responsible for the same mission area and had control of half the assets. Asset allocation was reflected on both stations. Human teammates were able to communicate freely through chat; they were not limited to the chat message set used in the human-autonomy team.

Besides the participant's control station and the teammate's control station (manned by the confederate acting as the human teammate for the human-human team and as the autonomous teammate for the human-autonomy team), an experimenter-manned Test Operator Console (TOC) was also used. The TOC displays mirrored the sandbox map of the participant's station with the addition of a display that was used to drive the scenario events. Any UV allocation performed by the team members was reflected on the TOC's displays (two Perceptive Pixel Inc. 27 in. LCD monitors @2560 x 1440 resolution).

3.4.2 Team Structure

In the operator-driven team structure, the participant operator received all taskings directly from chat and decided which tasks to complete versus which tasks to assign to the teammate (either human or autonomous). All tasks were populated in the Task Manager as the participant's responsibility but the participant was also able to use the Task Manager to assign tasks or subtasks to the teammate by selecting the lightning bolt icon to the right of the task icon (see Figure 6). Participants were thoroughly trained on which task types the teammate could accomplish. There was also a help menu (left monitor) that could be referenced during the trials.

The role-driven team structure automatically assigned tasks to a teammate based on predefined roles, such that the teammate could complete tasks without the participant's assignment. Participants were trained that a task would only be assigned to the autonomous or human teammate if it was within the scope of its capabilities or authorization and if that task had been predetermined to be the responsibility of that teammate (high level roles for the operator and teammate are shown in Table 2). All tasks were populated in the Task Manager, regardless of the ownership, allowing the participant to be aware of what tasks were assigned to the teammate.

Table 2: Task Assignments for the Role-Driven Team Structure.

Participant	Teammate
• Subset of NBDEs • RAMs • Queries related to task completion	• Subset of NBDEs • Intruder Events • Queries related to vehicle status or allocation

For both team structures, participants changed task allocation between their teammate and themselves using the Task Manger (new tasks) or the Active Play Manager and chat (ongoing tasks). The teammates utilized chat messages and the Task Manager to communicate limitations or constraints to the participant (e.g., chat messages that tasks were outside their capabilities or authority and sending tasks back to the participant in the Task Manager along with a chat message as to why tasks could not be completed, such as "no available assets").

3.5 Procedure

Each participant took approximately eight hours to complete the experiment. Training and two of the experimental trials were conducted in the morning with one team structure, followed by the third and fourth experimental trials in the afternoon with the other team structure. Upon arrival, participants were given a brief overview of the study and then read and signed an informed consent document. A background questionnaire was also administered asking age, gender, vision, hearing, and video gaming experience. Lastly, participants were asked to fill out the Desirability for Control Scale [36], Propensity to Trust Machines Questionnaire [37], and the Ten-Item Personality Inventory [38].

3.5.1 Training

Initial training consisted of the participants being given a detailed overview of the IMPACT system layout, displays, and control functionality. Training covered all the abilities of the simulation testbed, as well as its limitations. Following formal training, participants were provided a review followed by approximately 30 minutes for them to practice interacting with the system by responding to a series of mission events

representative of what they would encounter during the experimental trials. This also provided an opportunity for participants to gain experience working with the teammate, to better understand the teammate's limitations and help develop a shared mental model. The experimenter was nearby monitoring each participant's actions in order to answer any questions about the system (such as why something happened or how to perform a certain action) or to intervene if necessary (for instance, if a situation occurred that a participant could not recover from, the experimenter provided an explanation and restarted the practice trial).

Subsequent training focused on how to interact and work with the assigned (human or autonomous) teammate. Participants were trained on which tasks their teammate was capable/authorized to complete, how to trade tasks between team members, and how to use chat to communicate with their teammate. This training consisted of procedures that were common to both team structures. Training on the specifics related to the two different team structures was given prior to the start of the corresponding experimental trials with that team structure (e.g., what tasks would be automatically assigned to their teammate for the role-driven team structure). An approximately 30-minute capstone scenario was completed before the experimental trials for each team structure to assess whether the participant had been trained to a sufficient level of performance. The capstone was self-paced and included every type of task they would experience during an experimental trial.

3.5.2 Experimental Trials

Participants completed four 30-minute trials, two with each of the two team structures (operator-driven and role-driven). For each team structure, the first 30-minute trial was at the low mission complexity level and the second 30-minute trial was at the high mission complexity level (see Table 1).

The following task performance data were collected: number of tasks completed, time to complete the tasks, and the correctness of tasks completed. Time to complete a task was measured as the time from when the task was received in chat to when all subtasks had been completed. Correctness of task completion was represented by an experimenter calculated summary score based on four subcomponents: (1) Did the participant select the correct location/target? (2) Did the participant select the correct play? (3) Did the participant choose the optimal vehicle or sensor? and (4) Did the participant meet the event's constraints? Accuracy and response times were collected for the participants' responses to system failures, environmental events, and commander queries. The number of tasks completed by each teammate (or jointly) and the number of tasks that were traded between the teammates, along with any associated task details (e.g., task type, task name, subtasks, etc.), were also collected. In addition to all system interactions (e.g., chats sent, buttons clicked, etc.), voice audio and video of the participant's sandbox display were recorded for all trials.

Subjective data were also recorded. A questionnaire was administered after trials were completed with each team structure. Its questions addressed the system's usability (SUS; [39]) and participants' situational awareness and trust (Checklist for Trust; [40]) for the trials just completed. At the completion of all experimental trials, a final questionnaire was given with items to compare the two team structures. An experimenter also observed participants during the scenarios and took note of any incidents of automation surprise or moments of confusion; these were later discussed in the post-mission debrief that followed questionnaire completion.

4.0 RESULTS / SUMMARY

Data collection and analyses are underway; reported results should be viewed as preliminary. Complete results will be documented in a follow-on publication. In this section, questionnaire results are first described with respect to participants' workload ratings, situation awareness assessments, and perceived performance on mission tasks. These data are examined for each team structure: operator-driven and role-driven. Next, initial

data pertaining to characteristics of the teamwork between the experimental participant and teammate (autonomous and human) are presented. This includes assessment of the team processes, and the degree of communication and coordination. Finally, example comments and questionnaire ratings pertaining to the collaborative interfaces employed to support teaming are provided.

4.1 Participants' Workload, Situation Awareness, and Perceived Performance

Participants' ratings indicated that workload was higher for the operator-driven team structure compared to the role-driven team structure. This result likely reflects the additional workload associated with managing task assignment in the operator-driven structure as compared to the role driven structure where tasks were automatically assigned to the teammate. Other contributing factors to the higher workload ratings for the operator-driven structure include the requirement for the participant to be aware of available assets for both team members (as resources were limited), as well as their teammate's capabilities, in order to appropriately assign tasks. Moreover, the participants' workload was impacted whenever the teammate sent tasks back to the participant. Some participants even commented that they preferred to complete as many tasks as possible themselves, until they ran out of assets, forcing them to assign tasks to their teammate. In contrast, participants' ratings indicated less workload associated with the role-driven team structure.

These initial results on workload differences between the two team structures are similar to previously reported results examining workload differences between management-by-consent and management-by-exception automation control schemes [41]. The management-by-consent scheme (operator required to make a consent response before autonomy completes the task) is similar to the operator-driven team structure used in the present experiment in that both required the operator to engage in additional decision making and associated control station inputs for task completion. In contrast, with the role-driven team structure and the management-by-exception scheme, the operator has less steps to complete: initial task assignment is automated in role-driven teams and tasks are automatically completed in management-by-exception schemes unless the operator decides to intervene. The similarities in workload ratings between a role-driven team structure and management-by-consent are noteworthy since participants in the present experiment only had to make high-level task/play commands and the simulation's autonomy capabilities handled the workload involved in vehicle allocation and route planning.

The results pertaining to the situation awareness ratings are also of interest. Past research has shown that operators often have decreased situation awareness when task completion is highly automated (i.e., the higher the automation level, the more the operator is "out of loop"). However, initial data in the present study show that despite task assignment being automated in the role-drive team structure, the participants rated their situation awareness higher compared to the operator-driven structure where they were more involved in task assignment. This can be accredited partially to the role-driven structure having a pre-established working agreement prior to the start of missions. Therefore, participants were already aware of what tasks would be automatically assigned to their teammate and those that would be assigned to themselves. Additionally, not having the extra workload managing task assignment provided participants spare capacity to monitor control station displays and this may have increased awareness of the current state of each UV and task completion. The role-driven team structure was also rated better in terms of both perceived individual and team performance. Moreover, participants' frequency in shedding the low priority, but demanding RAM tasks was less when the team structure was role-driven, compared to operator-driven, perhaps also reflecting the spare capacity available without the management overhead of making task assignments. Both team structures, though, showed that shedding RAM tasks was more frequent in the high complexity missions, compared to the low complexity missions.

4.2 Teamwork: Communication and Coordination in Task Sharing

Experimenter observations and task sharing data indicated that when participants became overwhelmed in the high complexity missions, they more frequently assigned tasks to their teammate even when the tasks could not be completed by the teammate. This resulted in the teammate sending those tasks back to the participants, increasing participants' workload even more. This strategy to rely on the teammate in the operator-driven team structure to send tasks back, rather than exert workload to judge whether the task should even be assigned to the teammate, likely reflects participants' inability to maintain awareness of the teammate's capabilities and asset availability during the high complexity missions. This result is also aligned with the lower ratings for the operator-driven team structure, compared to role-driven, on questionnaire items addressing collaborative climate.

The rate of chat communications was slightly higher when the teammate was human compared to being autonomous. However, even in the human-human team configuration, chat communications were fairly limited. These results may reflect features of the training, design of the control station, and participants' tasks. Participants were trained on teammate capabilities and tasks that could be handled by the teammate prior to the experimental trials, eliminating the need for chat communications to obtain this information. Also, the symbology used on the displays was designed to facilitate quick retrieval of status information (e.g., available assets and task assignment) and the Task Manager play-based interface facilitated efficient task assignment. Thus, control station design may have made it more likely that needed information could be efficiently acquired from the displays rather than by taking the time to employ the chat system. Future research will examine the frequency and content of participant-teammate communications with the control station modified to also accommodate voice communications for information exchange. The use of missions with more dynamic, unexpected changes may also prompt more frequent communications to better evaluate the adequacy of communication interfaces.

4.3 Interface Feedback

The refined Task Manager was found to be useful in both team structures regardless of whether the teammate was autonomous or human. Questionnaire comments indicated that the overall look and feel was effective as well as the interface's functionality in supporting task management (e.g., knowing the tasks assigned to each team member and whether they have been completed or not). However, initial questionnaire ratings suggest that the Task Manager supported coordination better for the role-driven team structure compared to the operator-driven structure in terms of understanding the teammate's actions. These ratings may also reflect that the teammate's actions were more defined in the role-driven structure.

Despite the positive ratings for the Task Manager interface, participants had comments on how it could be improved. Some feedback indicated a desire for a more salient indication of task priority, perhaps adding color-coding to the task icons in addition to the row locations used to differentiate task priority. Participants' feedback also indicated that having more task status information would be useful in the Task Manager and Task Log. Desired improvements include task progression status and a mechanism whereby all chat messages/information related to a specific task, beyond the initial chat that created the task (e.g., updates, all clears, etc.), are easily accessed.

5.0 FUTURE RESEARCH

Mixed-initiative systems allowing for autonomy to share in task completion for reasons such as workload management are not inherently good or bad. For the benefits of a mixed-initiative team to be realized, however, the system's design must ensure human-autonomy coordination breakdowns are mitigated and tools are provided

that promote effective teamwork [42]. Thus, interfaces [21] need to be refined in parallel to advancements in the capabilities of autonomy, especially with the vision of a human operator and intelligent agent dynamically collaborating to problem solve and share task completion in a manner similar to effective human teams.

Advancement of intelligent agent technologies is also needed to support reliable operation, provide better autonomy transparency, and support human-autonomy collaboration [5]. The fact that the present experiment had to employ a human confederate to serve as the autonomous teammate illustrates the need for more capable intelligent agent support. Moreover, improvements are essential to enable more dynamic task distribution and trading to help equalize workload across the team and capitalize on the capabilities of each team member. Enhancements are also required such that the autonomy exhibits stable performance in a range of operating conditions and mission events, notifying the human teammate when it is operating outside of its competency bounds, just as an operator would do in a team of humans.

Future development of technology that better supports peer-to-peer type communication between human-autonomy team members is also necessary. For instance, identification of effective interface functionality and bi-directional communication technologies is needed to better support task negotiations between human and autonomy team members. This would allow human-autonomy teams to create working agreements pertaining to task sharing and determine under what conditions task swapping should be entertained or even automated. Design efforts to improve collaboration between a human operator and an autonomous teammate should result in interfaces that inform efforts that are addressing the vision of a well-situated team of multiple humans working together with multiple autonomous systems.

6.0 REFERENCES

[1] Baker, D.P., Day, R. and Salas, E. (2006). Teamwork as an essential component of high-reliability organizations. *Health Services Research, 41*, 1576-1598.

[2] Salas, E., Sims, D.E., and Burke, S.C. (2005). Is there a 'big five' in teamwork? *Small group research,36*, 555-599.

[3] Bradshaw, J.M., Feltovich, P.J., Johnson, M.J., Bunch, L., Breedy, M.R., Eskridge, T. and Uszok, A. (2008). Coordination in human-agent-robot teamwork. *2008 International Symposium on Collaborative Technologies and Systems*, 467-476.

[4] Christoffersen, K. and Woods, D.D. (2002). How to make automated systems team players. *Advances in Human Performance and Cognitive Engineering Research*, 1-12.

[5] Johnson, M., Bradshaw, J.M., Hoffman, R.R., Feltovich, P.J. and Woods, D.D. (2014). Seven cardinal virtues of human-machine teamwork: Examples from the DARPA robotic challenge. *IEEE Intelligent Systems*, *29*(6), 74-80.

[6] Klein, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R. and Feltovich, P.J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*, 91-95.

[7] DeChurch, L.A. and Mesmer-Magnus, R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology*, *95*(1), 32-53.

[8] 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: The Journal of the Human Factors and Ergonomics Society*, *49*(2), 243-256.

[9] Urban, J.M., Bowers, C.A., Monday, S.D. and Morgan Jr., B.B. (1995). Workload, team structure, and communication in team performance. *Military Psychology*, *7*, 123-139.

[10] Cannon-Bowers, J.A. and Salas, E. (2001). Reflections on team cognition. *Journal of Organizational Behavior*, *22*, 195-202.

[11] Salas, E. E. and Fiore, S.M. (2004). Team cognition: Understanding the factors that drive process and performance. *American Psychological Association*.

[12] Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E. and Cannon-Bowers, J.A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273.

[13] Atkinson, D.J., Clancey, W.J. and Clark, M.H. (2014). Shared awareness, autonomy and trust in human-robot teamwork. *Artificial Intelligence and Human-Computer Interaction (AIAA) Spring Symposium Series*.

[14] Ososky, S., Schuster, D., Jentsch, F., Fiore, S., Shumaker, R., Lebiere, C. and Stentz, A. (2012). The importance of shared mental models and shared situation awareness for transforming robots from tools to teammates. *International Society for Optics and Photonics, SPIE Defense, Security, and Sensing, Vol. 8387*.

[15] Espinosa, A., Kraut, R., Slaughter, S., Lerch, J., Herbsleb, J. and Mockus, A. (2002). Shared mental models, familiarity, and coordination: A multi-method study of distributed software teams. *ICIS 2002 Proceedings*, 39.

[16] Jonker, C.M., Van Riemsdijk, M.B. and Vermeulen, B. (2011). Shared mental models. *Coordination, Organizations, Institutions, and Norms in Agent Systems*, *6*, 132-151.

[17] Kleinman, D.L. and Serfaty, D. (1989). Team performance assessment in distributed decision making. *Proceedings of the Symposium on Interactive Networked Simulation for Training*, 22-27.

[18] Michan, S. and Rodger, S. (2000). Characteristics of effective teams: A literature review. *Australian Health Review*, *23*(3), 201-208.

[19] Rothwell, C. and Shalin, V. (2017). Human-Machine articulation work: Functional dependency dialogue for human-machine teaming. *19th International Symposium on Aviation Psychology*, 270.

[20] Harbers, M., Jonker, C. and Van Riemsdijk, B. (2012). Enhancing team performance through effective communication. *Proceedings of the 4th Annual Human-Agent-Robot Teamwork Workshop*, 1-2.

[21] Kilgore, R. and Voshell, M. (2014). Increasing the transparency of unmanned systems: Applications of ecological interface design. *International Conference on Virtual, Augmented and Mixed Reality*, 378-389.

[22] Urban, J.M., Weaver, J.L., Bowers, C.A. and Rhodenizer, L. (1996). Effects of workload and structure on team processes and performance: Implications for complex team decision making. *Human Factors: The*

Journal of the Human Factors and Ergonomics Society, 38(2), 300-310.

[23] Bowers, C.A., Urban, J.M. and Morgan Jr., B.B. (1992). The study of crew coordination and performance in hierarchical team decision making. Technical Report No. TR-92-01. Orlando: University of Central Florida.

[24] La Porte, T. and Consolini, P. (1998). Theoretical and operational challenges of "high-reliability organizations": Air-traffic control and aircraft carriers. *International Journal of Public Administration, 21*, 847-852.

[25] Draper, M., Rowe, A., Douglass, S., Calhoun, G., Spriggs, S., Kingston, K. … Frost, E. (2018). Realizing autonomy via intelligent adaptive hybrid control: Adaptable autonomy for achieving UxV RSTA team decision superiority (also known as Intelligent Multi-UxV Planner with Adaptive Collaborative Control Technologies (IMPACT). (2018). Final Report. AFRL Technical Report (AFRL-RH-WP-TR-2018-0005). Wright-Patterson Air Force Base, OH.

[26] Kingston, D., Rasmussen, S. and Mears, M. (2009). Base defense using a task assignment framework. *AIAA Guidance, Navigation, and Control Conference*, 6209.

[27] Douglass, S. (2013). Learner models in the large-scale cognitive modelling (LSCM) initiative. In R. Sottilare, A. Graesser, X. Hu and H. Holden (Eds.), *Design Recommendations for Intelligent Tutoring Systems, Volume 1*, 111-126. Orlando, FL: U.S. Army Research Laboratory.

[28] Dzieciuch, D., Reeder, J., Gutzwiller, R.S., Gustafson, E., Coronado, B., Martinez, L., Croft, B. and Lange, D.S. (2017). Amplifying human ability through autonomics and machine learning in IMPACT. *Proceedings Society of Photo-Optical Instrumentation Engineers (SPIE), Micro- and Nanotechnology Sensors, Systems, and Applications IX, Volume 10194*, p. 101941Y. International Society for Optics and Photonics.

[29] Draper, M., Calhoun, G., Hansen, M., Douglass, S., Spriggs, S., Patzek, M., Rowe, A. …Frost, E. (2017). Intelligent multi-UxV planner with adaptive collaborative/control technologies (IMPACT). *Proceedings of the International Symposium on Aviation Psychology*. Dayton OH: Wright State University. https://isap.wright.edu/isap-pubs.

[30] Roscoe, A.H. and Ellis, G. A. (1990). A subjective rating scale for assessing pilot workload in flight: A decade of practical use. Technical Report No. RAE-TR-90019. United Kingdom: Royal Aerospace Establishment.

[31] Miller, C.A. and Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49*(1), 57-75.

[32] Mersch, E.M., Behymer, K.J., Calhoun, G.L., Ruff, H.A. and Dewey, J.S. (2016). Game-based delegation interface design for unmanned vehicles: Color coding and icon row assignment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60*, 122-126.

[33] Calhoun, G.L., Ruff, H.A., Behymer, K.J. and Mersch, E.M. (2017). Operator-autonomy teaming interfaces to support multi-unmanned vehicle missions. In P. Savage-Knepshield and J. Chen (Eds.),

Advances in Human Factors in Robots and Unmanned Systems: Advances in Intelligent Systems and Computing, 499, 113-126. Switzerland: Springer.

[34] Calhoun, G., Ruff, H., Behymer, K. and Frost, E. (2018). Human-autonomy teaming interface design considerations for multi-unmanned vehicle control. *Theoretical Issues in Ergonomics Science, 19*, 321-352.

[35] Hansen, M., Calhoun, G., Douglass, S. and Evans, D. (2016). Courses of action display for multi-unmanned vehicle control: A multi-disciplinary approach. *Artificial Intelligence and Human-Robot Interaction (AIAA) Fall Symposium Series*.

[36] Burger, J.M. and Cooper, H.M. (1979). The desirability of control. *Motivation and Emotion, 3*(4), 381-393.

[37] Merritt, S.M. and Ilgen, D.R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors, 50*(2), 194-210.

[38] Gosling, S.D., Rentfrow, P.J. and Swann, W.B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504-528.

[39] Brooke, J. (1996). SUS: a 'quick and dirty' usability scale. In P. Jordan, B. Thomas, I. McClelland, and B. Weerdmeester (Eds.), *Usability Evaluation in Industry*. Bristol, PA: Taylor and Francis Inc.

[40] Jian, J.Y., Bisantz, A.M. and Drury, C.G. (1998). Towards an empirically determined scale of trust in computerized systems: Distinguishing concepts and types of trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 42*(5), 501-505.

[41] Ruff, H.A., Narayanan, S. and Draper, M.H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence: Teleoperators & Virtual Environments, 11*(4), 335-351.

[42] Borst, C., Flach, J.M. and Ellerbroek, J. (2015). Beyond ecological interface design: Lessons from concerns and misconceptions. *IEEE Transactions on Human-Machine Systems, 45*(2), 164-175.

Bi-directional Communication in Human-Autonomy Teaming

Robert J Shively
NASA-Ames Research Center
Moffett Field, CA
USA

robert.j.shively@nasa.gov

Gilles Coppin
Telecom Bretagne
Breast
FRANCE

gilles.coppin@imt-atlantique.fr

Joel Lachter
NASA-Ames Research Center
Moffett Field, CA
USA

joel.lachter@nasa.gov

ABSTRACT

Communication between autonomous agents and humans is critical for effective Human Automation Teaming (HAT). This paper investigates the Human Factors issues, technological challenges and implementation issues that are associated with bi-directional communication to facilitate comprehension. In HAT, as well as human teams, good communication is paramount to coordinate task planning and execution. It is important that all necessary information be communicated without excessive chatter that can divert attentional resources from task performance. Human autonomy teams face increased challenges in this novel environment such as the definition of a mechanism to communicate with the autonomy (keyboard input or natural language), and defining a shared language that both humans and automation can understand. A paradigm shift is required to move from using autonomous agents as tools to treating them as teammates. Because communication is central to effective teaming, it would be useful to standardize methods for operators to communicate with automation that can be reused with autonomous systems of varying types. Such methods might be analogous to the system of windows, menus, and files that became the standard metaphors for interaction with computers on the 1980s. In this paper we identify issues that bi-directional communication may address and issues that need to be addressed for effective communication between the operator and autonomous agents. Bi-directional communication can help with understanding the capability of the automation, self-confidence in particular solutions, provide transparency to help calibrate trust and through those mechanism even help with agility to reduce brittleness. As mentioned, an effective channel of communication must be established and the content of information must be defined to provide shared situation awareness. Consideration must be given to both what needs to be communicated (context, options, authority), and the manner in which it is best communicated (how and when). A "bi-directional communications manager" is proposed, that is capable of tracking goals, plan progress, and user mental state, to determine when communication is necessary. Such a communications manager could be

customized for use with a wide variety of autonomous systems. Systems with such a manager would be far more interactive than current automation, encouraging iterative planning in which both automation and human operators contribute as well as joint execution and monitoring, in much the same way that crew resource management proscribes procedures that assure both pilots monitor the situation and vet plans in current commercial aviation. As autonomous systems develop, it is clear that humans need ways to understand and team with them effectively. At the same time, current automation has proven brittle, requiring human intervention in off-nominal situations. More comprehensive communication between humans and automation can help solve both of these problems. We believe that a standardized communication system for HAT will increase usability and adoption of advanced automation.

1.0 HUMAN AUTONOMY TEAMING (HAT)

For hundreds of years automation has been taking on new tasks that increase exponentially in complexity. Recently these advances appear to be closing in on a new threshold, one where the automation takes on tasks that require it to act autonomously in the world. Autonomous cars promise to save us the drudgery of our daily commutes and autonomous aircraft promise to move us and our stuff around without traffic constraints. However, there are still hurdles to overcome. Such systems are brittle, working properly within some bounded space for which they have been programmed, then failing when parameters fall outside that space. They lack transparency; human operators often do not know what the automation is doing or why. Operators often do not know when to trust automation, relying on it to handle conditions it cannot, or not taking advantage of it to handle conditions it can. When the automation performs a task, the operator is often less aware of the system state, causing confusion if the operator needs to step in. Compounding this, as automation does more of the work, operators become less practiced and then are unable to perform correctly when the automation fails.While each of these issues is troubling by itself, they often manifest together. An operator, over-trusting the system, does not realize that some parameter has gone out of bounds. Because it is out of bounds, the automation either quits or is no longer reliable. The out of practice operator must then try to regain situation awareness using opaquely presented information. This situation has been responsible for a number of accidents (e.g., Air France 447 [1], Korean Air 801 [2]). Similar concerns have been identified by others [3-7]. A number of authors have suggested that these concerns are ameliorated by developing interfaces and procedures analogous to guidelines for improving teamwork among humans [4, 7, 8, 9]. This area of research has been termed “Human-Autonomy Teaming” or “HAT.” [7,9] HAT is a simple concept: treat automation as a teammate, rather than a tool. This will help people get the most out of the automation and work with it more naturally. However, as with many simple concepts, the implementation can be complex. The next section outlines some of the key attributes that will allow users to move toward treating automation as a teammate.

1.1 Key Attributes of HAT

1.1.1 Pilot directed interface

Previous work has focused on task allocation between the humans and automation. However, this static relationship can lead to non-optimal performance. At the same time, if the automation re-allocates tasks, the operator can become disoriented. A well planned and understood allocation strategy coupled with an operator directed dynamic allocation of tasks based on workload, time pressure and other important factors allows a much more agile, flexible system. One potential enabler of such a dynamic system of task allocation is the concept of a play [10]. A play encapsulates goals, tasks, and a task allocation strategy appropriate for a particular situation. Operators can call a play to quickly adapt to a new situation. A pilot directed interface also facilitates

understanding by explicit discussion of goals (as opposed to intent inferencing), as well as confidence, and rationale.

1.1.2 Transparency

Lyons [11] argues that the system and its operators should be able to recognize what the human/automated teammate is doing and why. He defines automation transparency as the enabler of such recognition. Transparency is more than simply making the information available to the human operator. Often the calculations done by automation do not correspond directly to those a human would do when performing the same task. To be transparent, the automation must bridge this gap by presenting information in ways that fit the operator's mental model. Transparency helps correctly calibrate trust in the sysem, which is a critical component, see below.

1.1.3 Bi-directional communication

Teammates often discuss options, brainstorm on solutions and openly discuss courses of action. For automation to be on a team, this bi-directional communication needs to exist. We see bi-directional communication as key to solving a number of the issues typically found in highly automated systems. Bi-directional communication can make systems less brittle. History has shown that humans, generally, are better able to adapt to unfamiliar situations, poorly structured problems, or situations with incomplete information. With HAT, the human can provide the missing information and context, apply judgment from experience in similar situations that would not be recognized as relevant by the automation (Christofferson and Woods [4] have shown how automation does not generalize well from one domain to another) and if necessary, override the decisions made by automation in these situations. Bi-directional communication can also help to build shared awareness. Alberts and colleagues [19] define shared awareness as a cognitive capability that builds on shared information. Shared awareness of an event can be developed through four mechanisms: (1) directly, (2) though independent sensors, (3) through information that is passed from one agent to another, and (4) through information that is shared and the fused results presented to two (or more) agents. The last two, sharing and fusing of information require a communication channel through which information can pass in either direction. And, while the same information could be gathered independently (mechanisms 1 and 2), to assure shared awareness, this information needs to be cross-validated or eventually the information will be out of sync. Several of the key attributes of HAT are enabled by bi-directional communication. It is of paramount importance, therefore to ensure that this communication is effectively supported.

2.0 THE STRUCTURE OF BI-DIRECTIONAL COMMUNICATION

2.1 Definition

When bi-directional communication is discussed in the context of HAT, it is meant to convey the kind of communication that occurs between human team members. In such communication, each participant provides new information that the other might not have, each can offer hypothesis and pose "what if" questions, and each can particpate in exchanges before a shared view is arrived at. This kind of two-way, natural bi-directional communication is the key to effective teaming between humans and in-turn, effective teaming between humans and autonomy. A few key components are required for this communication and these are discussed below.

2.2 Channel

Some mechanism must be available to host the communication, that is, some kind of two-way channel. Ideally, this would be as natural as possible for the human operators. In some cases the most natural interface might be a natural language. However, this is not required. For some purposes machine speech recognition and generation are not up to the task. Further, speech is not always the best way of communicating, even among humans, who often point, draw diagrams, or physically guide each other. Applying pressure to a throttle or drawing a path on a sketch pad may be more natural ways of communicating particular ideas or intent. The critical aspect is that the human and the automation share a channel that they use to communicate. They need to able to input information, ask questions and otherwise interact through this channel.

2.3 Language

Bi-directional communication requires some ability to have a common cognitive understanding and for both the human and automation to communicate their intentions [12]. To have true bi-directional communication, a shared language will be required. Because computers and humans do not share an internal "language of thought", translation will be necessary. Ideally, most of the translation will fall to the automation; however, it may be necessary for the human to learn some formalization. The "round trip" nature of bi-directional communication provides a useful check that such translations have been successful. Figure 1 provides an example where the operator gives an initial command using a mix of natural language and gestures which the automation then interprets**.** The resulting interpretation is then displayed for the operator to validate. A shared communication channel allowing both implicit and explicit communication between the human and agent, is required.

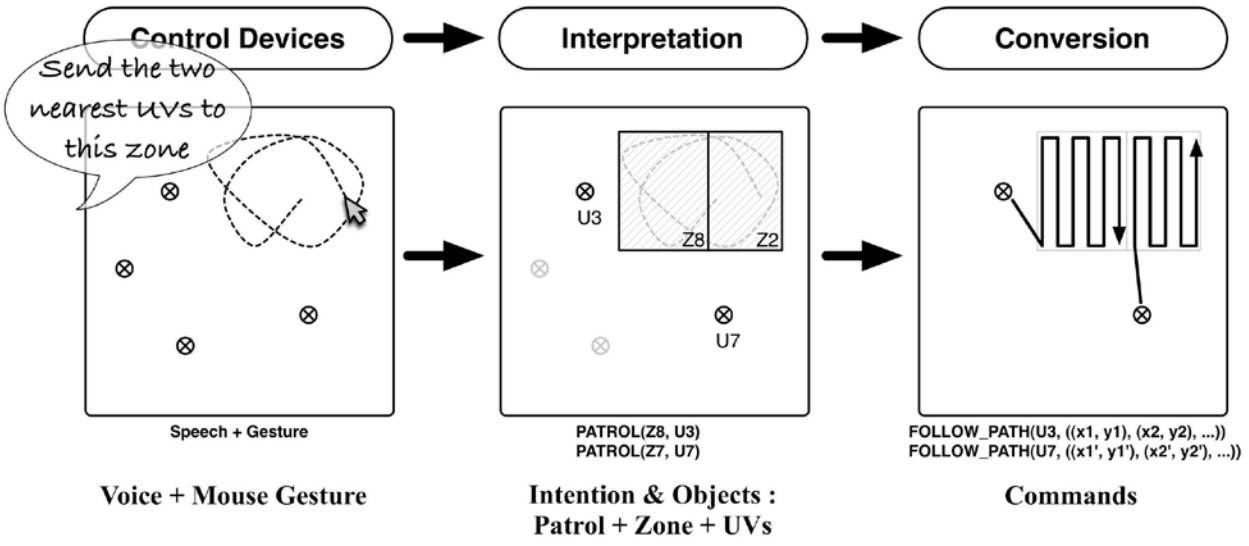


Figure 1: Communication Flow.

An example of such a language is the cockpit control language (CCL) developed by Riley [13]. CCL was based on the structured communications pilots use to talk to controllers and thus allowed pilots to easily translated controller instructions into commands for the flight management system. Such a codified, agreed upon language and syntax that would be understandable to humans as well as automation is critical to effective HAT.

2.4 Temporal Coordination

Most of the systems currently being discussed require real-time communication, or, at least, minimal delay. This can be critical to being able to effectively use teaming to make joint decisions. Systems operating in some environments such as under water or in deep space can have significant delays forcing the automation to largely operate independently.

2.5 Symmetry

The backbone of bi-directional communication is that both agents (human and computer) can input information, request information, and monitor the other agent; in short, the system needs to support symmetrical communication.

2.6 Examples

2.6.1 Everyday example: Navigation systems

We are probably all familiar with the navigation systems on our phone. They differ somewhat, but share many qualities. In general, these systems are extremely helpful and many of us don't know how we traveled without them. However, with HAT, they could be dramatically more useful, like a human navigator with knowledge of the area, traffic, weather, construction and all the factors that can affect your drive.

Currently, these systems will tell you that traffic is ahead and possibly the cause; an accident or a police action, for example. Then the system may (or may not) provide you with an alternate route (usually just one). That is about the extent of the (largely) one-way communication.

If you had a dialog with the system, you could ask several types of questions, bounce ideas or hypothesis off and interact more fully and naturally with the system. Perhaps you'd like to know where the accident is and ask how long taking an alternate route that you define would take. Or, you might have a prefered way to drive to work, and that doesn't show up as a recommendation, you can ask why not. This ability will allow you to treat this automation like you would treat a human teammate. Why are you saying that? What about this route?

The components of this bi-directional dialog are discussed below.

2.6.1.1 Channel

In this case, the channel may be verbal, a very natural channel for the human. This might be esspecially useful while driving or otherwise occipied. However, that requires automatic speech recognition and a natural language interface. While ideal for a human, the maturlity of this technology may limit its effectiveness. A keyboard and visual display interface may be serve the same function (althought not while driving!)

2.6.1.2 Language

The language may be normal spoken word, which again is optimal for humans. However, for the automation to parse and understand the semantics, it's useful to define syntax and even approved vocabulary.

2.6.1.3 Temporal coordination

This application would require real-time or near realtime porcessing and interface.

2.6.1.4 Symmetry

As indicated, this application would require the ability of both agents (human and automated) to be able to initiate, respond and build-on communications.

2.6.2 Aviation Example

Investigations as NASA recently looked at single pilot operations in commercial airlines. Questions included the workload of the single pilot and how automation may reduce that workload. In one instance, a flight was en route to San Francisco and a medical emergency arose which required a diversion to an alternate airport. The pilot was aided by an emergency landing planner system that when faced with a diversion would recommend an alternative airport. In this example, the system recommends diversion to Cheyenne, Wo. However, like many current systems, this is the end of the one-way communication.

This leaves the pilot without knowing why Cheyenne was chosen over Denver. Or how close other alternatives were - were alternatives just separated by a few numbers on a huge scale? What if the pilot knows that Denver has better medical facilities or has better company facitlities? How can he/she input that data to the system. Again, there is a need for a dialog between the agents in the system. The pilot should be able ask these questions, input data, ask what if questions and generally treat the automation like he/she would a teammate.

The components of this bi-directional dialog are discussed below.

2.6.2.1 Channel

In this case, the channel may be verbal, a very natural channel for the human. However, the auditory channel is quite busy when flying, so a keyboard/visual display may be more compatible or perhaps a graphic editing tool coupled with the keyboard. The pilot might circle an area, then type in “what airports are avialalbe in this area?” This type of hybrid interface can be very effective.

2.6.2.2 Language

Aviation brings with it its own grammar and syntax, for example when a pilot speaks to air traffic control. Building on that would be compatible with the humand and much eaiser for the automation to understand.

2.6.2.3 Temporal coordination

This application would require real-time or near real-time processing and interface.

2.6.2.4 Symmetry

As indicated, this application would require the ability of both agents (human and automated) to be able to initiate, respond and build-on communications.

3.0 DISCUSSION

In order to communicate, the system must be first able to compute all the necessary elements such as the level of confidence (of its own proposals of decision or action), a course of action that can legitimate its communication and related decisions, its "own" expression of goals and decision consequences. This goes through the ability of learning from the user (classification of current functional state and cognitive state of the user) and from the situation. All these abilities mostly rely on Artificial Intelligence and Learning.

New capacities of autonomous systems lead to a paradigm shift as the system has now the possibility of building its own representations of world (the situation), self (system self evaluation) and the partners (learning from the user), from which it can propose or take decisions. A first step before decision concerns the communication of these representations towards the user as a teammate would do, while traditional systems are only mirroring the unique user's knowledge and representations.

The existence of these knowledge representations opens the perspective of using many of the features of human dialogs (etiquette, emotion, all performative - or perlocutionary - linguistic acts) and to develop efficient communication in the framework of HAT.

For the locutionary acts, we can refer to an existing framework proposed proposed by authors such as [Hoc, Chauvin, Flemisch] for human- human cooperation. Especially, teaming considered as cooperation needs to rely on a jointly structured ground (common ground or COFOR for Hoc) addressing views on external situation, hints on internal situation (functional states of team members) and elements of meta-knowledge (such as respective roles of team members), and to communicate bidirectionally on these elements.

As systems become more and more autonomous, HAT will become more and more critical and bi-directional dialog is key to impllmenting effective HAT.

4.0 REFERENCES

[1] Langewiesche, W. The human factor. *Vanity Fair*. Retrieved from http://www.vanityfair.com/business/2014/10/air-france-flight-447-crash on September 7 2018 (2014)

[2] National Transportation Safety Board. *Controlled Flight Into Terrain, Korean Air Flight 801, Boeing 747-300, HL7468, Nimitz Hill, Guam, August 6, 1997*. Aircraft Accident Report NTSB/AAR-00/01. Washington, DC. (2000)

[3] Parasuraman, R., Riley, V.: Humans and Automation: Use, Misuse, Disuse, Abuse. Hum. Factors. 39, 230–253 (1997)

[4] Christoffersen, K., Woods, D.D.: How to Make Automated Systems Team Players. In: Advances in Human Performance and Cognitive Engineering Research, vol. 2, 1–12. Emerald Group Publishing Limited (2002)

[5] Lee, D.D.: Review of a Pivotal Human Factors Article: "Humans and Automation: Use, Misuse, Disuse, Abuse. Hum. Factors. 50, 404–410 (2008)

[6] Parasuraman, R., Manzey, D.H.: Complacency and Bias in Human Use of Automation: An Attentional

Integration. Hum. Factors. 52, 381–410 (2010)

[7] Endsley, M.R.: From Here to Autonomy: Lessons Learned from Human-Automation Research. Hum. Factors. 59, 5–27 (2017)

[8] Chen, J.Y.C., Barnes, M.J.: Human-Agent Teaming for Multi-Robot Control: A Literature Review. Army Research Lab Technical report, ARL-TR-6328 (2013)

[9] Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V., & Johnson, W. W. Why Human-Autonomy Teaming? To be presented at the 8th International Conference on Applied Human Factors and Ergonomics (AHFE 2017), Los Angeles, CA. (2017)

[10] Miller, C.A., Parasuraman, R.: Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. Hum. Factors 49, 57–75 (2007)

[11] Lyons, J.B.: Being transparent about transparency: a model for human robot interaction. In: AIAA Spring Symposium Series (2013)

[12] Chen, J.Y.C., Barnes, M.J., Harper-Sciarini, M.: Supervisory control of multiple robots: human performance issues and user interface design. IEEE Trans. Syst. Man Cybern.–Part C: Appl. Rev. 41, 435–454 (2011)

[13] Riley, V., DeMers, B., Misiak, C., & Schmalz, B. The cockpit control language: A pilot-centered avionics interface. *Air & Space Europe, 1*(1), 48–50 (1999)

Human Autonomy Teaming using Cooperative Automation, Interaction Patterns and Image Schemes

Marcel C. A. Baltzer, Daniel López, and Frank Flemisch
Fraunhofer FKIE, Dept. Balanced Human Systems Integration
Fraunhoferstraße 20, D-53343 Wachtberg
GERMANY

Marcel.Baltzer@fkie.fraunhofer.de

ABSTRACT

Human Autonomy Teaming needs clear patterns that can be understood by all human and machine team members in order to become really cooperative. The proposed contribution addresses this challenge by linking the paradigm of cooperative automation with interaction patterns and image schemes. Image schemes derived in psychology and describe the fundamental linking principles of human (embodied) cognition and the world. As an example, cooperative automation, interaction pattern and image schemes are applied to heavy military ground vehicles. Military vehicles like heavy trucks, tanks or excavators face the challenge of being sufficiently guarded against enemy fire and being safely manoeuvrable. The latter includes driving capability but also vision in order to see where the vehicle is driving. Current military vehicles therefore face a trade-off between optimal vision and optimal armour resulting in more obstruction of the driver's sight and the necessity of an assisting or automating co-driver when using a heavily armoured truck, tank or excavator. In order to overcome such limitations the human co-driver can be replaced by a cooperative automation that supports the driver in the driving tasks as an autonomous but cooperative team member in a cooperatively guided vehicle. To efficiently communicate information between co-automation and the crew, interaction patterns and AR was applied.

1.0 INTRODUCTION

Cooperation becomes necessary when two or more actors with the ability to autonomous behaviour are supposed to work towards a common goal. A successful cooperation is characterized by a consistent, commonly accepted result generated by the actors. If those actors do not work cooperatively, the best result would be two results and the worst none at all. Cooperation is not useful in all situations, e.g., in trivial tasks or tasks with a low resource demand, or in competitive circumstances. Still cooperation makes sense when several aspects of a task require a multiplicity of abilities or expertise like in interdisciplinary teams that work on complex tasks. Furthermore, cooperation makes sense, if the workload for a single actor gets too high and could be reduced by a better balance between several actors.

When talking about cooperative movement, the cooperation or collaboration of at least two actors with the goal of changing the current location is addressed. In Cooperative Guidance and Control a human operator works together with a cooperative automation, which we call co-automation or co-system. Cooperation is enabled by successful interaction. In our concept, interaction is facilitated using an Interaction Mediator [1] see Figure 1.

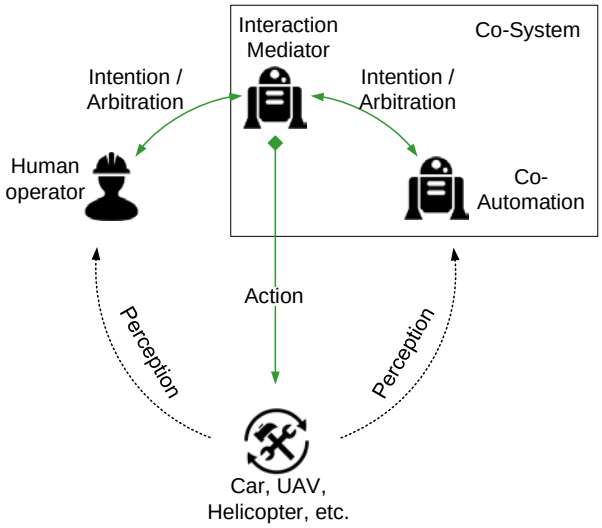


Figure 1: Cooperative Guidance and Control, cf. [2]

The automation of such a cooperatively guided manipulator, like e.g., a highly automated car, UAV or helicopter, can either assist the driver, e.g., to maintain an energy efficient state or even take over more advanced movement tasks such as lateral and longitudinal control in varying degrees [3]. In the military domain, these aspects can be extended with system qualities like safety and survivability. When driving for a certain time in a convoy, the workload of many drivers can be reduced to only the workload of the leading vehicle driver, where the others can observe the environment for threats.

1.1 Cooperative Automation

[4] already distinguished "Cooperation in action", cooperation in plan and meta-cooperation, which is not directly concerned by current control. Shared control seems to focus on the common task or function on the operational level, e.g. control, while cooperation adds the way to take increasingly into account the other agent and the other levels (see Figure 2).

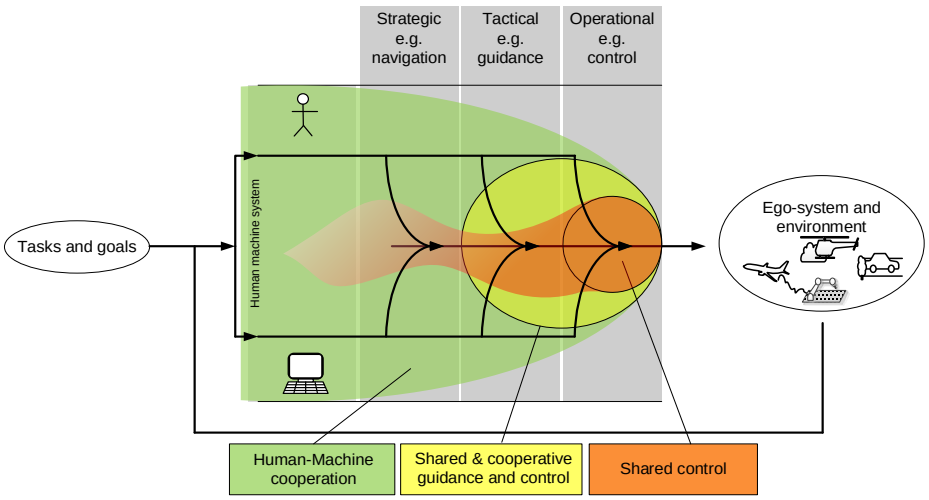


Figure 2: Relationship between the Shared control, shared and cooperative Guidance and

control, human-machine cooperation (cf. [5])

More than to know what the other is doing, cooperation allows to have a model of the partner to know how it is possible to cooperate with it/him/her. A cooperative agent (that has know-how-to-cooperate) can gather information about the other, analyse this information to decide about their cooperation. Such activity can again go directly back to the operational level, e.g., with shared control. But such cooperative activity can and should be prepared at the tactical and strategic levels.

An effective way to create cooperation between human and automation, is that human and machine have the same internal model of what has happened, what is currently happening and what will happen. This kind of compatible internal models is called cognitive or inner compatibility [6].

Known models to improve inner compatibility were proposed by [7], [8], [9], and [10], that distinguish navigation, guidance and control level of the driving task. [11] combined these to a simplified model of human information processing while guiding a vehicle with the guidance level split into long term (manoeuvre) and short term (trajectory) planning in order to enable technical realisation (see Figure 3).

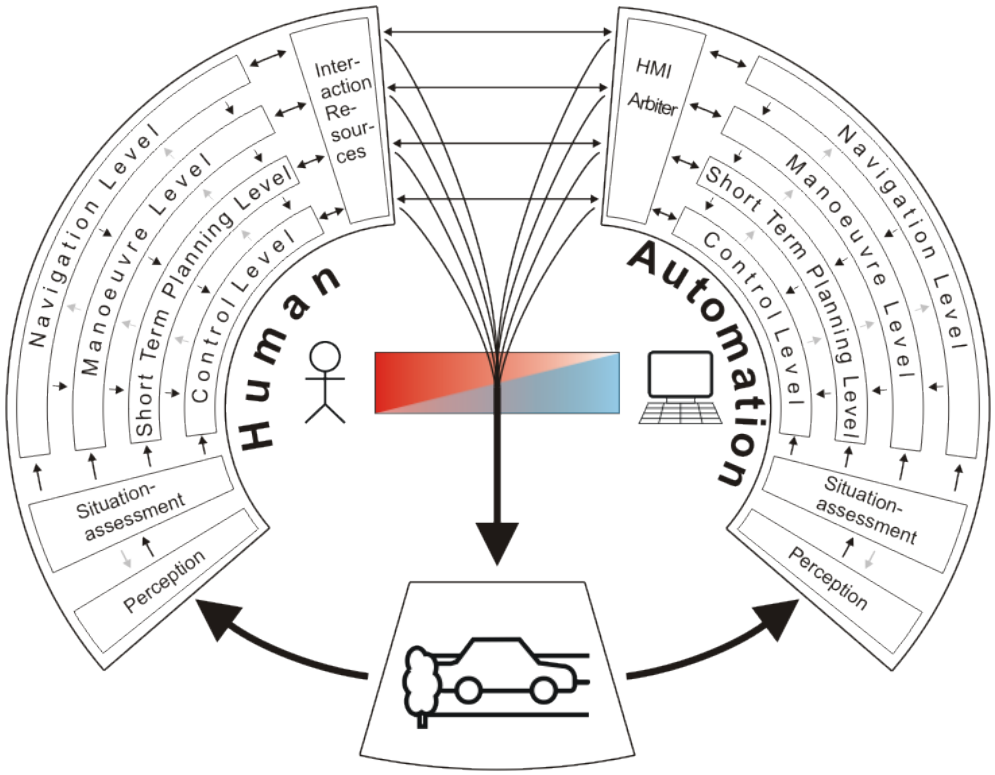


Figure 3: Human and Cooperative Automation (cf. [11])

In this model compatibility results from implementing the structure of human planning and acting into the automation. On the human side perception works over the sensory modalities, whereas the automation needs suitable sensors. Situation assessment is the interpretation of the situation and its projection into the future.

The planning phase can be differentiated into four levels: Navigation, representing the route through the road network; Manoeuvre, semantically determining what next manoeuvre to perform in order to comply to the

route; Short term planning, to generate the trajectory that is according to the manoeuvre; Control, representing the specific commands (steer angle, gas) to conform with the respective trajectory. In order to generate common actions, the intention of human and machine need to be communicated and arbitrated via the interaction resources by the HMI Arbiter on all 4 levels. How these actions work is determined by the respective automation level.

The automation spectrum shows multiple possibilities of task distribution. Examples of mode configurations are seen in Figure 4.

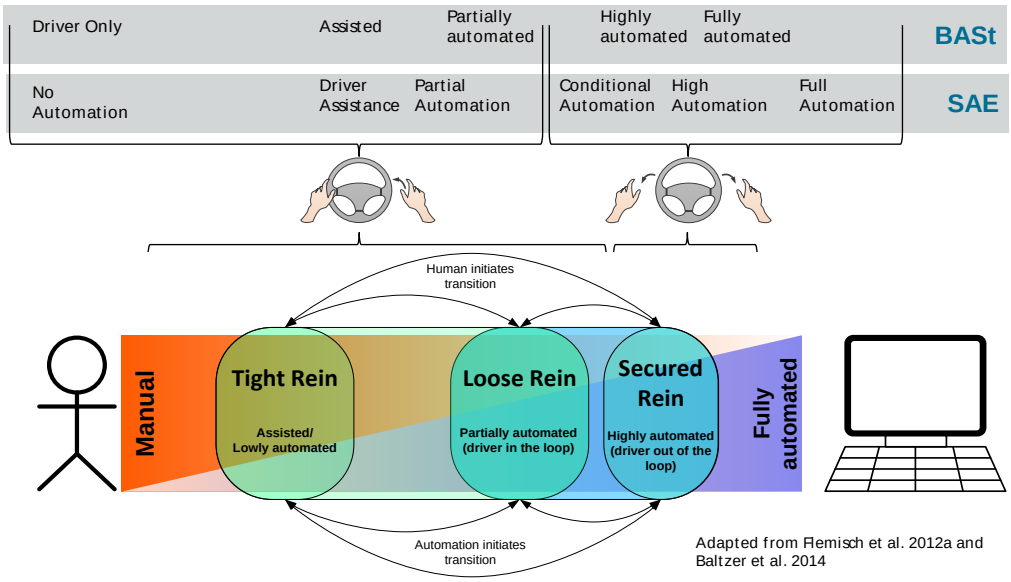


Figure 4: Possible Mode Configurations in Cooperative Guidance and Control (cf. [1])

As a means to create intuitive interaction between human and co-automation, the interaction mediator uses interaction patterns that will be explained in the following.

1.2 Interaction patterns for Cooperative Guidance and Control

Interaction patterns are solutions to recurring interaction problems described by their use cases. Interaction Patterns are described by the flow of information from the Source domain that results in the change of the user´s understanding of one of more internal target states (see Figure 5). These states can be e.g. an understanding of power or urgency. This flow of information can happen in many ways, we use image schemas to guide this information in a more natural way that can be related to previous embodied experiences.

Image Schemas originate from our experiences with humans, objects and events in our environment. When visiting restaurants for example we start to generalize things and develop abstract generic expectations about what we can expect from other restaurants. Therefore, when somebody tells us from their restaurant visit, we do not need to know all details like having paid etc., since we have our own schema of such a restaurant visit filling the missing information with our own experiences.

In another sense Image Schemas are not specific to a certain sensory modality [13], [14]. The so created ambiguity of being abstract (they are schematic) and not abstract (they are embodied) can be solved by understanding image schemas as abstract repeating representations of dynamic patterns of embodied

interaction with our environment, that structure our understanding of how the world works. Their strength for human machine interaction design lies in their metaphorical extension (e.g., importance, power, etc.), so they can transmit more meaning than just affordances [15].

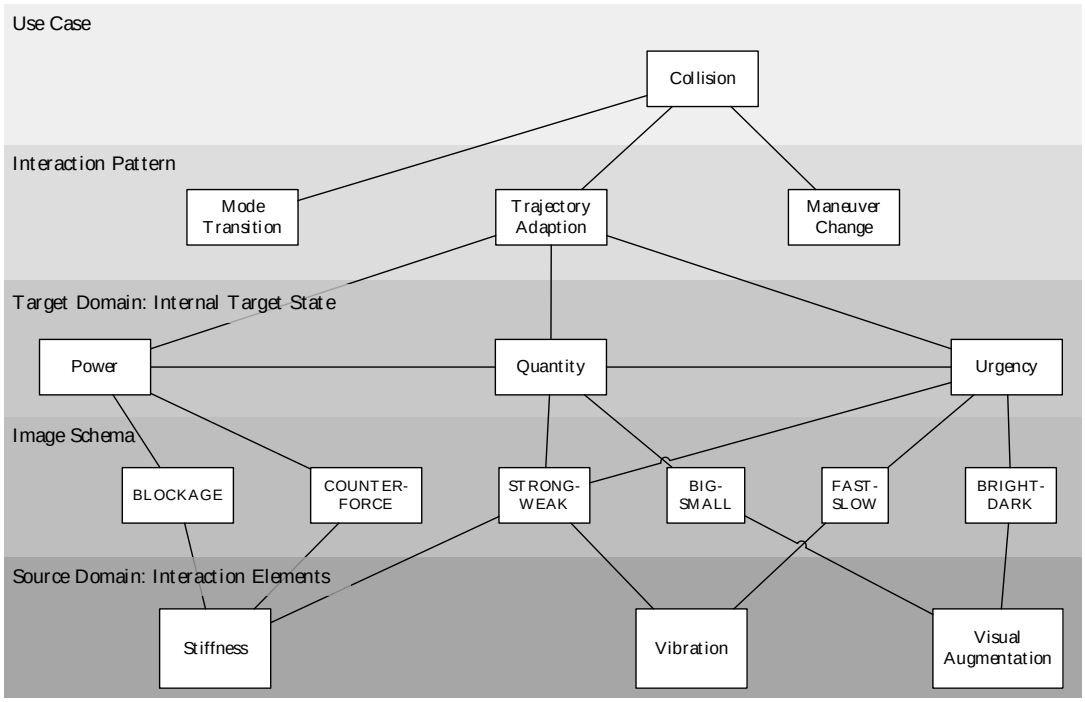


Figure 5: Information flow related to interaction patterns for Cooperative Guidance and Control (c.f. [12])

A simple example of an Image Schema is the CONTAINER schema that is built from the structural elements of an INTERIOR, a BOUNDARY and the EXTERIOR [16]. Human understand their body as a CONTAINER with their skin (BOUNDARY) delimiting their inner organs (INTERIOR) from the environment (EXTERIOR) [16]. This schema can be used for buildings, bottles and a Local Area Network. In terms of design they can be used to transfer meaning between the source domain, e.g., sensory modalities, and the target domain, e.g., speed of action, progress, importance or intensity.

Image Schemas work as a translation between specific bodily experiences from the source domain, i.e., via interaction elements and the abstract target domain, i.e., internal target states. Interaction elements can be described as implementations that generate a certain image schema, e.g., a large stiffness in the gas pedal creates the understanding of a BLOCKAGE etc.

Interaction patterns must have a solution: an actor reaction, this reaction can be an acknowledgement, a movement or a trigger for another action. The more obvious the actor´s reaction, the easier it is to see if a pattern works as intended.

An Implementation of such an interaction pattern is presented in the following section.

2.0 INTERACTION PATTERN "COLLISION AVOIDANCE"

2.1 Name

Collision Avoidance.

2.2 Problem description

The main problem addressed by the interaction pattern is to prevent collisions with objects in the vehicle's proximity. There are at least two tension poles from the perspective of the ego-system between which a balance needs to be achieved: approaching the obstacle(s) and keeping at distance of (or deviating around) the obstacle(s). The awareness might be low, perception might be reduced. The problem can be visualized, see Figure 6.

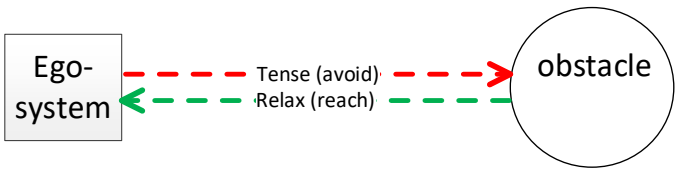


Figure 6: Abstract problem description for the interaction pattern "collision avoidance". Upper red arrow represents tensing action due to approaching the obstacle. Lower green arrow represents relaxing action due to receding from the obstacle. The balancing action is to avoid the obstacle and reach a safe distance (cf. [19]).

2.3 Solution

A stepwise escalation can lead to a higher awareness that the danger in the current situation is increasing when continuing with the current action or non-action. Also escalation offers the opportunity to react on user action, hence a positive reaction to behaviour adaption towards the non-danger tension pole. Depending on the relative speed and the escalation phase a respective obstacle avoidance manoeuvre should be chosen, e.g. an implemented automated lane change pattern or an implemented emergency brake pattern. To make both avail-able a mode transition pattern is necessary to decouple the driver.

The escalation should be triggered depending on the TTC in situations with rather high relative velocities, and triggered depending on distance in situations with rather low relative velocities. The result of the pattern implementation is an improved awareness of the target domains Variation and Importance. Variation of the current system state and Importance to change the current behaviour.

2.4 Consequences

The interaction pattern addresses a reduction in collisions with near objects and accordingly will reduce lateral displacement from the centre of the lane. Due to a reduced task load, the Situation Awareness of drivers will be enhanced. The solution focuses on adaption management, therefore certain internal target states, like Variation and Importance, are addressed. These domains are addressed, since an action needs to be made to avoid a dangerous situation that is going to happen in the current course of (non)action. Challenging consequences are the aspect of overtrust or overreliance in a non-perfect system. Also connected to non-perfect systems are wrong escalations due to falsely detected objects that may negatively affect acceptance. Finally an overreaction by the user could be observed when escalation phases are too small.

2.5 Implementation examples

The target domain of Variation can be addressed using the image schema PATH. Structural elements of PATH are a start, an end and a direction, [14] and [15]. The PATH schema also includes a series of locations [16] that can be interpreted as escalation steps. Symbolic qualities addressing Importance can be implemented with colour codes following the pattern from traffic lights, where red means stop or danger and green means go or safe [17]. These colour codes can be emphasized using the BRIGHT-DARK image schema [18].

The start and end locations or phase limits need to be determined with relevant variables. In a situation with rather high relative velocities, e.g. following a moving car, distance is not sufficient information to determine the need for behaviour adaptation. In situations with high relative velocities the need for reaction is farther away than in situations with low relative velocities. Therefore a combination of relative velocity and distance as time to collision (TTC) seems to be a valid concept to determine escalation boundaries in situations with rather high relative velocities. In more static situations, e.g. parking or slowly driving in a narrow road, the TTC becomes unfeasible due to extremely low relative velocities and the need for very close approaches, so that distance seems to be a useful concept in such situations. Another variable that takes the current driver's behaviour into account is the current steering angle. Therefore depending on the situation the same pattern can be used with different escalation variables and boundaries.

The respective implementations for a collision avoidance pattern in a situation with rather low relative velocities are visualized in Figure 7, for forward collisions, and Figure 8, for side collisions or road departure.

The PATH image schema is implemented having three distinctive steps or locations: Low urgency, medium urgency and high urgency. The respective variable when a certain urgency level is reached depends on the distance between the ego vehicles bounding box and the obstacle's bounding box as well as the current steering angle. Depending on the ego vehicle's action a Variation of Importance is defined by the escalation phase.

In Figure 7, we see the implementation of the collision avoidance pattern in a forward collision avoidance assistance when approaching a parked truck.

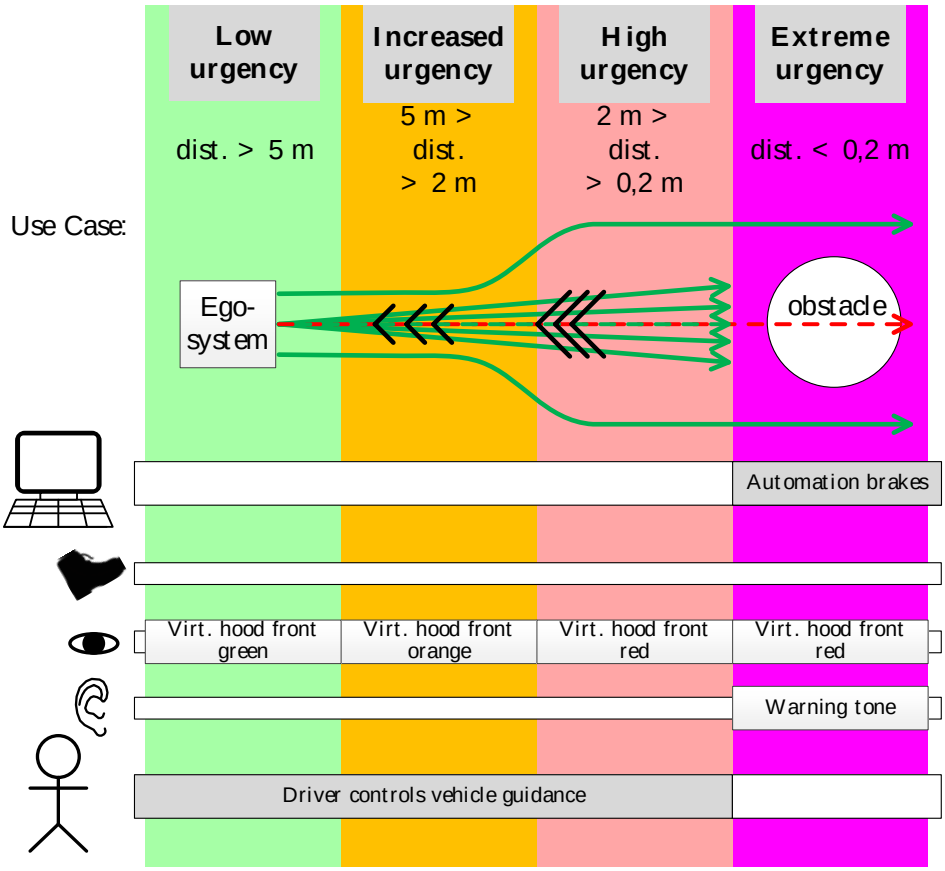


Figure 7: "Collision Avoidance Pattern" to prevent collisions in forward vicinity (cf. [19]).

When there is 5 m distance between ego-vehicle and obstacle the forward hood turns from green to orange, except if the driver starts to steer to the right. When forward distance is lower than 2 m and the ego vehicle and obstacle are in the same lane, forward hood turns from orange to red except if the driver starts to steer to the right. If the driver initiates reverse and the distance increases the interaction pattern deescalates, respectively.

A similar collision avoidance pattern can be implemented to be used as a side collision avoidance assistance either for parked vehicles at the side or to prevent road departure (see Figure 8).

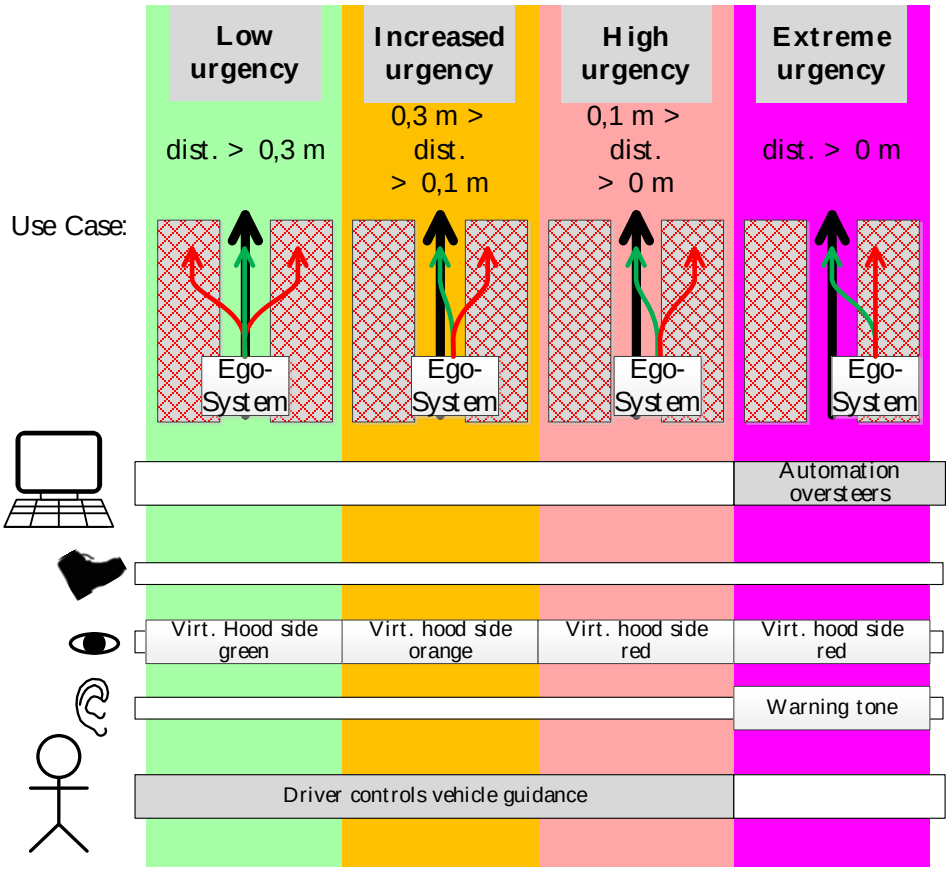


Figure 8: "Collision Avoidance Pattern" to prevent side collisions or behave like a virtual gravel trap when departing from a road (cf. [19]).

Again not only the current relation of the vehicle towards a side obstacle or road-side end defines the escalation, but also the driver's action. If the driver already steers to the left when departing to the right, the pattern will deescalate according to the steering model of the vehicle and the respective time of lane departure (TLC).

3.0 STUDY

To test the collision avoidance interaction pattern for different system qualities, different scenarios were used.

3.1 Study design

The design of the experiment was a within subjects design with three repeated measurements. 18 military drivers took part. The mean age of the participants was 32 (SD=6,3). Every run included a training of 3 minutes to get used to the setup. After every condition they filled out the NASA-TLX questionnaire. At the end of the experiment the systems were evaluated in a semi-structured interview.

The NASA Task Load Index (NASA-TLX) [20] is an assessment tool that rates perceived task load in order to assess a task. The task load is divided into six subscales. They are rated for each task within a 100-points

range with 5-point steps.

- Mental Demand: How mentally demanding was the task?
- Physical Demand: How physically demanding was the task?
- Temporal Demand: How hurried or rushed was the pace of the task?
- Performance: How successful were you in accomplishing what you were asked to do?
- Effort: How hard did you have to work to accomplish your level of performance?
- Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

Additionally quantitative data for displacement of the centre of the lane and the amount of collisions with infrastructure, cars and mines was logged during the test runs. Different scenarios were used to find out about the usefulness of the different test conditions in terms of performance and situation awareness (driving data), and task load (NASA-TLX). The respective scenarios were a city scenario with other traffic participants, like pedestrians and other vehicles, and an off-road part. In order to prevent sequence effects, two different maps with different scenario sequences were built and the test conditions permutated.

Respectively, comparable interaction patterns were used in the off-road part, where a very narrow path needed to be followed through a mine field.

3.2 Setup

The participants drove in a random order with a cabin, a monitor and a monitorsystem with augmentation condition (see Figure 9).

The study was conducted in a generic static driving simulator running the professional driving simulation software SILAB from the Würzburger Institut für Verkehrswissenschaften. As driving interfaces an active steering wheel and active gas and brake pedals from the company SENSODRIVE were used and a Sidestick from the company Stirling Dynamics as gear stick.

Regarding the visual interfaces, there was a training condition and three test conditions (see Figure 9).

In the training condition Figure 9 a) the simulation was visualized via a cave set-up representing three large projection screens that were arranged in a 90 degree angle to the sides and to the front. Additionally two 13” LCD 720p monitors were used as rear-view mirrors. A third 13” LCD 720p monitor was used to visualize speedometer and tachometer.

In the cabin condition Figure 9 b) a wooden vehicle frame was added to the training condition to introduce ambient occlusion at A and B-pillars.

a) Training **b) Cabin** **c) Monitorsystem**

d) Monitorsystem with interaction pattern

Figure 9: Training and test conditions during the experiments

In the monitor condition Figure 9 c) and d), the wooden frame was replaced by a monitor array of five 13” LCD 720p monitors that cover 160 degree of the driver’s horizontal field of view. Rear-view mirrors as well as ramp mirrors were integrated as picture-in picture (PiP) in the forward left and forward right screens (see Figure 9 d)).

Additionally in the monitorsystem with augmentation condition Figure 9 d) de-pending on the actions of the driver and the respective situation, interaction pat-terns escalated or deescalated. As mentioned before, the basic architecture of the generic assistance and automation system is based on the concept of interaction mediation and cooperative guidance and control of highly automated vehicles [1], [3]. In the respective study only visual assistance is given via the screens. Therefore the final escalation step of the collision avoidance patterns Figure 7 and Figure 8, when control was shifted from human to automation, was not considered.

The hypothesis of this study is that the human behaviour adaptation using interaction patterns via a camera monitorsystem with augmentation will improve the situation awareness, the driver performance and will reduce the overall task load.

3.3 Evaluation

As mentioned before, the concepts were tested in a simulator experiment, in which 18 military drivers took part. The mean age of the participants was 32 (SD=6,3). All of them have a car and a truck driving licence, eight a motorcycle licence, and three a tank licence. 11 of the participants use their vehicle for pri-vate purpose daily, 1 participant 3-5 times a week. 50 % of the participants had very little simulator experience, 50 % little or rather little. Eleven persons asses their driving style as safe/experienced, three as dynamic/sportive/brisk, 4 as cautious. The experience with driving assistance systems, e.g. lane departure warning system, was very low. Only with adaptive cruise control systems 50 % of the par-ticipants have extensive experience.

Figure 10 and Table 1 show the results of the NASA-TLX

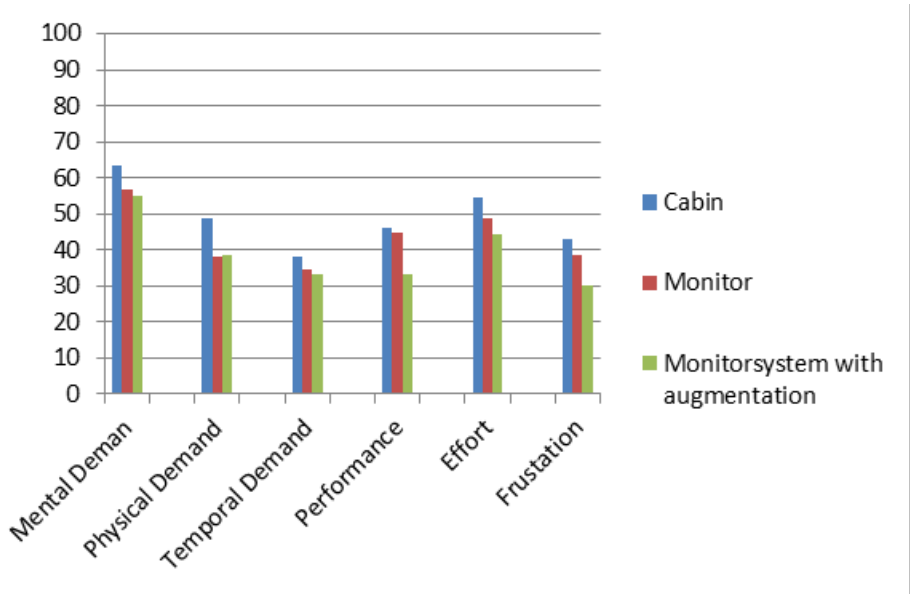


Figure 10: NASA-TLX diagram.

Table 1: NASA TLX Scores.

	Cabin M (SD)	Monitor M (SD)	Monitorsystem with augmentation M (SD)
Mental Demand	63,33 (25,95)	56,94 (24,20)	55 (23,31)
Physical Demand	48,61 (21,13)	38,06 (18,16)	38,61 (21,41)
Temporal Demand	38,33 (18,15)	34,72 (20,11)	33,06 (15,54)
Performance	**46,11* (19,52)**	**44,72* (19,89)**	**33,33* (14,65)**
Effort	54,72 (23,29)	48,61 (25,19)	44,44 (24,85)
Frustration	43,06 (20,23)	38,61 (23,06)	30,28 (18,90)

*p<0,05

The mental demand for driving with the cabin was higher rated than when driving with the monitor and monitorsystem with augmentation condition. The physical demand was also rated as the highest between the concepts. The monitorsystem with augmentation had the lowest temporal demand, effort and frustration. The participants were satisfied with their performance the most after driving the monitorsystem with augmentation. For the performance there was a statistically significant difference between all conditions. A repeated measures ANOVA showed a difference, $F(2, 34) = 7{,}055$ $p = .003$, partial $\eta^2 = .293$. A Bonferroni-corrected post-hoc test showed a significant difference between the cabin and the monitor condition (.009, 95%-CI[-22.63, -2.93]). Also there was a difference between the monitor and the monitorsystem with augmentation condition (.005, 95%-CI[-19.45, -3.3]).

The results from lateral displacement from the road centre are displayed in Figure 11 and Table 2.

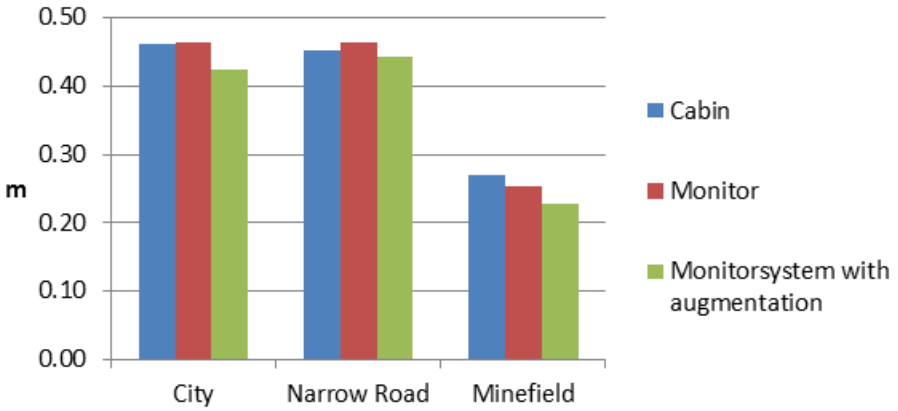


Figure 11: Displacement from the centre of the lane.

Table 2. Displacement from the centre of the lane.

	Cabin *M (SD)*	Monitor *M (SD)*	Monitorsystem with augmentation *M (SD)*
City	0,46 (0,14)	0,46 (0,12)	0,42 (0,16)
Narrow road	0,45 (0,07)	0,46 (0,09)	0,44 (0,09)
Minefield	0,27 (0,11)	0,25 (0,09)	0,23 (0,09)

The lowest lateral displacement was in all parts of the scenarios with the monitorsystem with augmentation. A statistical significance could not be found between the conditions.

In Figure 12, Figure 13 and Figure 14 show the different amount of collisions with the respective elements. Collisions with infrastructure, cars and mines were mainly caused in the cabin. The probability of a collision is higher, but also the amount of absolute collisions. The infrastructure used in this study was made of houses, walls, traffic lights and traffic signs.

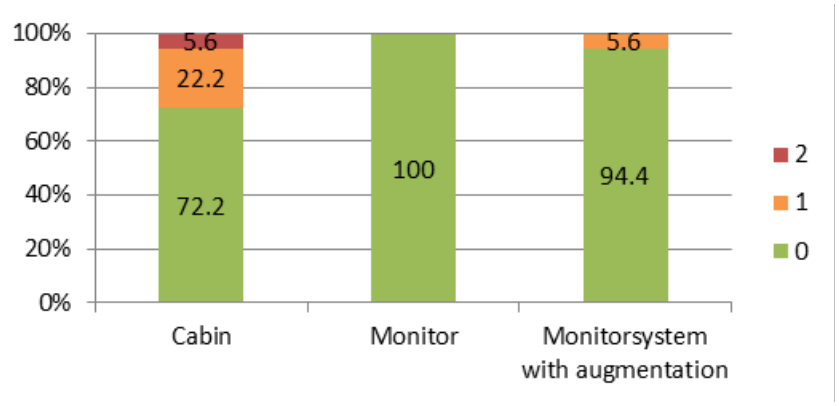


Figure 12: Collisions with infrastructure.

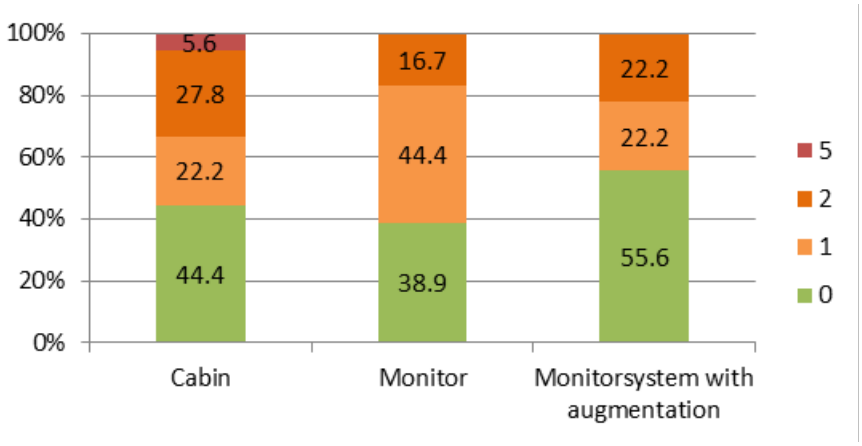


Figure 13: Collisions with cars.

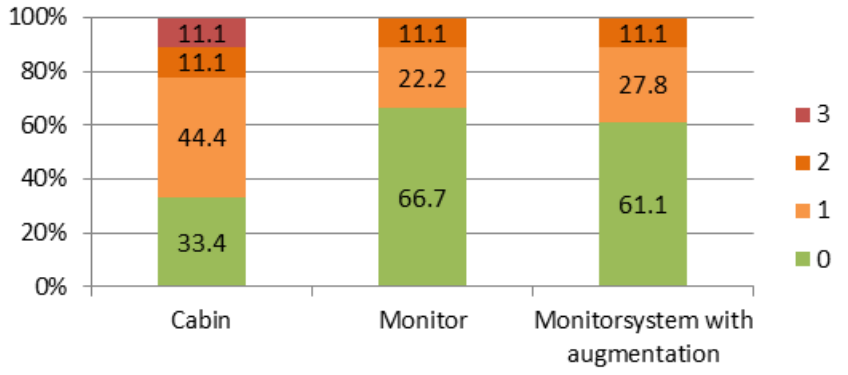


Figure 14: Collisions with mines.

4.0 CONCLUSION AND OUTLOOK

This publication described how interaction patterns and image schemas could be used in Human Autonomy Teams with a cooperative automation. This approach was developed in close interplay with the concepts of cooperative guidance and control, where a cooperative automation and a human control a machine together, and of augmented reality, where a natural representation of the world, e.g. in form of a video stream, is enriched with dynamic symbology. The concept was applied to armoured vehicles, instantiated as interaction pattern “collision avoidance”, implemented in a fix based simulator, and tested with professional operators. The results were a good mix of encouragement and lessons learned, both for the methodical approach of pattern based human machine interaction, and for the application of AR-based cooperative guidance and control.

The use of interaction patterns for collision avoidance, which was tested in the study, showed different aspects of improvement. One of the main objectives to reduce the number of collisions with vehicles in the near proximity could be reached. Also the displacement was always lower. As a result the driving performance could be enhanced. Furthermore the task load could be reduced, as the driving with the augmentation caused the lowest temporal demand, effort and frustration. Also the participants were satisfied with their performance the most.

A possible opportunity for improvement is the integration of live eye tracking to move the drivers focus e.g. from one screen to another or to point out the critical collision area. Physiological metrics might be useful for patterns in other situations, e.g. combining higher assistance and automation degrees with a mode transition pattern when the task load of drivers is too high. Also the used visual figure of the hood could be improved in terms of size and form. Additionally multimodal extension of the patterns with haptic or acoustic feedback could be evaluated.

Regarding the AR-based cooperative guidance and control: We gained an increasing understanding of how this cooperative interplay between an automation and humans can be organized, and patterns are an excellent way to describe this organization. We have encouraging results on a couple of patterns, and especially with the link to image schemas we increasingly understand why some patterns work differently and better than others. With all optimism, we are far from having optimal patterns, and far from having more than a first glimpse of this vast design space of human machine cooperation, and technology based reality augmentation.

Regarding the overall approach of linking patterns with image schemas: For us this is the most promising way to link everything that the community has learned already about specific patterns and specific image schemas, and to make this available in the design process. We have a first understanding how this link can be done, however we are far from having an optimal way to do this linking of patterns and image schemas efficiently. More research, and especially more joint effort is needed to organize and combine the increasing knowledge that is being built up in different spots in the community, and to make this available in the specific design and engineering situation of real products, so that it can improve the increasingly complex human machine systems, not only in the far future, but right here, right now.

The use of interaction patterns showed different aspects of improvement. One of the main objectives to reduce the number of collisions with vehicles in the near proximity could be reached. Also the displacement was always lower. As a result the driving performance could be enhanced. Furthermore the task load could be reduced, as the driving with the augmentation caused the lowest temporal demand, effort and frustration. Also the participants were satisfied with their performance the most.

Using Interaction patterns is the most promising way to link everything that the community has learned already about human autonomy teaming. Specific patterns and specific image schemas have to be tailored together, linked to specific tasks, and to be made available in the design process as early as possible. We

have a first understanding how this link can be done, however we are far from having an optimal way to do this linking of patterns and image schemas efficiently. More research, and especially more joint effort is needed to organize and combine the increasing knowledge that is being built up in different spots in the community, and to make this available in the specific design and engineering situation of real products, so that it can improve the increasingly complex human machine systems.

5.0 ACKNOWLEDGEMENTS

Contents of this work were gained from studies funded by the German Federal Ministry of Defence and the Deutsche Forschungsgesellschaft DFG.

6.0 REFERENCES

[1] M. Baltzer, E. Altendorf, S. Meier and F. Flemisch (2014). “Mediating the interaction between human and automation during the arbitration processes in cooperative guidance and control of highly automated vehicles: Basic concept and first study”. In N. A. Stanton, S. J. Landry, G. Di Bucchianico and A. Vallicelli (Eds.), Advances in Human Aspects of Transportation Part I, USA, AHFE, pp. 439-450.

[2] F. O. Flemisch, J. Schomerus, J. Kelsch und U. Schmuntzsch (2005) “Vom Fahrer zum Reiter? Zwischenbericht 2005 auf dem Weg von der H-Metapher zum H-Mode für Bodenfahrzeuge“. In VDI-Kongress "Fahrer im 21. Jahrhundert", Braunschweig.

[3] F. Flemisch, H. Winner, R. Bruder and K. Bengler (2016): “Cooperative Guidance, Control, and Automation”. In Winner, H.; Hakuli, S.; Lotz, F. & Singer, C. (Eds.) Handbook of Driver Assistance Systems, pp. 1471-1481. Springer, Switzerland

[4] A. Loiselet and J. M. Hoc (2001), “La gestion des interférences et du référentiel commun dans la coopération: implications pour la conception”. Psychologie Française, 46, pp. 167-179.

[5] F. Flemisch, D. Abbink, M. Itoh, M.-P.-. Pacaux-Lemoine and G. Weßel (2016), “Shared control is the sharp end of cooperation: Towards a common framework of joint action, shared control and human machine cooperation”, IFAC-PapersOnLine, 49, pp. 72-77.

[6] F. Flemisch, J. Kelsch, C. Löper, A. Schieben und J. Schindler (2008). “Automation spectrum, inner / outer compatibility and other potentially useful human factors concepts for assistance and automation“. In D. de Waard, F. Flemisch, B. Lorenz, H. Oberheid und K. Brookhuis (Eds.) Human Factors for assistance and automation, , Maastricht, the Netherlands, Shaker, pp. 1-16.

[7] R. Bernotat (1970) “Plenary Session: Operation Functions in Vehicle Control Anthropoteehnik in der Fahrzeugfuhrung”. Ergonomics, 13, pp. 353-377.

[8] T. B. Sheridan (1976). “Toward a General Model of Supervisory Control“. In Monitoring Behavior and Supervisory Control, Springer, pp. 271-281.

[9] E. Donges (1982). “Aspekte der Aktiven Sicherheit bei der Führung von Personenkraftwagen“, Automobil-Industrie, 27, pp. 183-190.

[10] R. Parasuraman, T. B. Sheridan and C. D. Wickens (2000). “A Model for Types and Levels of Human

Interaction with Automation", Transactions on Systems, Man, and Cybernetics, pp. 286-297.

[11] C. Löper, J. Kelsch and F. O. Flemisch (2008). "Kooperative, manöverbasierte Automation und Arbitrierung als Bausteine für hochautomatisiertes Fahren". Braunschweiger Symposium zum Thema Automatisierungs-, Assistenzsysteme und eingebettete Systeme für Transportmittel (AAET).

[12] M.C.A. Baltzer; G. Weßel; D. López and F. Flemisch (2017). "Interaction Patterns for Cooperative Guidance and Control Automation mode transition in highly automated truck convoys", IEEE International Conference on Systems, Man, and Cybernetics.

[13] G. Lakoff (1987). Women, fire and dangerous things, Chicago: University of Chicago Press.

[14] M. Johnson (1987). The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason, Chicago: University of Chicago Press.

[15] J. Hurtienne (2009). Image Schemas and Design for Intuitive Use - Exploring new Guidance for User Intrerface Design. Dissertation, Berlin: Technische Universität Berlin.

[16] G. Lakoff (1989). "Some empirical results about the nature of concepts", Mind and Language, 4, no. 1-2, pp. 103-129.

[17] R.E. Bolz (1973). "CRC handbook of tables for applied engineering science". CRC press, Boca Raton

[18] C. Baldauf (1997). "Metapher und Kognition: Grundlagen einer neuen Theorie der Alltagsmetapher", P. Lang, Frankfurt am Main

[19] M.C.A. Baltzer, C. Lassen, D. López and F. Flemisch (2018) "Behaviour Adaptation using Interaction Patterns with Augmented Reality Elements", Augmented Cognition: Intelligent Technologies, Springer International Publishing.

[20] Hart, S.G. & Staveland, L.E. (1988). "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research", Advances in Psychology 52, 139–183.

Design Patterns and Ontologies for Situated Human-Agent Collaboration at Organization, Team and Interaction Level

Mark A. Neerincx[1,2] and Jurriaan van Diggelen[1], Martin L. Stolk[1]
[1]TNO Perceptual and Cognitive Systems, Kampweg 5, 3769 DE Soesterberg
[2]Delft University of Technology, Faculty of EEMCS, Van Mourik Broekmanweg 6, 2628 XE Delft
THE NETHERLANDS

<mark.neerincx, jurriaan.vandiggelen, martin.stolk>@tno.nl

ABSTRACT

For the integration of robots into the operations of rescue, reconnaissance, security and defence teams, there is a need for theoretical and empirical grounded human-agent collaboration models. This paper proposes the development of organization, team and interaction design patterns with corresponding ontologies, as re-usable building blocks of human-agent collaboration. A situated Cognitive Engineering (sCE) methodology supports the incremental development and implementation of these building blocks (i.e., ePartner abilities and behaviours) into a perceptual-cognitive agent framework. Core cognition concepts are formalized as ontology design patterns (e.g., "workload" and "situation awareness") and related to the concerning domain ontology (e.g., disaster response with concepts like "point of interest" and "urgency"). In parallel, relevant design patterns are identified or constructed with the corresponding ontologies, at the organization (e.g., norms), team (e.g., work agreements) and interaction level (e.g. explanation). The ontologies are implemented into the prototype and tested on their expected outcomes (i.e. the claims). The European TRADR project (www.tradr-project.eu) provides an example in which an agent-based system for robot-assisted disaster response is developed, where response team dynamically adapt the human-robot teamwork (i.e., task allocation and coordination) to the momentary context.

1.0 INTRODUCTION

Ongoing progress in artificial intelligence, robotics, sensors and network technology feeds the development of software agents that can act as teammates or so-called *ePartners.* These ePartners are applied in defense, safety and security missions [1],[2],[3] and operate robustly in hazardous and chaotic environments. They are supposed to show responsible, agile and flexible behaviors in response to foreseen *and* unforeseen events. Furthermore, their behaviors should be fluently integrated into the overall human-agent teamwork, and harmonized with the activities and expectations of the human teammates. To establish the envisioned human-ePartner collaboration, ePartners require an understanding of human's social, cognitive, affective and physical behaviors, and the ability to engage in partnership interactions (such as explanations of task performances, and the establishment of joint goals and work agreements).

To cope with the design complexity of these systems, theoretically and empirically grounded models are needed that capture these social, cognitive, affective and physical processes. This paper proposes organization, team and interaction design patterns with corresponding ontologies, that can be used as re-usable building blocks for human-agent collaboration in the design process. These patterns are abstractions and generalizations from successful individual design solutions. In our approach, humans and ePartners are part of a joint cognitive system that operates in a specific set of situations. In the system development phase, relevant *situation cognition* models have to be identified or constructed as core concepts of the human-ePartner collaboration. A *domain* model is constructed to explicate how cognition operates in the concerning situated collaborations. Subsequently, the *design patterns* are selected or constructed, which explicate how the actors of the joint cognitive system collaborate successfully in this domain.

As a running example, we will regard an agent-based system for robot-assisted disaster response, as used in the European TRADR project (www.tradr-project.eu). Within this application a response team needs to dynamically adapt the task allocation and coordination to the momentary context [4]. Figure 1 shows the Unmanned Ground Vehicles (UGV) and Unmanned Aerial Vehicles (UAV) that were deployed in the TRADR-project (the photos were taken during a realistic field evaluation of the TRADR-system).

Figure 1: Unmanned Ground Vehicle (UGV; left) and Unmanned Aerial Vehicle (UAV; right) during a field evaluation in the TRADR-project.

Figure 2 presents a part of the disaster response team and its structure (in this example with three human team-members, their ePartners and two robots). The team leader and her ePartner have the lead of two "units", each consisting of an operator, ePartner and robot. The domain model and cognition models of, for example workload, underpin the specification of design patterns at three levels. At the highest level, *organization* design patterns are specified about normative behaviors, such as the work and rest schedules of an organization to comply with a workload norm. At the intermediary level, successful collaborative behaviors of specific *teams* are specified, such as successful work agreements on the task (workload) allocations of a team leader, the two operators, the three ePartners and the two robots in a specific situation. At the lowest level, the *interaction* design patterns specify how the information is communicated between the actors, such as the user interfaces for sharing workload information.

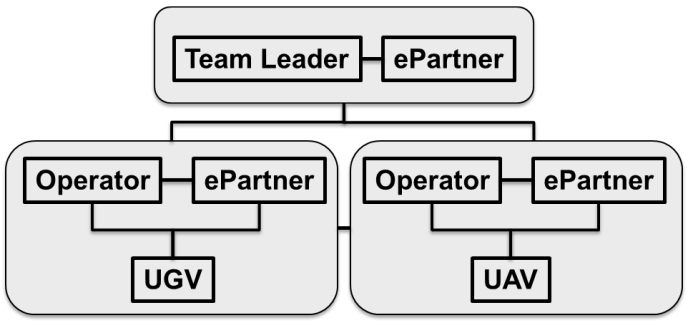


Figure 2: Part of the TRADR disaster response team with a UAV and a UGV.

Section 2 gives a brief overview of a situated Cognitive Engineering (sCE) method for the incremental development of the design patterns with the corresponding domain and cognition models (e.g., "workload" and "situation awareness"), which underpin system's functionality to collaborate. Section 3 discusses a perceptual-cognitive framework to implement this functionality. Subsequently, section 4 elaborates on the

development and application of the design patterns, entailing an extendable, theory and empirical grounded, knowledge base of the ePartner functions and behaviors. Section 5 contains the conclusions of this paper.

2.0 SITUATED COGNITIVE ENGINEERING OF E-PARTNERS

Knowledge about the *operational demands*, *human factors* and *technology* (see Fig 3), underpin the requirements specification of the human-agent system. The sCE methodology supports the acquisition, construction and usage of knowledge from these three foundation components.

First, *ontologies* are used as formal, explicit specifications of the knowledge that the agents need to have about the domain ("operational demands") and the human factors (see Fig 3). An ontology is an explicit representation of declarative knowledge. It is structured around concepts, properties, and relations, which allows for automated reasoning (about this structure). For our purposes, the ontology encompasses consensual knowledge that is important for humans-agent collaboration. Also here, we aim at an extendable set of building blocks, i.e., Ontology design Patterns (OP; [5]). They distinguish knowledge at the upper ('classes'), and lower level ('individuals'). To systematically incorporate human factors knowledge, relevant concepts of situated cognition are formalized as an ontological model, such as situation awareness [6] and workload [7]. For example, Harbers and Neerincx [7] provide a workload model, which distinguishes cognitive load, affective load and mental effort, and can reason about task allocations. The upper level workload ontology (or parts of the ontology) has been studied in and applied to several domains: disaster response [8], space missions [9], ship control [10], and train control [7]. During operation, the domain ontology is added to instantiate the concerning concepts, e.g. for the computational reasoning (such as task allocation based on "urgency" and "workload"). Part of the domain ontology for disaster response is shown in figure 4 (for more information on the domain ontology, see [11]). Note that the ontology refers to human factors concepts (like workload) that are specified in the "Situated Cognition" ontologies.

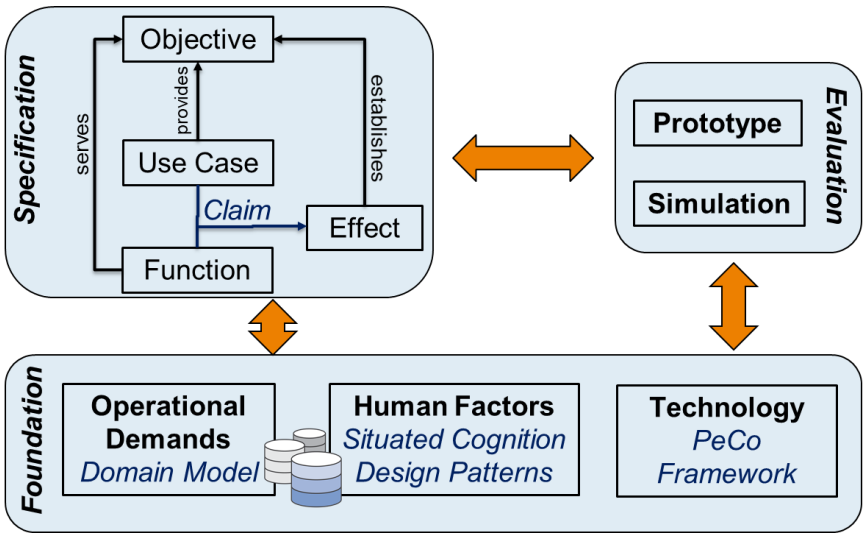


Figure 3: Situated Cognitive Engineering entails an iterative grounding ("foundation"), requirements derivation with its rationale ("specification"), and prototype and simulation testing ("evaluation").

Second, to build on past and current agent systems, and to relate the concerning agent behaviors to each other, it is important to understand and share agents' *design rationales*. In the specification or the functional requirements (see Fig 3), *claims* explicate the design rationale of a specific function: The expected effect of this functionality in a (set of) use case(s). In other words, claims describe the effects of agents behavior on

the condition and performance of (the ensemble of) team actors; the effects can entail an expected positive trade-off of the advantages (upsides) over the disadvantages (downsides). Explicating possible downsides is important, because they have to be checked in the evaluation. For example, a contrastive explanation function for a task reallocation advice might be expected to enhance the trust of a team-leader in the advising agent, with only a minor increase of workload. Claims have to be testable and, if appropriate, refer to stakeholders' values.

Third, to generalize over functions and claims, so-called "*design patterns*" are used to capture the "generic" design rationale by explicating common solutions for classes of problems (e.g., as formulated in a use case [12]). As far as possible, proven models and solutions have to be looked for and maintained as an accessible library in the foundation component. Section 4 will elaborate on the identification and construction of design patterns.

Fourth, the design rationales with their claims are evaluated via simulation and prototyping, resulting in approval, rejection or refinement of the concerning models. In TRADR, the ontology, team and interaction design patterns have been developed and tested in several cycles, applying simulations and realistic field tests with a complete team, including unmanned ground vehicles and unmanned aerial vehicles.

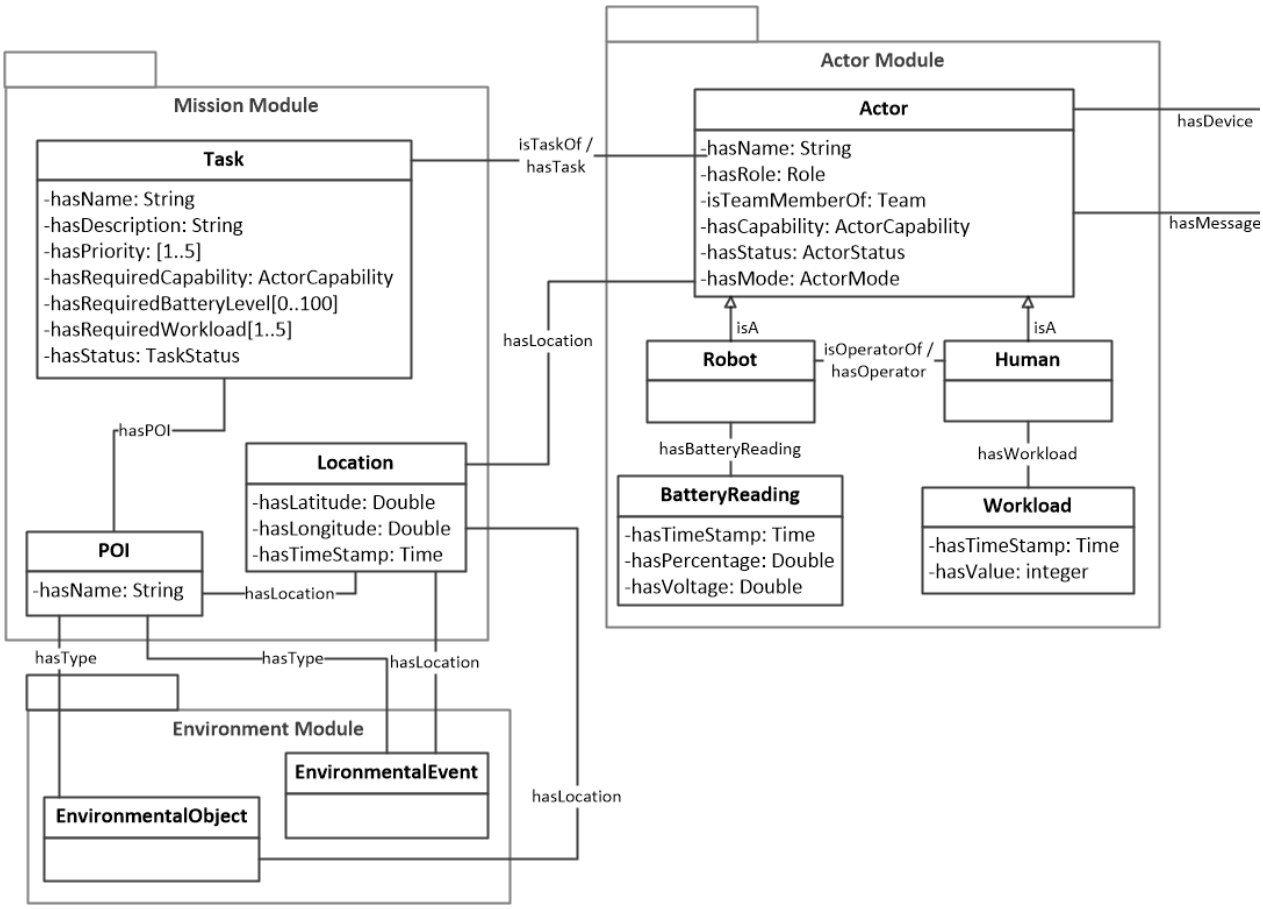


Figure 4: Part of the TRADR ontology (domain model of sCE in Fig 3; more information on the TRADR ontology, see [11]).

3.0 PERCEPTUAL AND COGNITIVE E-PARTNER FRAMEWORK

As a team-member, ePartners need to acquire and show understanding of the meaning, weighting and implementation method of the joint goals they are committed to. This type of understanding and commitment is being built and communicated via three partnership interactions in Human-Agent Teams. Currently, a generic perceptual-cognitive (PeCo) framework-suite is being developed for this purpose ([15], [16], see Fig. 5). This PeCo-framework distinguishes three components: An Objective, Work Agreement and Explanation component.

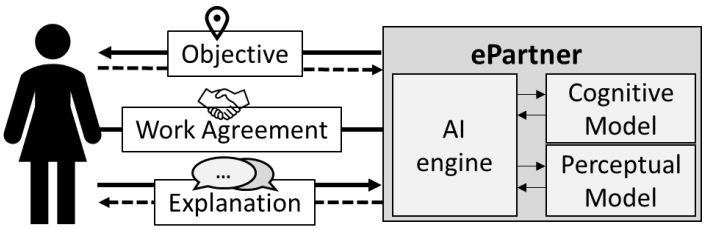


Figure 5: Three human-agent partnership interactions: objective sharing, harmonizing activities via work agreements and explaining task outcomes.

First, active partnerships require the setting, progress monitoring, refinement and adjustment of joint *objectives*, based on a shared understanding, awareness and communication of the objectives and their situated urgencies. Stakeholders' values, strategic intentions and momentary context conditions determine the urgency and feasibility of the objectives to pursue. The choices of the objectives can require re-assessment and adjustment to deal with changing circumstances, possibly just in advance or even during the work processes. Objective ontology design patterns [5] [17] [18] and utility elicitation methods [19] are being developed to support the setting of joint objectives. Adaptivity is an important capability of human-machine teams, which may involve creativity to find a new work process (possibly with an adjusted objective) when the current process is expected to fail. Humans perceive, associate and reason differently than machines; consequently, the perspective-taking and solution space for new (possibly wicked) problems is larger when machines and humans share their perceptions, associations and reasoning.

Second, the proposed human-agent partnerships involve explicit, univocal and possibly adjustable, Work Agreements (WA) on how the tasks should be allocated and performed [16]. Similar to the modeling of the objectives, an ontology design pattern has been developed for work agreements, defining the knowledge required for a team member to set, reason about and adjust work agreements. The WA-ontology defines core concepts with their relations encompassing the knowledge to specify, activate, monitor, and reason about work agreements: ***<creditor, debtor, antecedent, consequent, lifespan, acceptance>***. This ontology was implemented and tested for human-robot partnerships in disaster response teams (i.e., in the TRADR-project). Such a team needs to dynamically adapt the task allocation and coordination to the momentary context. A disaster response field test showed that the fire-brigade officers could specify work agreements as required, which, subsequently, brought forward the desired adaptive team behavior of the concerning robot.

Third, we developed explanation models for ePartners that should invoke adequate trust calibration (i.e., to mitigate over- or under-reliance) and improve the learning of the team-members. Currently, a generic perceptual-cognitive explanation (PeCo) framework is being developed for this purpose [15]. The perceptual level provides an Intuitive Confidence Measure [18] and identifies the "foil" that can be used in a contrastive explanation [19]. Such an explanation fits with human explanation dialogues, centering on the question "Why this output (the fact) instead of that output (the foil)?" It reduces the factors in the explanation to the ones that are of interest to the human, so that it can be better interpreted by the human. The cognitive level provides the beliefs, goals and emotions for explanations [15]. Initial evaluations showed that humans

can understand the explanations that PeCo will provide, but further personalization and contextualization is needed.

4.0 DESIGN PATTERNS FOR HUMAN-E-PARTNER COLLABORATION

As mentioned in section 2, design patterns explicate the rationale of the chosen design solution as a common solution for a class of design problems. The following procedure aims at a common pattern library for human-agent collaboration:

1. *Identify key design problems for specific functionality*
2. *Search for available design patterns*
3. *If no pattern can be found,*
   and if it is a general, recurrent design problem:
   - *Start with a Proto Pattern, a pattern "in construction", i.e., a design problem and solution documented in a pattern form (yet lacking empirical grounding)*
4. *Provide different instantiations (examples)*
5. *Test, refine and validate these examples*
6. *If successful:*
   - *Make the Design Pattern accessible in library (of best practices)*

In our approach, the pattern language comprises a structure, template and ontology. Patterns can be constructed at different abstraction levels with different types of relationships, forming the building blocks of the research and development of team agents. We distinguish three pattern levels: Organization (e.g., work and rest schedules), team (e.g., constructive vs. destructive) and interaction (e.g., direct feedback).

Organizations often maintain norms for the joint activities of the workers. These norms can be explicated and formalized as policies: Explicit enforceable constraints on human and agents performance in a given situation, as authorization ("permissions") and obligation [23]. This way, policies govern the behavior of specific agent roles within the organization to enforce normative behaviors (i.e., they apply to a set of actors). The normative model allows to reason about (and to verify) obligations, permissions, and prohibitions, based on deontic logic [25]. Example policies for human-robot collaborations come from the space [23] and automotive domain [24]. In the last domain, the policies specified when an autonomous car should provide the driver with personalized supportive information, and when to hand over control to the driver. Both policies were linked to the driver's workload ([24], cf. the workload ontology of section 2). In these examples and in the TRADR project, organization design patterns have not yet been formulated. However, we have identified some "Proto Patterns" (see above) that need to refined and validated, such as the rest and work schedule pattern mentioned in the introduction of this paper, and design patterns for dealing with hazardous substances in populated areas. Following such patterns, the fire fighters and robots would operate within the safety and health norms of the organization. Van Diggelen et al. [13] provide a design pattern language for teamwork, which seems to be a suitable starting point to describe the normative joint activities at the level of the organization.

Team design patterns specify generic reusable behaviors of actors for supporting effective and resilient teamwork [13]. These patterns can be used to describe the dynamics of teamwork over longer periods of time. They can be constructive (i.e. leading to a more coherent team) or destructive (i.e. leading to a less coherent team), which can be observed in the behavior of an individual or group of actors. During a field test, for example, the TRADR team showed an individual, destructive micro-management pattern of the team-leader. By formalizing and implementing this pattern in the PeCO-framework, the agent can mitigate this behavior in future situations.

Interaction design patterns describe the shaping of the (multimodal) human-agent interactions, such as the dialogue acts for work agreements in TRADR [16]. Figure 6 presents part of the Interaction Design patterns for Human-Agent Collaboration (HAC) that have been specified so far. It distinguishes patterns for a Tailored Situation View (SV), Adjustable Work Agreements (WA) and Harmonized Interactive Notifications (IN) (see [13] for more details on these patterns).

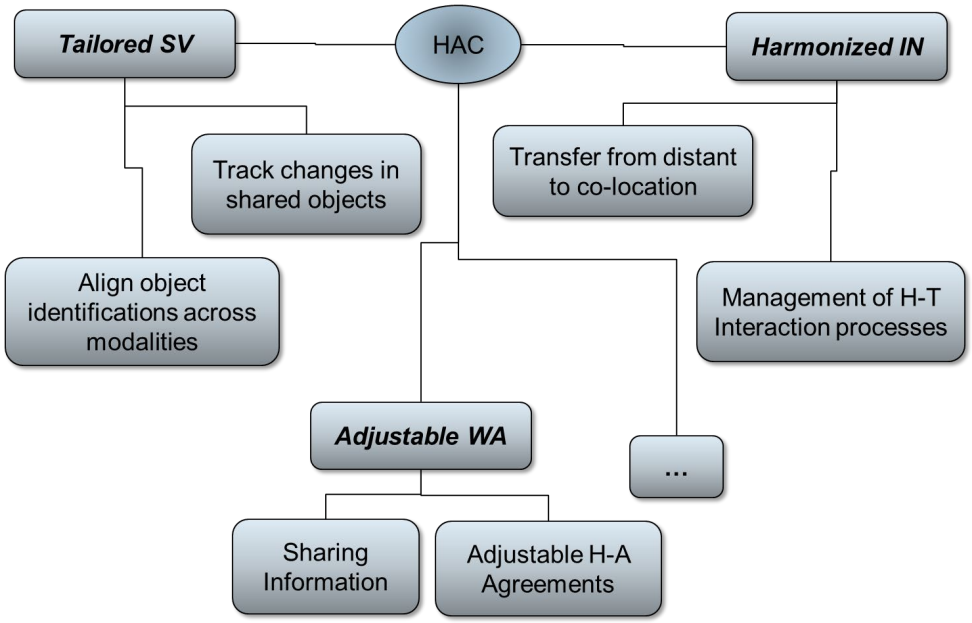


Figure 6: Part of the Interaction Design patterns for Human-Agent Collaboration (HAC), distinguishing patterns for a Tailored Situation View (SV), Adjustable Work Agreements (WA) and Harmonized Interactive Notifications (IN) (see [13] for more details on the patterns).

5.0 CONCLUSIONS

For the integration of robots into the operations of rescue, reconnaissance, security and defense teams, there is a need for theoretical and empirical grounded human-agent collaboration models. This paper proposes the development of organization, team and interaction design patterns with corresponding ontologies, as re-usable building blocks of human-agent collaboration. A situated Cognitive Engineering (sCE) methodology supports the incremental development and implementation of these building blocks (i.e., ePartner abilities and behaviors) into a perceptual-cognitive agent framework. Core cognition concepts are formalized as ontology design patterns (e.g., "workload" and "situation awareness") and related to the concerning domain ontology (e.g., disaster response with concepts like "point of interest" and "urgency"). In parallel, relevant design patterns are identified or constructed with the corresponding ontologies, at the organization (e.g., norms), team (e.g., work agreements, and interaction level (e.g. explanation). The ontologies are implemented into the prototype and tested on their expected outcomes (i.e. the claims).

A key proposal of this paper is to build on past and current agent systems via the specification and sharing of *design rationales*. This rationale includes the explication of testable claims ***and*** the derivation of design patterns. The design patterns are formulated at three abstraction levels as building blocks of the research and development of team agents or ePartners: Organization, team and interaction. A second key proposal is to use *ontologies* that formalize human factors concepts and relate them to the domain concepts. These ontologies underpin the knowledge base ("beliefs") and reasoning of the TRADR agents, for example for proposing, adjusting, accepting, fulfilling or rejecting a work agreement. By formalizing and implementing

constructive and destructive team patterns in the knowledge base, the agent can detect and support constructive behaviors and mitigate destructive behaviors.

Following the proposed methodology, robots can evolve as partners in disaster response by instantiating design patterns for the (a) sharing objectives, beliefs and experiences, (b) committing to norms and work agreements, and (c) uptake and learning of explanations and feedback (Fig 5).

ACKNOWLEDGEMENTS

This work was partly funded by the EU FP7 TRADR project (grant no. 60963) and the Dutch Ministry of Defence's research program (project V1719). We thank the fire fighters from the "Gezamenlijke Brandweer Rotterdam" in the Netherlands for participating in this study.

REFERENCES

[1] Neerincx, M. A. (2011). Situated cognitive engineering for crew support in space. *Personal and Ubiquitous Computing*, *15*(5), 445-456

[2] Bosse, T., Breebaart, L., Diggelen, J. V., Neerincx, M. A., Rosa, J., & Smets, N. J. (2017). Developing ePartners for human-robot teams in space based on ontologies and formal abstraction hierarchies. *International Journal of Agent-Oriented Software Engineering*, *5*(4), 366-398.

[3] Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V., & Johnson, W. W. (2017, July). Why Human-Autonomy Teaming?. In *International Conference on Applied Human Factors and Ergonomics* (pp. 3-11). Springer, Cham, Germany.

[4] Kruijff-Korbayová, I., Colas, F., Gianni, M., Pirri, F., de Greeff, J., Hindriks, K., Neerincx, M.A., Ögren, P., Svoboda, T. & Worst, R. (2015). TRADR project: Long-term human-robot teaming for robot assisted disaster response. *KI-Künstliche Intelligenz*, *29*(2), 193-201.

[5] Gangemi, A. and Presutti, V. (2009). Ontology design patterns. In *Handbook on ontologies*, 221-243. Springer, Berlin, Heidelberg, Germany.

[6] Smets, N. J., Neerincx, M. A., Jonker, C. M., & Båberg, F. (2017, March). Ontology-based situation awareness support for shared control. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 289-290). ACM.

[7] Harbers, M., & Neerincx, M. A. (2017). Value sensitive design of a virtual assistant for workload harmonization in teams. *Cognition, Technology & Work*, *19*(2-3), 329-343.

[8] Colin, T. R., Mioch, T., Smets, N. J., & Neerincx, M. A. (2012). Estimating an operator's cognitive state in real time: A user modeling approach. In *RO-MAN* (pp. 627-633).

[9] Cohen, I., den Braber, N., Smets, N. J., van Diggelen, J., Brinkman, W. P., & Neerincx, M. A. (2016). Work content influences on cognitive task load, emotional state and performance during a simulated 520-days' Mars mission. *Computers in Human Behavior*, *55*, 642-652.

[10] Neerincx, M. A., Kennedie, S., Grootjen, M., & Grootjen, F. (2009). Modeling the cognitive task load and performance of naval operators. In *International Conference on Foundations of Augmented Cognition* (pp. 260-269). Springer, Berlin, Heidelberg.

[11] Looije, R., Neerincx, M. A., & Hindriks, K. V. (2017). Specifying and testing the design rationale of social robots for behavior change in children. *Cognitive Systems Research*, *43*, 250-265.

[12] Schulte, A., Donath, D., & Lange, D. S. (2016). Design patterns for human-cognitive agent teaming. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 231-243). Springer, Cham.

[13] Neerincx, M. A., van Diggelen, J., & van Breda, L. (2016). Interaction design patterns for adaptive human-agent-robot teamwork in high-risk domains. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 211-220). Springer, Cham, Germany.

[14] Mioch, T., Peeters, M.M.M. and Neerincx, M.A. (2018). Improving Adaptive Human-Robot Cooperation through Work Agreements. *RO-MAN2018 27th IEEE International Symposium on Robot and Human Interactive Communication*.

[15] Saad, E., Hindriks, K. V., & Neerincx, M. A. Ontology Design for Task Allocation and Management in Urban Search and Rescue Missions. *ICAART 2018*.

[16] Van Diggelen, Neerincx, M.A., Peeters, M.M.M., and Schraagen, J.M.S. (in press). Developing effective and resilient human-agent teamwork using team design patterns. *IEEE Intelligent Systems*.

[17] Mioch, T., Ledegang, W., Paulissen, R., Neerincx, M. A., & van Diggelen, J. (2014). Interaction design patterns for coherent and re-usable shape specifications of human-robot collaboration. In *Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems* (pp. 75-83). Association for Computing machinery (ACM).

[18] Neerincx, M.A. , Van der Waa, J. , Kaptein, F. and Van Diggelen, J. (2018). Using perceptual and cognitive explanations for enhanced human-agent team performance. *EPCE 2018, LNAI 10906*, pp. 1–11, Springer.

[19] Peeters, M.M.M. and de Visser, E.J. (2018). "Longitudinal trust development in human-agent-robot teams," in *Workshop on Longitudinal Human-Robot Teaming, held at the Human-Robot Interaction Conference*, February 2018, Chicago, IL, USA.

[20] González-Ortega, J., Radovic, V., and Insua, D.R. (2018). Utility elicitation. In *Elicitation*, pp. 241-264. Springer, Cham

[21] Van der Waa, J., Diggelen, J. and Neerincx, M.A. (2018). The design and validation of an intuitive certainty measure. *Proceedings of IUI 2018 Workshop on Explainable Smart Systems*.

[22] Van der Waa, J., Robeer, M..J., Van Diggelen, J., Brinkhuis, M.J.S. and Neerincx, M.A. (2018). Contrastive Explanations with Local Foil Trees. *IJCAI 2018 Workshop on Human Interpretability in Machine Learning (WHI)*.

[23] Van Diggelen, J. M. Bradshaw, J.M., Johnson, M. Uszok, A. and Feltovich, P.J. (2010). Implementing collective obligations in human-agent teams using KAOS policies. In: *Coordination, Organizations, Institutions and Norms in Agent Systems V*. Springer, pp. 36–52.

[24] Mioch, T., Kroon, L., & Neerincx, M. A. (2017). Driver readiness model for regulating the transfer from automation to human control. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (pp. 205-213). ACM.

[25] Wieringa, R.J. and Meyer, J.-J. C. (1993). Applications of deontic logic in computer science: A concise overview. *Deontic logic in computer science*, pp. 17–40, 1993.

Design Patterns and Ontologies for Situated Human-Agent Collaboration at Organization, Team and Interaction Level

Mark A. Neerincx[1,2] and Jurriaan van Diggelen[1], Martin L. Stolk[1]
[1]TNO Perceptual and Cognitive Systems, Kampweg 5, 3769 DE Soesterberg
[2]Delft University of Technology, Faculty of EEMCS, Van Mourik Broekmanweg 6, 2628 XE Delft
THE NETHERLANDS

<mark.neerincx, jurriaan.vandiggelen, martin.stolk>@tno.nl

ABSTRACT

For the integration of robots into the operations of rescue, reconnaissance, security and defence teams, there is a need for theoretical and empirical grounded human-agent collaboration models. This paper proposes the development of organization, team and interaction design patterns with corresponding ontologies, as re-usable building blocks of human-agent collaboration. A situated Cognitive Engineering (sCE) methodology supports the incremental development and implementation of these building blocks (i.e., ePartner abilities and behaviours) into a perceptual-cognitive agent framework. Core cognition concepts are formalized as ontology design patterns (e.g., "workload" and "situation awareness") and related to the concerning domain ontology (e.g., disaster response with concepts like "point of interest" and "urgency"). In parallel, relevant design patterns are identified or constructed with the corresponding ontologies, at the organization (e.g., norms), team (e.g., work agreements) and interaction level (e.g. explanation). The ontologies are implemented into the prototype and tested on their expected outcomes (i.e. the claims). The European TRADR project (www.tradr-project.eu) provides an example in which an agent-based system for robot-assisted disaster response is developed, where response team dynamically adapt the human-robot teamwork (i.e., task allocation and coordination) to the momentary context.

1.0 INTRODUCTION

Ongoing progress in artificial intelligence, robotics, sensors and network technology feeds the development of software agents that can act as teammates or so-called *ePartners*. These ePartners are applied in defense, safety and security missions [1],[2],[3] and operate robustly in hazardous and chaotic environments. They are supposed to show responsible, agile and flexible behaviors in response to foreseen *and* unforeseen events. Furthermore, their behaviors should be fluently integrated into the overall human-agent teamwork, and harmonized with the activities and expectations of the human teammates. To establish the envisioned human-ePartner collaboration, ePartners require an understanding of human's social, cognitive, affective and physical behaviors, and the ability to engage in partnership interactions (such as explanations of task performances, and the establishment of joint goals and work agreements).

To cope with the design complexity of these systems, theoretically and empirically grounded models are needed that capture these social, cognitive, affective and physical processes. This paper proposes organization, team and interaction design patterns with corresponding ontologies, that can be used as re-usable building blocks for human-agent collaboration in the design process. These patterns are abstractions and generalizations from successful individual design solutions. In our approach, humans and ePartners are part of a joint cognitive system that operates in a specific set of situations. In the system development phase, relevant *situation cognition* models have to be identified or constructed as core concepts of the human-ePartner collaboration. A *domain* model is constructed to explicate how cognition operates in the concerning situated collaborations. Subsequently, the *design patterns* are selected or constructed, which explicate how the actors of the joint cognitive system collaborate successfully in this domain.

As a running example, we will regard an agent-based system for robot-assisted disaster response, as used in the European TRADR project (www.tradr-project.eu). Within this application a response team needs to dynamically adapt the task allocation and coordination to the momentary context [4]. Figure 1 shows the Unmanned Ground Vehicles (UGV) and Unmanned Aerial Vehicles (UAV) that were deployed in the TRADR-project (the photos were taken during a realistic field evaluation of the TRADR-system).

Figure 1: Unmanned Ground Vehicle (UGV; left) and Unmanned Aerial Vehicle (UAV; right) during a field evaluation in the TRADR-project.

Figure 2 presents a part of the disaster response team and its structure (in this example with three human team-members, their ePartners and two robots). The team leader and her ePartner have the lead of two "units", each consisting of an operator, ePartner and robot. The domain model and cognition models of, for example workload, underpin the specification of design patterns at three levels. At the highest level, *organization* design patterns are specified about normative behaviors, such as the work and rest schedules of an organization to comply with a workload norm. At the intermediary level, successful collaborative behaviors of specific *teams* are specified, such as successful work agreements on the task (workload) allocations of a team leader, the two operators, the three ePartners and the two robots in a specific situation. At the lowest level, the *interaction* design patterns specify how the information is communicated between the actors, such as the user interfaces for sharing workload information.

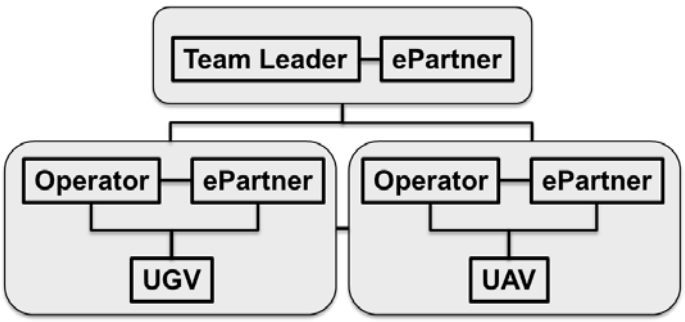


Figure 2: Part of the TRADR disaster response team with a UAV and a UGV.

Section 2 gives a brief overview of a situated Cognitive Engineering (sCE) method for the incremental development of the design patterns with the corresponding domain and cognition models (e.g., "workload" and "situation awareness"), which underpin system's functionality to collaborate. Section 3 discusses a perceptual-cognitive framework to implement this functionality. Subsequently, section 4 elaborates on the

development and application of the design patterns, entailing an extendable, theory and empirical grounded, knowledge base of the ePartner functions and behaviors. Section 5 contains the conclusions of this paper.

2.0 SITUATED COGNITIVE ENGINEERING OF E-PARTNERS

Knowledge about the *operational demands*, *human factors* and *technology* (see Fig 3), underpin the requirements specification of the human-agent system. The sCE methodology supports the acquisition, construction and usage of knowledge from these three foundation components.

First, *ontologies* are used as formal, explicit specifications of the knowledge that the agents need to have about the domain ("operational demands") and the human factors (see Fig 3). An ontology is an explicit representation of declarative knowledge. It is structured around concepts, properties, and relations, which allows for automated reasoning (about this structure). For our purposes, the ontology encompasses consensual knowledge that is important for humans-agent collaboration. Also here, we aim at an extendable set of building blocks, i.e., Ontology design Patterns (OP; [5]). They distinguish knowledge at the upper ('classes'), and lower level ('individuals'). To systematically incorporate human factors knowledge, relevant concepts of situated cognition are formalized as an ontological model, such as situation awareness [6] and workload [7]. For example, Harbers and Neerincx [7] provide a workload model, which distinguishes cognitive load, affective load and mental effort, and can reason about task allocations. The upper level workload ontology (or parts of the ontology) has been studied in and applied to several domains: disaster response [8], space missions [9], ship control [10], and train control [7]. During operation, the domain ontology is added to instantiate the concerning concepts, e.g. for the computational reasoning (such as task allocation based on "urgency" and "workload"). Part of the domain ontology for disaster response is shown in figure 4 (for more information on the domain ontology, see [11]). Note that the ontology refers to human factors concepts (like workload) that are specified in the "Situated Cognition" ontologies.

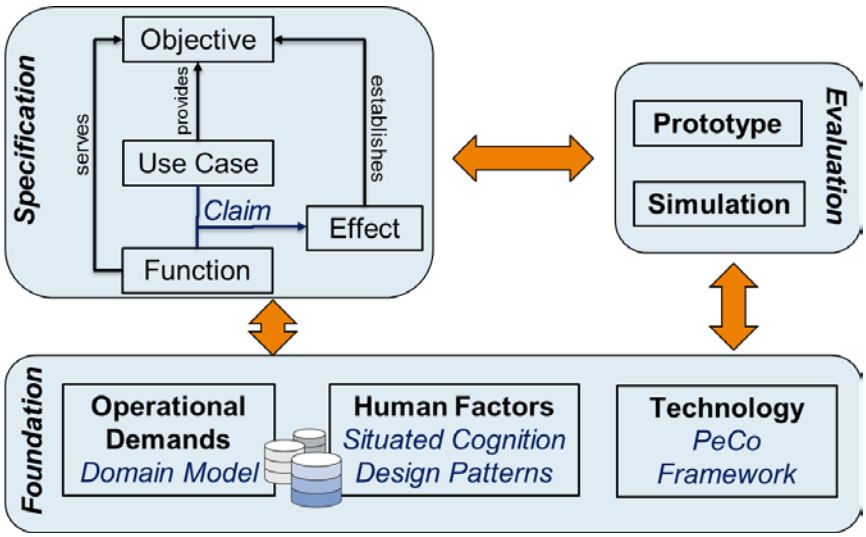


Figure 3: Situated Cognitive Engineering entails an iterative grounding ("foundation"), requirements derivation with its rationale ("specification"), and prototype and simulation testing ("evaluation").

Second, to build on past and current agent systems, and to relate the concerning agent behaviors to each other, it is important to understand and share agents' *design rationales*. In the specification or the functional requirements (see Fig 3), *claims* explicate the design rationale of a specific function: The expected effect of this functionality in a (set of) use case(s). In other words, claims describe the effects of agents behavior on

the condition and performance of (the ensemble of) team actors; the effects can entail an expected positive trade-off of the advantages (upsides) over the disadvantages (downsides). Explicating possible downsides is important, because they have to be checked in the evaluation. For example, a contrastive explanation function for a task reallocation advice might be expected to enhance the trust of a team-leader in the advising agent, with only a minor increase of workload. Claims have to be testable and, if appropriate, refer to stakeholders' values.

Third, to generalize over functions and claims, so-called "*design patterns*" are used to capture the "generic" design rationale by explicating common solutions for classes of problems (e.g., as formulated in a use case [12]). As far as possible, proven models and solutions have to be looked for and maintained as an accessible library in the foundation component. Section 4 will elaborate on the identification and construction of design patterns.

Fourth, the design rationales with their claims are evaluated via simulation and prototyping, resulting in approval, rejection or refinement of the concerning models. In TRADR, the ontology, team and interaction design patterns have been developed and tested in several cycles, applying simulations and realistic field tests with a complete team, including unmanned ground vehicles and unmanned aerial vehicles.

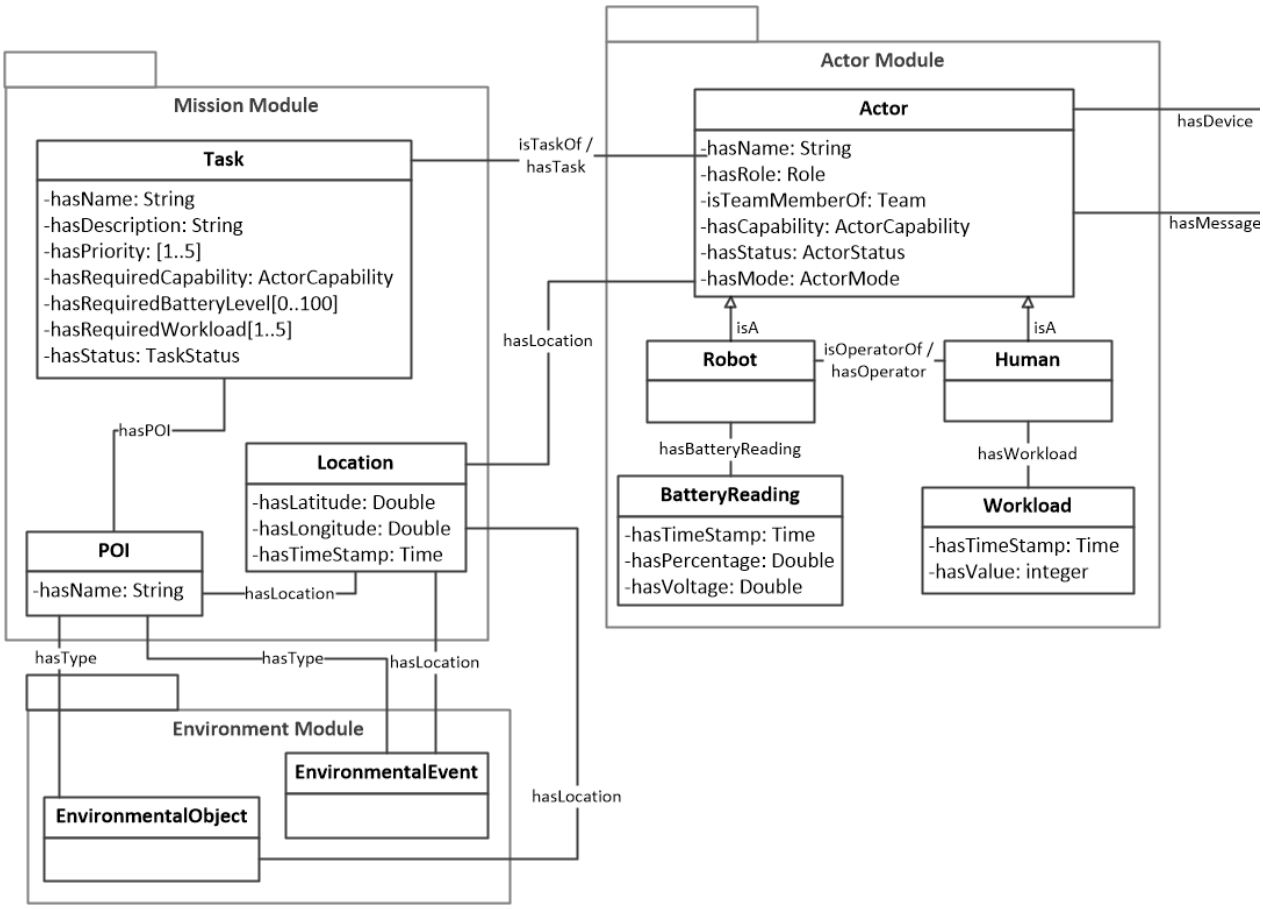


Figure 4: Part of the TRADR ontology (domain model of sCE in Fig 3; more information on the TRADR ontology, see [11]).

3.0 PERCEPTUAL AND COGNITIVE E-PARTNER FRAMEWORK

As a team-member, ePartners need to acquire and show understanding of the meaning, weighting and implementation method of the joint goals they are committed to. This type of understanding and commitment is being built and communicated via three partnership interactions in Human-Agent Teams. Currently, a generic perceptual-cognitive (PeCo) framework-suite is being developed for this purpose ([15], [16], see Fig. 5). This PeCo-framework distinguishes three components: An Objective, Work Agreement and Explanation component.

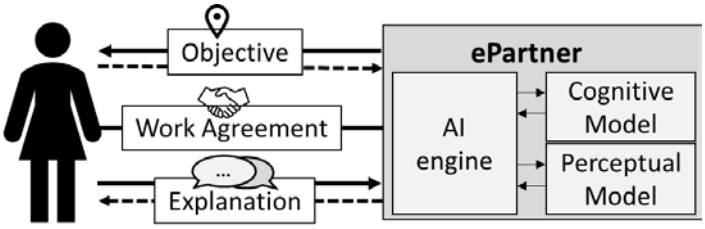


Figure 5: Three human-agent partnership interactions: objective sharing, harmonizing activities via work agreements and explaining task outcomes.

First, active partnerships require the setting, progress monitoring, refinement and adjustment of joint *objectives*, based on a shared understanding, awareness and communication of the objectives and their situated urgencies. Stakeholders' values, strategic intentions and momentary context conditions determine the urgency and feasibility of the objectives to pursue. The choices of the objectives can require re-assessment and adjustment to deal with changing circumstances, possibly just in advance or even during the work processes. Objective ontology design patterns [5] [17] [18] and utility elicitation methods [19] are being developed to support the setting of joint objectives. Adaptivity is an important capability of human-machine teams, which may involve creativity to find a new work process (possibly with an adjusted objective) when the current process is expected to fail. Humans perceive, associate and reason differently than machines; consequently, the perspective-taking and solution space for new (possibly wicked) problems is larger when machines and humans share their perceptions, associations and reasoning.

Second, the proposed human-agent partnerships involve explicit, univocal and possibly adjustable, Work Agreements (WA) on how the tasks should be allocated and performed [16]. Similar to the modeling of the objectives, an ontology design pattern has been developed for work agreements, defining the knowledge required for a team member to set, reason about and adjust work agreements. The WA-ontology defines core concepts with their relations encompassing the knowledge to specify, activate, monitor, and reason about work agreements: <***creditor, debtor, antecedent, consequent, lifespan, acceptance***>. This ontology was implemented and tested for human-robot partnerships in disaster response teams (i.e., in the TRADR-project). Such a team needs to dynamically adapt the task allocation and coordination to the momentary context. A disaster response field test showed that the fire-brigade officers could specify work agreements as required, which, subsequently, brought forward the desired adaptive team behavior of the concerning robot.

Third, we developed explanation models for ePartners that should invoke adequate trust calibration (i.e., to mitigate over- or under-reliance) and improve the learning of the team-members. Currently, a generic perceptual-cognitive explanation (PeCo) framework is being developed for this purpose [15]. The perceptual level provides an Intuitive Confidence Measure [18] and identifies the "foil" that can be used in a contrastive explanation [19]. Such an explanation fits with human explanation dialogues, centering on the question "Why this output (the fact) instead of that output (the foil)?" It reduces the factors in the explanation to the ones that are of interest to the human, so that it can be better interpreted by the human. The cognitive level provides the beliefs, goals and emotions for explanations [15]. Initial evaluations showed that humans

can understand the explanations that PeCo will provide, but further personalization and contextualization is needed.

4.0 DESIGN PATTERNS FOR HUMAN-E-PARTNER COLLABORATION

As mentioned in section 2, design patterns explicate the rationale of the chosen design solution as a common solution for a class of design problems. The following procedure aims at a common pattern library for human-agent collaboration:

1. *Identify key design problems for specific functionality*
2. *Search for available design patterns*
3. *If no pattern can be found,*
   and if it is a general, recurrent design problem:
   - *Start with a Proto Pattern, a pattern "in construction", i.e., a design problem and solution documented in a pattern form (yet lacking empirical grounding)*
4. *Provide different instantiations (examples)*
5. *Test, refine and validate these examples*
6. *If successful:*
   - *Make the Design Pattern accessible in library (of best practices)*

In our approach, the pattern language comprises a structure, template and ontology. Patterns can be constructed at different abstraction levels with different types of relationships, forming the building blocks of the research and development of team agents. We distinguish three pattern levels: Organization (e.g., work and rest schedules), team (e.g., constructive vs. destructive) and interaction (e.g., direct feedback).

Organizations often maintain norms for the joint activities of the workers. These norms can be explicated and formalized as policies: Explicit enforceable constraints on human and agents performance in a given situation, as authorization ("permissions") and obligation [23]. This way, policies govern the behavior of specific agent roles within the organization to enforce normative behaviors (i.e., they apply to a set of actors). The normative model allows to reason about (and to verify) obligations, permissions, and prohibitions, based on deontic logic [25]. Example policies for human-robot collaborations come from the space [23] and automotive domain [24]. In the last domain, the policies specified when an autonomous car should provide the driver with personalized supportive information, and when to hand over control to the driver. Both policies were linked to the driver's workload ([24], cf. the workload ontology of section 2). In these examples and in the TRADR project, organization design patterns have not yet been formulated. However, we have identified some "Proto Patterns" (see above) that need to refined and validated, such as the rest and work schedule pattern mentioned in the introduction of this paper, and design patterns for dealing with hazardous substances in populated areas. Following such patterns, the fire fighters and robots would operate within the safety and health norms of the organization. Van Diggelen et al. [13] provide a design pattern language for teamwork, which seems to be a suitable starting point to describe the normative joint activities at the level of the organization.

Team design patterns specify generic reusable behaviors of actors for supporting effective and resilient teamwork [13]. These patterns can be used to describe the dynamics of teamwork over longer periods of time. They can be constructive (i.e. leading to a more coherent team) or destructive (i.e. leading to a less coherent team), which can be observed in the behavior of an individual or group of actors. During a field test, for example, the TRADR team showed an individual, destructive micro-management pattern of the team-leader. By formalizing and implementing this pattern in the PeCO-framework, the agent can mitigate this behavior in future situations.

Interaction design patterns describe the shaping of the (multimodal) human-agent interactions, such as the dialogue acts for work agreements in TRADR [16]. Figure 6 presents part of the Interaction Design patterns for Human-Agent Collaboration (HAC) that have been specified so far. It distinguishes patterns for a Tailored Situation View (SV), Adjustable Work Agreements (WA) and Harmonized Interactive Notifications (IN) (see [13] for more details on these patterns).

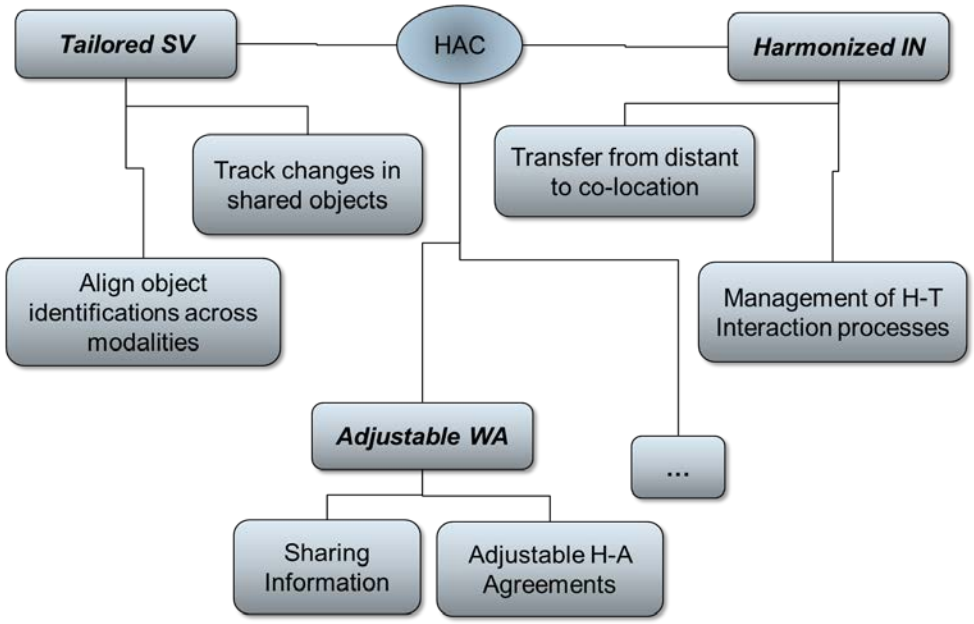


Figure 6: Part of the Interaction Design patterns for Human-Agent Collaboration (HAC), distinguishing patterns for a Tailored Situation View (SV), Adjustable Work Agreements (WA) and Harmonized Interactive Notifications (IN) (see [13] for more details on the patterns).

5.0 CONCLUSIONS

For the integration of robots into the operations of rescue, reconnaissance, security and defense teams, there is a need for theoretical and empirical grounded human-agent collaboration models. This paper proposes the development of organization, team and interaction design patterns with corresponding ontologies, as re-usable building blocks of human-agent collaboration. A situated Cognitive Engineering (sCE) methodology supports the incremental development and implementation of these building blocks (i.e., ePartner abilities and behaviors) into a perceptual-cognitive agent framework. Core cognition concepts are formalized as ontology design patterns (e.g., “workload” and “situation awareness”) and related to the concerning domain ontology (e.g., disaster response with concepts like “point of interest” and “urgency”). In parallel, relevant design patterns are identified or constructed with the corresponding ontologies, at the organization (e.g., norms), team (e.g., work agreements, and interaction level (e.g. explanation). The ontologies are implemented into the prototype and tested on their expected outcomes (i.e. the claims).

A key proposal of this paper is to build on past and current agent systems via the specification and sharing of *design rationales*. This rationale includes the explication of testable claims ***and*** the derivation of design patterns. The design patterns are formulated at three abstraction levels as building blocks of the research and development of team agents or ePartners: Organization, team and interaction. A second key proposal is to use *ontologies* that formalize human factors concepts and relate them to the domain concepts. These ontologies underpin the knowledge base (“beliefs”) and reasoning of the TRADR agents, for example for proposing, adjusting, accepting, fulfilling or rejecting a work agreement. By formalizing and implementing

constructive and destructive team patterns in the knowledge base, the agent can detect and support constructive behaviors and mitigate destructive behaviors.

Following the proposed methodology, robots can evolve as partners in disaster response by instantiating design patterns for the (a) sharing objectives, beliefs and experiences, (b) committing to norms and work agreements, and (c) uptake and learning of explanations and feedback (Fig 5).

ACKNOWLEDGEMENTS

This work was partly funded by the EU FP7 TRADR project (grant no. 60963) and the Dutch Ministry of Defence's research program (project V1719). We thank the fire fighters from the "Gezamenlijke Brandweer Rotterdam" in the Netherlands for participating in this study.

REFERENCES

[1] Neerincx, M. A. (2011). Situated cognitive engineering for crew support in space. *Personal and Ubiquitous Computing, 15*(5), 445-456

[2] Bosse, T., Breebaart, L., Diggelen, J. V., Neerincx, M. A., Rosa, J., & Smets, N. J. (2017). Developing ePartners for human-robot teams in space based on ontologies and formal abstraction hierarchies. *International Journal of Agent-Oriented Software Engineering, 5*(4), 366-398.

[3] Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V., & Johnson, W. W. (2017, July). Why Human-Autonomy Teaming?. In *International Conference on Applied Human Factors and Ergonomics* (pp. 3-11). Springer, Cham, Germany.

[4] Kruijff-Korbayová, I., Colas, F., Gianni, M., Pirri, F., de Greeff, J., Hindriks, K., Neerincx, M.A., Ögren, P., Svoboda, T. & Worst, R. (2015). TRADR project: Long-term human-robot teaming for robot assisted disaster response. *KI-Künstliche Intelligenz, 29*(2), 193-201.

[5] Gangemi, A. and Presutti, V. (2009). Ontology design patterns. In *Handbook on ontologies*, 221-243. Springer, Berlin, Heidelberg, Germany.

[6] Smets, N. J., Neerincx, M. A., Jonker, C. M., & Båberg, F. (2017, March). Ontology-based situation awareness support for shared control. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 289-290). ACM.

[7] Harbers, M., & Neerincx, M. A. (2017). Value sensitive design of a virtual assistant for workload harmonization in teams. *Cognition, Technology & Work, 19*(2-3), 329-343.

[8] Colin, T. R., Mioch, T., Smets, N. J., & Neerincx, M. A. (2012). Estimating an operator's cognitive state in real time: A user modeling approach. In *RO-MAN* (pp. 627-633).

[9] Cohen, I., den Braber, N., Smets, N. J., van Diggelen, J., Brinkman, W. P., & Neerincx, M. A. (2016). Work content influences on cognitive task load, emotional state and performance during a simulated 520-days' Mars mission. *Computers in Human Behavior*, *55*, 642-652.

[10] Neerincx, M. A., Kennedie, S., Grootjen, M., & Grootjen, F. (2009). Modeling the cognitive task load and performance of naval operators. In *International Conference on Foundations of Augmented Cognition* (pp. 260-269). Springer, Berlin, Heidelberg.

[11] Looije, R., Neerincx, M. A., & Hindriks, K. V. (2017). Specifying and testing the design rationale of social robots for behavior change in children. *Cognitive Systems Research*, *43*, 250-265.

[12] Schulte, A., Donath, D., & Lange, D. S. (2016). Design patterns for human-cognitive agent teaming. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 231-243). Springer, Cham.

[13] Neerincx, M. A., van Diggelen, J., & van Breda, L. (2016). Interaction design patterns for adaptive human-agent-robot teamwork in high-risk domains. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 211-220). Springer, Cham, Germany.

[14] Mioch, T., Peeters, M.M.M. and Neerincx, M.A. (2018). Improving Adaptive Human-Robot Cooperation through Work Agreements. *RO-MAN2018 27th IEEE International Symposium on Robot and Human Interactive Communication.*

[15] Saad, E., Hindriks, K. V., & Neerincx, M. A. Ontology Design for Task Allocation and Management in Urban Search and Rescue Missions. *ICAART 2018.*

[16] Van Diggelen, Neerincx, M.A., Peeters, M.M.M., and Schraagen, J.M.S. (in press). Developing effective and resilient human-agent teamwork using team design patterns. *IEEE Intelligent Systems*.

[17] Mioch, T., Ledegang, W., Paulissen, R., Neerincx, M. A., & van Diggelen, J. (2014). Interaction design patterns for coherent and re-usable shape specifications of human-robot collaboration. In *Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems* (pp. 75-83). Association for Computing machinery (ACM).

[18] Neerincx, M.A. , Van der Waa, J. , Kaptein, F. and Van Diggelen, J. (2018). Using perceptual and cognitive explanations for enhanced human-agent team performance. *EPCE 2018, LNAI 10906*, pp. 1–11, Springer.

[19] Peeters, M.M.M. and de Visser, E.J. (2018). "Longitudinal trust development in human-agent-robot teams," in *Workshop on Longitudinal Human-Robot Teaming, held at the Human-Robot Interaction Conference*, February 2018, Chicago, IL, USA.

[20] González-Ortega, J., Radovic, V., and Insua, D.R. (2018). Utility elicitation. In *Elicitation*, pp. 241-264. Springer, Cham

[21] Van der Waa, J., Diggelen, J. and Neerincx, M.A. (2018). The design and validation of an intuitive certainty measure. *Proceedings of IUI 2018 Workshop on Explainable Smart Systems*.

[22] Van der Waa, J., Robeer, M..J., Van Diggelen, J., Brinkhuis, M.J.S. and Neerincx, M.A. (2018). Contrastive Explanations with Local Foil Trees. *IJCAI 2018 Workshop on Human Interpretability in Machine Learning (WHI).*

[23] Van Diggelen, J. M. Bradshaw, J.M., Johnson, M. Uszok, A. and Feltovich, P.J. (2010). Implementing collective obligations in human-agent teams using KAOS policies. In: *Coordination, Organizations, Institutions and Norms in Agent Systems V*. Springer, pp. 36–52.

[24] Mioch, T., Kroon, L., & Neerincx, M. A. (2017). Driver readiness model for regulating the transfer from automation to human control. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (pp. 205-213). ACM.

[25] Wieringa, R.J. and Meyer, J.-J. C. (1993). Applications of deontic logic in computer science: A concise overview. *Deontic logic in computer science*, pp. 17–40, 1993.

On Human-Assisted Multiagent-Based Planning

Dr. Leah Kelley and Dr. Michael Ouimet
SPAWAR Systems Center Pacific
53560 Hull Street, San Diego, CA 92152-5001
USA

kelleylc@spawar.navy.mil ouimet@spawar.navy.mil

ABSTRACT

One of the goals of the U.S. Third Offset Strategy is to develop a truly collaborative human-machine fighting force. Currently, the state of the art in the field is a single soldier or sailor providing direct supervisory control of a machine agent, requiring the soldier's full attention. A truly collaborative force requires capable machine agents able to generate and execute courses of action without constant, direct human supervision. At the same time, the human teammates must still be able to affect the machine agents' decision-making, and vice versa. This paper addresses the topic of human-machine interaction, communication and cooperation applied to Artificial Intelligence planning.

Automated machine-based agents interact with the world in two ways: they react to external stimuli or deliberately execute previously generated courses of action. Reactive behaviours, such as a ground vehicle swerving to avoid obstacles, happen in extremely short time frames and address short term needs. Not only is there little time for human intervention during execution, there is also little need, since the decision making complexity for execution of short term, immediate goals (e.g. avoid the obstacle in the path) is low. In contrast, generation of deliberative courses of action is more complex, and requires goal-based decision making: given a causal model of the environment, a set of objectives, and a belief about its current state, a machine-based agent can formulate a course of action – a plan. This form of goal-based decision-making is called "Planning" by the Artificial Intelligence (AI) community.

The AI Planning community has developed planning algorithms over the past several decades. Many of these algorithms work well when applied to well-constrained planning problems, where assumptions of turn taking and a static world hold. Even games as complex as Heads-Up No-Limit Texas Hold 'Em poker are well-constrained when compared with the dynamic, uncertain warfighting environment. When the current planning algorithms are applied to real-world, military-relevant problems, the domain space the planning agent must reason over explodes such that its planner can't find a solution at all, or by the time it finds a solution, the time to execute it has passed.

In this research, we hypothesize that human knowledge can be injected into the AI planning process to help pare down the search space and assist the machine agents' planning by providing additional intelligence in the form of constraints and goals dynamically during plan generation and execution. We start with a simple representation of a naval-relevant planning problem and systematically increase complexity by removing underlying assumptions that make the problem tractable. As we remove assumptions, we find ways to inject human intelligence into the planning process, enabling the machine agent to reason over more complex scenarios. We apply this method to multi-agent mobile target search scenarios in simulation, and have developed ways for a human user to interact with several AI agents' planning without requiring the user to understand the complex mathematics behind the planning algorithms. This paper addresses our methodology, development of a human-user interface, and development of an experimental framework for measuring the effectiveness of human intelligence injection into AI planning through human subject trials.

1.0 INTRODUCTION

Recently, the U.S. Military has expressed a desire for a fully collaborative human-machine fighting force, and also seeks to find ways to enable a single or small group of soldiers or sailors to oversee many machine agents, e.g. in a robotic swarm [1]. Currently, at best a single soldier or sailor oversees a single machine agent, and in many cases a machine agent requires several humans for oversight and mission support. In this paradigm, the human operator is required to make very low-level decisions for the agent, such as supplying a fine waypoint trajectory for its motion, or even tele-operating it. The human also has to closely monitor the machine's behaviour, and may be required to take emergency action on the machine's behalf. Such tight supervision is not practical when overseeing more than a few agents [2].

Several technical advances are needed in order to transition from today's single robot-single human operator paradigm to a truly collaborative multi-machine, multi-human force. These include, but are not limited to, advances in machine sensing, perception and reasoning, and advances in command and control interfaces for facilitating human interaction with and supervision of multiple agents. This research focuses on enhancing machine reasoning and decision-making by incorporating human decision-making into machine goal-based reasoning, as well as on effective means of communicating machine agents' decision-making and reasoning to human operators for effective teaming.

When a machine agent interacts with the world, it can respond to external stimuli, as a mobile robot avoids obstacles or as a digital assistant searches for, and then answers, a query. However, if a machine agent is to achieve a longer-term goal, such as reach a destination across a room, efficiently perform a series of tasks, or play a game such as Chess, it has to generate a course of action to achieve its goal. This goal-based reasoning process is called "Planning" by the Artificial Intelligence (AI) community. If humans are to effectively team with machine agents, and are to effectively supervise many machine agents, they need to be able to influence the machines' planning processes and vice versa.

AI Planning is an active area of research and has been over the past several decades [3, 4]. Much of the work has been focused on domain-independent planning algorithms that are capable of solving wide varieties of planning problems in different application domains. To use such planning algorithms, one must provide the planner with a model of the application domain, a planning problem definition, a goal to achieve and an initial starting condition. This requires that the user not only understand the application domain and the problem, but must also have understanding of how to formulate them so that the solver can interpret and solve the problem. Additionally, many of the existing planners require that the domain be modelled using discrete representations of variables, that the assumption that the domain remains static during planning holds, that the agent knows all relevant facts about the domain and that the agent's sensing is perfect. Though the AI planning community continues to improve planning algorithms and has addressed some of the above limitations, they still cannot handle all real-world complexities, especially as found in highly dynamic, uncertain domains relevant to successful warfighting.

1.1 Research Approach

This research develops a methodology to mature AI planning algorithms to make them applicable to realistic, military-relevant scenarios. Given a simplified version of a military-relevant problem, we apply an appropriate AI planning algorithm to it, and see how it performs. If the algorithm is able to generate a plan for the simplified problem, we remove one of the simplifying assumptions and re-run the planner. If the planner is still successful, we remove another simplifying assumption. We continue removing assumptions until the planner is unable to generate a solution. At this point, we then find ways to inject human intelligence/help into the planning so that it is able to generate solutions. We then remove another assumption, find where the planner is unsuccessful again, and find additional ways to inject human

intelligence and help into the planning process. Using this systematic approach, we enable AI planners to address more realistic planning problems (see Figure 1).

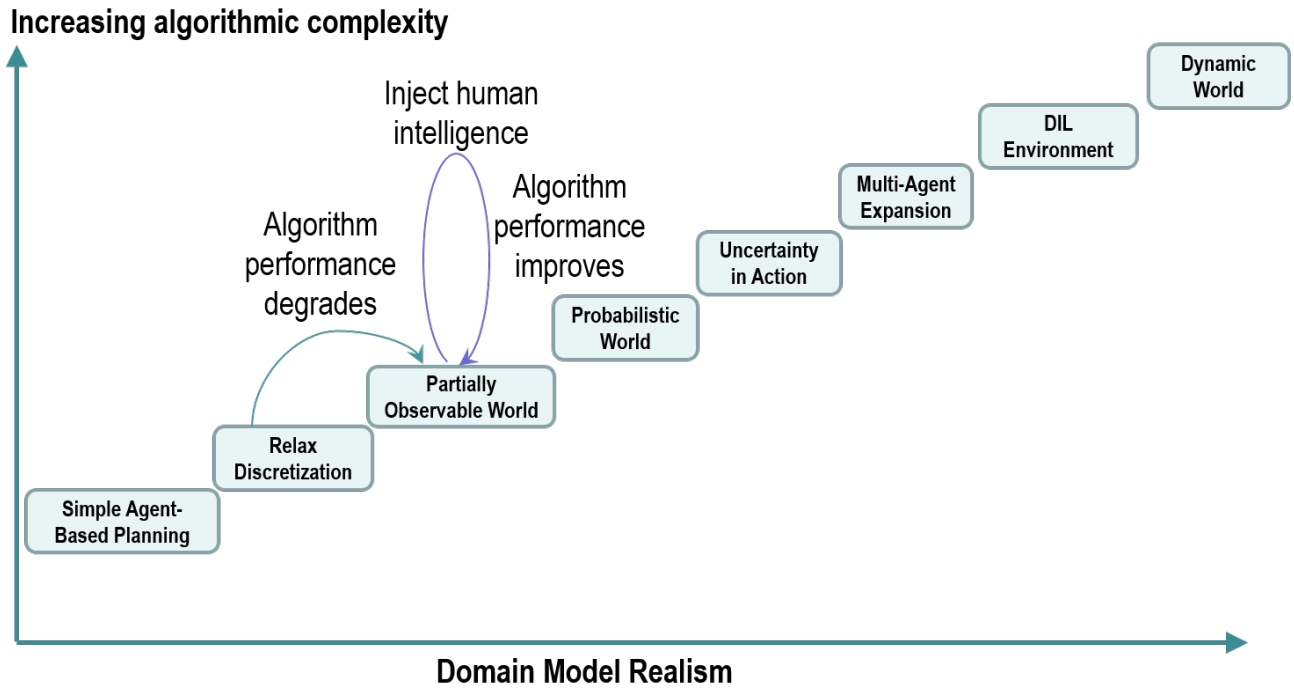


Figure 1: Assumption removal and human intelligence injection process

This research also addresses methods for communicating the AI agent's goals, beliefs and needs to a human supervisor so that s/he can assist the planning agent. Unlike previous work in semi-autonomous systems, which focuses on methods to assist a robot during task execution [5], this work has similarity with mixed-initiative planning [6-8], where a human and AI agent interact in order to jointly form a plan before task execution. The mixed-initiative planning work cited focuses on tight interactions between the machine and the human, where the machine agent poses a series of questions to a human and explicitly asks for human input, and hence the need to create complex domain models is circumvented by having the human present in the planning as an active participant. However, as with tight supervisory robot control fielded today, a tight-knit and highly interactive planning approach doesn't scale to situations where one human must supervise many machine AI agents, as in swarming applications. This research seeks to find ways to facilitate human assistance to the AI agents without requiring the agents to explicitly request help, and also without requiring tight supervision of each agent from a human operator, so decision-making is less tightly coupled than in mixed initiative planning. The goal is to find ways to couple the human abilities of quick decision-making under highly dynamic situations with the computational strengths of AI planning agents so that together humans and machines can accomplish complex tasks neither can do (either well, quickly or at all) without help from the other. To facilitate such cooperation and interaction, the agents' goals, beliefs and needs have to be displayed to the human in an easily interpretable manner, and the human needs to be able to provide information in a form the machine agents can understand and use.

The rest of this paper is organized as follows. Section 2 describes the current state of the art of AI Planning, as well as our prior work on injecting human intelligence into the planning process of a single AI agent in a naval-relevant scenario. Section 3 introduces our notion of hierarchical planning and describes the extension of our single human-single agent approach to a single human-multi-AI agent persistent search scenario. Section 4 describes our approach to and development of a human subject test and experimental framework.

As of the time of writing, the team is still awaiting permission to officially begin the human subjects test. Section 5 concludes the paper and discusses expected future outcomes and directions.

2.0 BACKGROUND AND PRIOR WORK

2.1 State of the Art of AI Planning

The AI community has developed several types of planning algorithms and techniques over the last several decades, with primary emphasis on developing domain-independent planners that can solve problems in a wide variety of application domains. Early methods, often referred to as "Classical Planning," use situational calculus to find a set of transitions from an initial state to a goal state given a domain model that encodes all possible state transitions using some form of formal logic, such as prepositional or first order logic [3]. These methods assume the world is static, time is atomic, only the agent causes changes to the state of the world, the agent's actions are deterministic, the agent has complete domain knowledge, the domain is not affected by external influences (i.e. if the initial conditions are completely known, then the action sequences generated will be completely and correctly predicted), and the planning goals are known. Such assumptions are valid even for complex games such as Chess and Go, but do not hold for many real-world problems.

State-space search methods have been used to solve classical planning problems. Such methods include tree-based and graph-based search, where states an agent and the environment are junctions or nodes, and transitions between states are "branches" or "edges." Search algorithms can start from either the initial state or goal state to build plans (forward search and backward search, respectively), and can either explore down one path (depth first) or explore all transitions from the starting node (breadth-first). The planning community has developed heuristics for choosing between alternatives, in order to speed up plan generation. A* search is one well-known best-first search method [3]. Other heuristic search methods include hill-climbing, simulated annealing, genetic algorithms and belief-state search for partially observable environments. Though early heuristic functions were encoded by hand, methods exist to generate heuristics for particular problems automatically by analysing the domains [4].

Planning problems can be formulated as constraint satisfaction problems (CSPs). In constraint satisfaction, a solver seeks to find an assignment of permissible values to a set of variables such that the assignment satisfies all of the constraints that dictate how the variables interact and/or combine. CSPs can be defined using prepositional logic, Boolean formulas or mathematical formulas. Search methods used for classical planning can be applied to constraint satisfaction problems. Additional methods for solving CSPs include constraint propagation, backtracking, back-jumping and model checking [3]. Though methods such as model checking can solve large CSPs, such problems still require assumptions of determinism (no uncertainty) and static domains with no external influences.

Domain-specific planning algorithms highly tailored to particular applications and problems also exist. For example, Google's Optimization Tools contain specialized algorithms that solve combinatorial optimization problems such as Traveling Salesman Problems, knapsack problems, linear assignment problems, bin packing problems, and vehicle routing problems, as well as more general constraint and linear programming solvers [9]. In order to use such solvers, one must cast one's planning problem into the particular form required by the planning algorithm, and the same assumptions used when encoding the domain must apply to any new planning problems one tries to solve.

The AI community has also developed planning methods that are able to cope with more complex scenarios. Hierarchical Task Networks (HTNs) divide complex planning problems into smaller and smaller sub-problems until sub-problems are fully decomposed into sets of primitive actions. Sub-problems, called higher level actions (HLAs), may be realizable by various combinations of primitive actions. When solving HTNs, a planner may reason over the higher-level actions, letting an agent determine the primitives to use

during solution execution [3]. The Simple Hierarchical Ordered Planner (SHOP) and SHOP2 HTN planners generate plans with steps in their execution order, reducing complexity [10], and have been extended to handle nonlinear and continuous effects [11] as well as user preferences [12]. Case-based planning methods use databases of previously generated plans to form solutions to new planning problems. This method is applicable to planning problems that exist within one application domain, and where new planning problems are assumed to be similar to previously solved problems. Reuse of previous solutions is intended to reduce computational complexity [13]. However, the assumption that new problems are similar to old problems is not necessarily valid.

Planning methods that handle uncertainty have been developed, and include contingency planning, Markov Decision Process (MDPs) and Partially Observable Markov Decision Process (POMDPs) iteration, and inference over statistically modelled processes. In contingency planning, the domain is modelled as an "And-Or" graph and online planners generate solutions using depth-based search, random walk, hill-climbing or Learning Real-Time A* (LRTA*) [3]. Fully observable, stochastic environments can be represented as Markov Decision Processes (MDPs), assuming uncertainties in the environment, sensory data and effects of agent actions can be quantified [3, 14-15], and optimal solutions to problems in such representations can be found using dynamic programming. POMDPs are extensions of MDPs to domains in which an agent cannot fully observe the environment, hence its current state is also represented as a probability distribution. Methods for handling POMDP planning include back-chaining [14], an extension of value iteration [15], and branch-and-bound and gradient ascent methods [16]. Inference reasoning algorithms have been used to solve partially observable stochastic processes modelled as Bayesian Non-Parametric Models (BNPs) [17]. Finally, methods for planning for imperfect information games have been developed, and have been successful at defeating all human opponents of Heads-Up No-Limit Texas Hold 'Em poker [18], using a combination of planning over abstraction (blueprint strategy computation), real-time sub-game solving, and self-improvement game play. Note that even in imperfect information games, the assumption of a static world while planning still holds.

Model-free Reinforcement Learning algorithms have had recent success in playing videogames, starting with DQN on a variety of Atari games [19]. This method directly learns from experience and does not have any explicit planning or model of the games. DQN has best performed on games that were more tactical than strategic though where there is a rich reward signal or reactions are key. However, recent work by OpenAI on Dota 5v5 has shown the algorithm PPO can play a complex, team-based adversarial game in real-time with partial observations [20, 21]. Even though no planning is being done, the agent is able to play competitively at a game where long-term strategy, teamwork, and tactics are all important by learning from previous data and only using the most recent states as the input to the action selectors. This approach is promising since it doesn't require explicit domain modelling in advance. However, it becomes difficult for a human to collaboratively work with an agent when s/he doesn't understand the agent's decision-making process.

As problem complexity increases, more complex domain models with large state space representation are required, regardless of the methods one intends to apply to solve the planning problem. When considering large state spaces, planning algorithms require significant time to perform computations or, in some cases, cannot generate plans at all. Plan generation is PSPACE-hard even when the set of reachable states is finite – to generate the shortest set of transition actions, the number of states that must be considered from initial to goal state is exponential in in the size of the planning problem description [4]. Planning computational complexity can range from constant time to NEXPTIME-complete, with increasing complexity corresponding to fewer simplifying assumptions [22]. This research addresses the challenge of applying the state of the art planning techniques, especially those that handle some degree of uncertainty, to naval-relevant problems that are not well constrained.

2.2 Single Human – Single Agent Cooperative Planning

As a starting point for this research, we selected an underwater object search task as a representative naval-relevant scenario, and applied the systematic relaxation approach described above to it. In its most simplified version, the underwater search task is a recasting of the "Wumpus World" game [3], where the Wumpus is replaced by an object of interest and pits are replaced by underwater obstacles. Note that unlike the Wumpus game, there is no treasure and here the AI agent simply wants to find the object and interact with it. The simplest version takes place on a square grid, and the agent knows where it starts, all locations of obstacles and the object of interest and has perfect sensing and motion. Under these circumstances, planning is easily accomplished using A* search or other tree or graph-based search methods, and no human assistance is necessary. However, if this search were to be expanded into 3-dimensional space and the agent's roll, pitch and yaw were also considered, the number of states would become too large for tree-based methods to find optimal paths in reasonable time and one would need to employ heuristics to get real-time but sub-optimal results.

We were able to remove several simplifying assumptions to this scenario without needing human assistance, simply by changing planning algorithms and domain representations. By casting the search problem (still in 2 dimensional space) as an MDP, the problem becomes a simultaneous localization and mapping (SLAM) problem in addition to an object search. The SLAM problem has been well studied in the robotics and AI communities [23]. In SLAM, the agent does not have to have perfect sensing, motion, or knowledge of obstacle locations, and can still successfully navigate and localize, while building a representation of the environment. This representation is more realistic since sensors are prone to false readings, unmodelled currents can affect agent motion and it is unlikely that all underwater object locations are known a priori. This representation still assumes the search space is discretized into a grid.

In the MDP representation used here, the agent uses a reward function distributed over the entire discretized space to decide where to move next, by moving toward locations where it will receive high reward and avoiding locations where rewards are low. Its overall goal is to find a policy that will maximize its reward over time. When no information is provided about the space in advance, the agent is incentivised to search randomly. As the agent explores and senses obstacles, the value of the reward in previously explored regions decreases, and, at sensed obstacle locations, the reward drops significantly. The reward function thus evolves as the agent navigates. Further details on the implementation on the MDP iteration planning applied in this research are available [24].

This formulation provides a human with opportunities to influence the agent's planning by interacting with the reward function. A human operator can set initial goals and keep out areas, high reward and low reward areas, respectively, and can make changes to goal locations and keep-out locations on the fly by changing values of the reward function while the agent navigates. Rather than requiring a human user, especially a non-expert user, to interact with the mathematical values directly, the team developed a graphical user interface to facilitate such interactions, using the Intelligent Multi-UxV Planner for Adaptive Collaborative/Control Technologies (IMPACT) System [25, 26] as a simulation and visualization tool and user interface. The agent's search area, current location, goals and detected obstacles are displayed onscreen in one of two ways: as a greyscale occupancy map (see Figure 2, left) or as a reward function map (see Figure 2, right). The greyscale representation shows the agent's position (agent icon), user-defined agent goals (red circles), detected obstacles (black areas), previously traversed areas free of obstacles (white areas) and unexplored areas (grey areas). The reward function is depicted as a heat map overlaid on the search area, where rewards range from red to blue through a rainbow colour scheme, with red depicting high rewards and blue depicting low rewards. In the heat map representation, the user-defined goals are shown in red, detected obstacles are shown in blue, and other areas primarily green, have intermediate reward values. The reward function is structured such that its value gradually decreases away from a user-defined goal location in all directions.

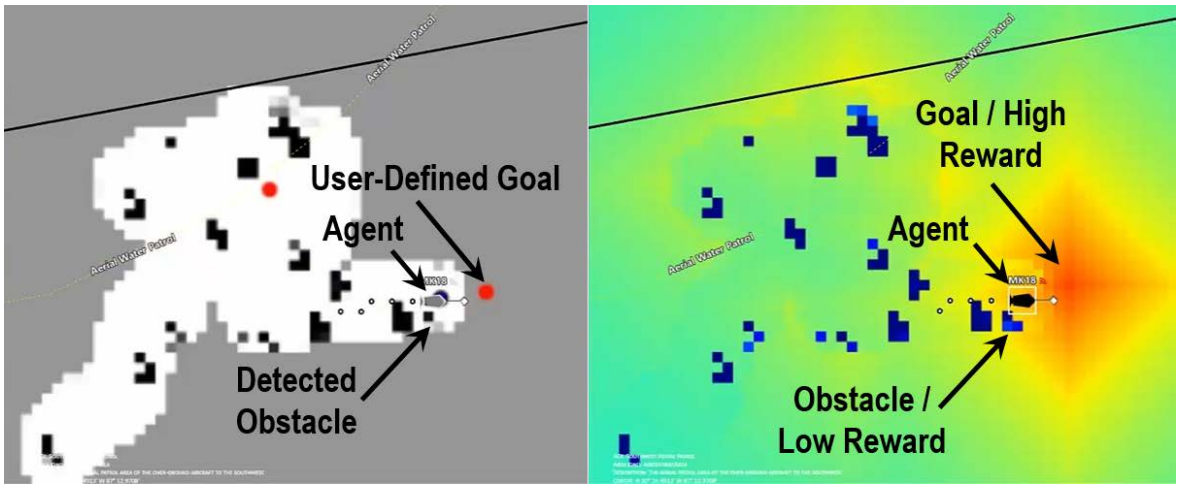


Figure 2: Occupancy map, left and reward function map, right [24], as displayed in IMPACT.

The user is able to interact with the agent's reward function using a custom task incorporated into IMPACT's Task Manager (see Figure 3). The user can toggle between the heat map (reward function map) and the greyscale map using the drop-down menu button. To change the intensity of the reward function, the user clicks on the appropriate button (Increase Intensity or Decrease Intensity) and then clicks on the search area. By changing the reward intensity graphically, the human can influence the agent's behaviour and navigation without understanding the mathematics behind the reward function or the agent's model of its motion, and can do so as the agent executes its motion.

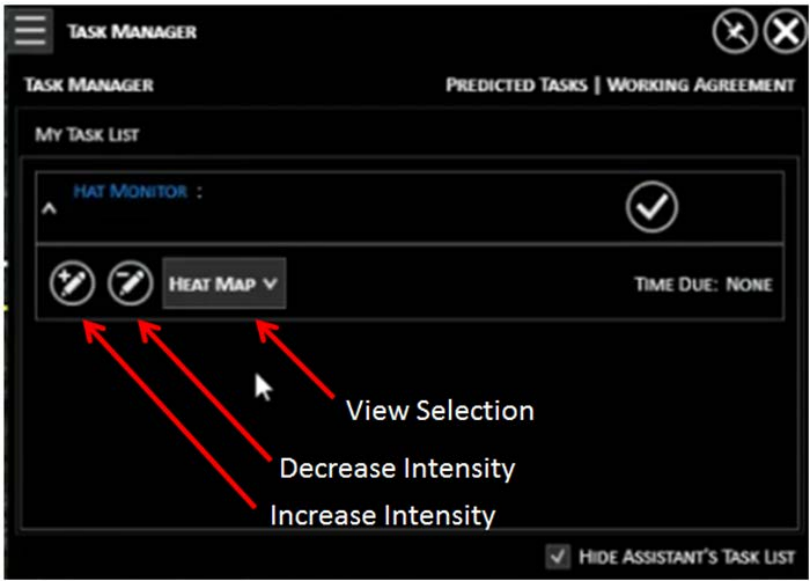


Figure 3: Task Manager controls for human operator, from [24].

3.0 HIERARCHICAL PLANNING WITH MULTIPLE AGENTS

In the single-agent mapping scenario, MDP iteration is used to fuse obstacle data and determine where the agent should go next. As the planning horizon, number and complexity of tasks, and number of agents increase, it becomes impossible to plan the whole mission at the lowest level of granularity (e.g. direct motor

control actions). Hierarchical planning allows one to solve a complex task by composing the solution from a nested hierarchy of increasingly abstracted, simplified problems. This approach can be used to guide a team of unmanned systems in mapping an area, persistent monitoring, and tracking moving targets. A concrete example is that a group of agents can monitor an area by first deciding how to split the area, potentially based on agents' different skills or battery charge (algorithm 1) and then letting each agent decide for itself the path to best monitor its assigned area (algorithm 2). At the lowest level, another planner governs the physical actuations required to maintain on course in the presence of unmodelled disturbances or unknown obstacles (algorithm 3).

There are at least two reasons for adopting a hierarchical approach: one to benefit the human and the other to benefit the autonomy. First, people have a limited attention bandwidth available and a person cannot directly control more than one vehicle at a time. Thus, in order for human-autonomy teaming with more than one robot, at least some level of hierarchy is required. Furthermore, it can be challenging for a person to even maintain situational awareness of 8-12 agents each doing their own tasks autonomously [2], let alone to aid by also directly tasking them or providing input on their paths. This leads us to conclude that people are most suited to aiding a team of robots by providing input at the higher levels of hierarchical abstraction.

As one introduces multiple agents into a strategic planning domain, the algorithmic complexity increase dramatically. The challenge comes largely from the coupling between each agents' actions. Because each agent affects the rest of the team as well as the global environment, care must be taken to ensure that all team members coordinate so they do not interfere with one another, not perform the same task twice. However, planning over a long time horizon for all agents at the lowest level of abstraction in a centralized way becomes computationally infeasible due to the combinatorial number of possible joint trajectories that must be evaluated. For this reason, we elect to solve a high level coordination/assignment problem to decide which agent(s) should do which task(s). Once this decision is made, the agents can each perform their own path planning and reactive planning to safely achieve their assigned goals. As tasks are finished or new ones arise, replanning can perform multi-agent assignment again, passing the updated goals to the lower level algorithms. A schematic of our model can be seen in Figure 4.

In addition to depicting the type of planning algorithms appropriate for use at the various levels of the hierarchy, Figure 4 depicts the levels at which various types of human intelligence are most appropriately inserted into the planning process. At the highest level, a human commander will define the overall mission goals, such as priority search locations in a multi-agent search scenario. The commander may also have preferences for which agents perform which tasks, or may specify that agents with particular abilities are suited to particular tasks. This information can be injected into task assignment planning. In addition to task assignment preferences, the human commander may also have additional information about activities or features in the search environment, such as cluttered/unsafe areas or other keep-out zones. Environmental information will be used by each agent as it plans its navigation routes, and such information may change over time, so the human commander requires methods to communicate this information to the appropriate agents. As stated earlier, it is difficult for one human to closely monitor more than 8-12 agents, so aside from occasional tele-operation (assuming the mobile agents have enough autonomous capability to handle their own well-being in most situations), human injection at the lowest levels of planning is not needed.

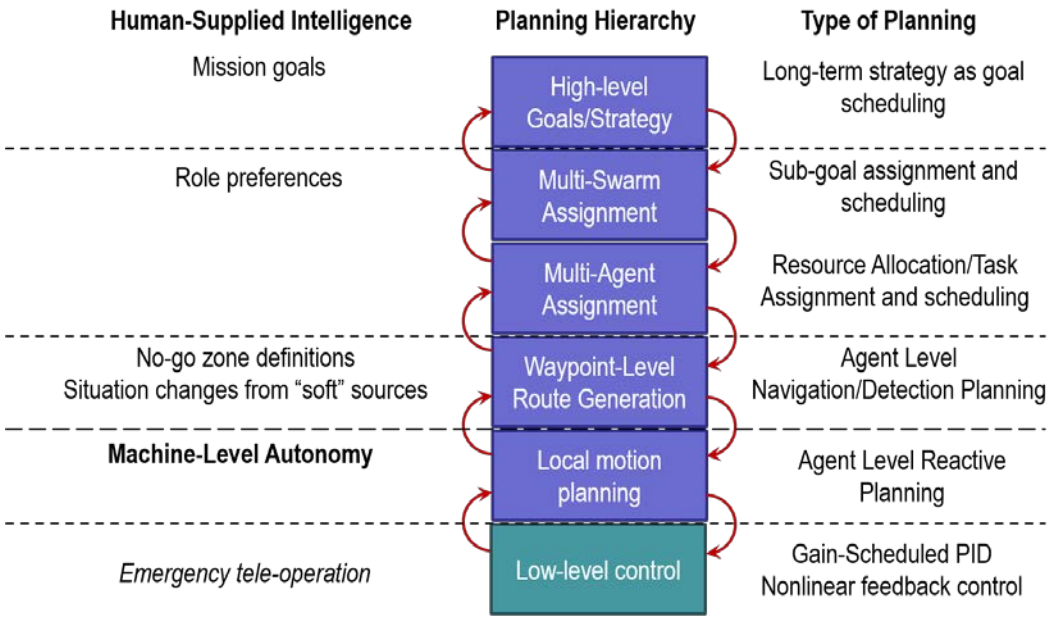


Figure 4: Hierarchical planning model, including areas for injection of human intelligence

3.1 Example: Multi-agent Persistent Monitoring

As an example of a multi-agent hierarchical planning problem, consider a 2-dimensional environment that requires persistent monitoring with multiple mobile machine agents, where the importance of maintaining watch over some locations is of greater importance than others. Following the hierarchy in Figure 4, a human commander defines locations that need to be observed, as well as assigning the level of importance to each location. The hierarchical planning process then starts, omitting the "Multi-Swarm Assignment" level in this case.

Voronoi Partitioning: Since some parts of the environment are more important than others, the highest level of the hierarchy partitions the environment into N continuous regions, one for each agent. The Voronoi partitioning algorithm [27] weights important regions more heavily so some agents may cover large swaths of unimportant areas, whereas other agents cover smaller, more important regions more frequently.

Linear Assignment: Once the optimal partitioning is determined, we use the Hungarian algorithm solution to the Linear Assignment problem [28] to match each agent with one region based on the distance to the areas' centroids. The global matching ensures that sum of distances between agents and their assigned regions is minimized.

Markov Decision Process: Each agent is then provided their region as well as the relative importance of the subareas in that region. The MDP algorithm uses the importance values as the reward for each potential location. When a location is visited, its importance is set to zero. Globally, the importance increases back over time at a rate dependent on how important a person deems that area.

4.0 HUMAN SUBJECT EXPERIMENTAL FRAMEWORK

The prior work described above established methods for non-expert users to interact with individual AI agents performing navigation and search tasks, and the team has introduced and implemented a hierarchical planning structure that can handle multi-agent persistent monitoring as well as multi-agent object search. However, to date the research team has been unable to quantify performance improvements and also has yet

to determine how well such methods scale to single-human multi-AI-agent scenarios. Additionally, the team has not gathered any data on the effectiveness of the heat map as a method for displaying an AI agent's goals and belief states to non-expert users. To gather such data, the team has developed a human subject experimental framework, test scenario and protocol, currently under review by the U.S. Navy's Human Protection Review Office.

The human subject study uses a heterogeneous multi-agent search scenario to determine whether the human input enables reduction of computational complexity and time required by the planning agents to generate successful assignments and motion paths that accomplish the search goal. A human subject plays the role of a human commander of a small fleet of 7 autonomous, unmanned surface vehicles (USVs) and 5 unmanned air vehicles (UAVs) in a fictitious, simulated coastal area of the world. The human-UxV fleet has been tasked to find an adversary who is somewhere in the region. Each USV has its own machine artificial intelligence (AI) agent capable of planning and executing its motions, using the MDP iteration method described above. The UAVs are carried by a "carrier ship" that serves as a mobile charging base, which the human participant can instruct to move to different locations. The participant can also instruct the carrier ship to deploy all of the UAVs, which then use their own MDP agents to navigate and perform search tasks while remaining close enough to the carrier ship to recharge their batteries as needed. The UAV deployment is accomplished through Voronoi partitioning. Finally, the participant has an AI assistant that can help assign the USVs to search areas in the region using a combination of Voronoi partitioning and the Hungarian method for linear assignment. Note that the human participant can decide to assign a subset of agents to specific search areas and can have the AI assistant take care of assignments for the remaining agents.

The experiment front-end is a web-based user interface that displays a map of the fictitious region where the experimental scenario occurs, icons of the autonomous agents the user can interact with on the map, as well as various GUI interfaces the user can click on (see Figure 5). This user interface is agnostic to the type of web browser a participant uses, hence one can perform the experiment from any computer on the network where the host server resides. A dedicated computer workstation hosts the front end, and also runs the experiment back-end simulations and AI planning algorithms, and serves as a data recording device.

The web-based user interface (Figure 5) contains a map of the fictitious littoral region in the upper right, a table of simulated unmanned vehicle agents on the left that includes information about each agent's state (e.g. fuel levels, current tasks assigned) and a timeline of events at the bottom. During the experiment, participants receive information about what is happening in the scenario in several ways. Agent locations are displayed on a map in the web interface as they move about the region. Their status, task assignments and power levels are displayed in a table, which dynamically updates as the agents perform tasks. A participant can click on an agent in the map, which then displays a heat-map for that agent overlaid on the main map that displays the agent's reward function over its assigned search region, similar to the heat map shown on the right side of Figure 3. As in the earlier IMPACT-based single-human-single-agent simulation, the human user can add or change the priorities of previously defined search locations by clicking on the appropriate increase or decrease buttons and then clicking on the map. A participant can also assign search areas to specific agents if they so desire, by dragging unassigned tasks to agents (Figure 5, left side), and by creating new tasks and then assigning them to agents. Such methods of interaction are intended to be intuitive, and easy to use with minimal instruction. As part of the pre-experiment brief, participants are provided with a 5-minute tutorial that describes the features of the user interface and provides a walk-through of the interactive functions.

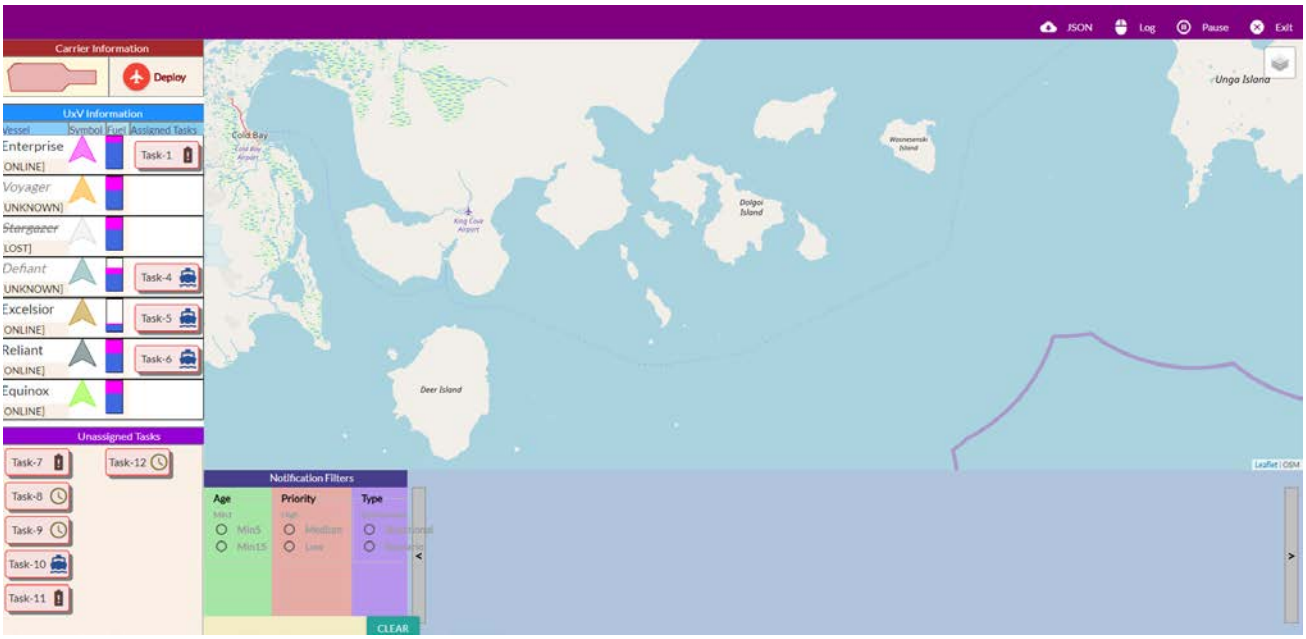


Figure 5: Human subject study web-based user interface

During the experiment, participants will also receive situationally relevant information in the event timeline at the bottom of the web-based interface (see Figure 5). Such information includes camera images from USVs and UAVs as they search the region, as well as messages about events happening in the region that the machine agents cannot understand. In this experiment, the UxVs are assumed to be able to detect objects, but cannot perform the high fidelity classification for distinguishing the adversary from other entities. Participants can adjust search goals and reassign tasks based on this information to help the UxVs find the adversary. Finally, participants can also deploy the UAV fleet to obtain higher-quality images (that will appear in the event timeline at the bottom of Figure 5) if they so choose, by clicking on the "Deploy" button shown in the upper left in Figure 5 under "Carrier Information."

The human subject study records participant performance by tracking performance metrics based on the inputs the participant provides in the web interface, as well as by tracking metrics in the back-end simulation, such as completion time and computational load. Actions taken by a participant to direct the agents (e.g. clicking on the map or other user interface elements in the GUI, change of view, examination of the screen, etc.) will be directly recorded, as will the AI planning agents' computations and actions. Immediately following an interaction with an agent (or agents), the experiment will pause and display a brief "pop-up" survey (multiple choice and short text-based answering) that asks the participant about the action just performed, and the reason the participant had for choosing to interact at the time. Such questions are designed to elicit participants' beliefs about the AI agents' behaviours, as well as to determine whether supplemental information was used to redirect/assist the AI agents. Additionally, pre-and post-experiment surveys are used to collect non-personal identifiable demographic information, such as whether or not a participant plays video games, and perceptions/understanding of the experiment, respectively. Post-experiment questioning will also be used to gauge the effectiveness of the tutorial provided to participants before beginning the experiment. The post-experiment survey also includes the NASA Task Load Index [29, 318], to be used as a supplemental measure of workload as well as to facilitate general comparison to known ranges of workload in other fields, to prior studies, and to other participant groups such as experts compared to novices.

Performance measures include the time to complete mission, the level of computational burden exhibited by the AI agents (e.g. time it takes an AI agent to generate a plan) and instances of agent stagnation during

target search. The number of algorithmic course changes detected versus no effect in algorithmic outcome will be measured through accuracy to ground-truth goal representation. Correlation of course changes with instances of human participant interaction, along with participant answers to pop-up survey questions, will be used to determine whether human intervention helped, hindered, or had no effect on the agent's search task. Analysis of a participant's answers on the questionnaire regarding their awareness of the experiment's purpose and understanding of the autonomous planning agents' actions will use the established methods of Endsley and Garland [31].

The following hypotheses about the effects human input have on the performance of AI planning algorithms will be tested using results of this experiment:

- The human-AI teams will successfully interdict the adversary more frequently than the AI alone.
- The human-AI teams will complete the interdiction mission more quickly than the AI alone.
- Those subjects with prior experience playing video games are expected to complete the interdiction task more frequently than those without prior gaming experience.
- Those subjects with prior experience playing video games are expected to complete the interdiction task more quickly than those without prior gaming experience.

Human subject tests are anticipated to start immediately upon protocol approval.

5.0 SUMMARY

In this paper we've described our approach to collaborative human-AI agent planning that can enable plan generation in highly dynamic, uncertain environments with few constraints, by injecting human intelligence into the AI agent's planning process. Our prior work focused on single-human-single-agent collaboration, and we previously demonstrated a method and developed a user interface that permits a non-expert user an intuitive means of interacting with the single agent both before and during plan execution. In this paper, we've presented our extension of the single-human-single agent collaborative planning for search tasks to single-human-multi-agent search and persistent monitoring scenarios. By constructing a hierarchical planning process using several AI agents, we decompose a highly complex scenario into smaller planning problems, where the human supplies situationally relevant information at the higher planning levels, while the individual agents take care of their own lower-level execution plans. By interpreting situationally relevant information that machine agents are incapable of understanding, the human collaborator can help guide the AI planners towards plan generation, speeding up the planning process. Human experience also serves to fill gaps in the AI agents' domain models.

We've developed methods for human interaction with AI planning agents that do not require the user to have expertise in the field of AI planning, or extensive training in order to engage and collaborate with the AI agents. Use of intuitive representations, such as color-coded occupancy and heat maps, pictorial bar charts, and dynamic displays, are intended to make the AI agents' goals, intentions and beliefs about their world states easily interpretable by non-expert users. Use of simple interaction mechanisms, such as adding or subtracting value from map locations by pointing and clicking, as well as dragging tasks into tables, are expected to be easy for non-experts to use to interact with the AI agents. We anticipate human test results will provide quantitative proof that human-influenced planning will result in better performance than if the AI agents perform the task with no additional assistance. We also anticipate qualitative results from survey questions will yield suggested improvements that can be incorporated into future user interfaces as well as suggestions for easier methods of interaction.

In addition to performing the human subject testing described here, future extensions of this work include addressing the problem of collaborative planning when AI agents do not have continuous communication

with the human supervisor as well as each other. Additional extensions include expanding the hierarchical planning structure to include scenarios requiring task scheduling for long-duration missions, as well as scenarios requiring multiple human operators collaborating with multiple AI agents. The multi-human-multi-agent scenarios introduce opportunities to incorporate methods for sharing unmanned assets already in development. Finally, this research has been conducted solely with simulated AI agents, so a natural extension is implementation of the AI planning agents on actual UxVs and ground station for real-world testing, experimentation and demonstration.

ACKNOWLDEGMENTS

The authors thank Luiz Martinez, Eric Gustafson and James Jen, the software development team instrumental in developing and building the human subject test framework, as well as integrating the single-human-single agent interaction mechanism into the IMPACT System Task Manager. The research presented here has been funded by SPAWAR Systems Center Pacific's In-House S&T Innovation Program. This paper is approved for public release.

REFERENCES

[1] Work, B. 2015. "The Third U.S. Offset Strategy and its Implications for Partners and Allies," January 28, 2015. Willard Hotel, Washington, DC.

[2] Wang, H., M. Lewis, P. Velagapudi, P. Scerri, and K. Sycara. 2009. "How search and its subtasks scale in N robots." In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 141-148). ACM. 9-13 March. La Jolla, California, United States.

[3] Russell, S., and P. Norvig. 2010. *Artificial Intelligence: A Modern Approach*. 3rd ed. Prentice Hall, Upper Saddle River, NJ.

[4] Long, D., and M. Fox. 2002. "Progress in AI Planning Research and Applications," *UPGRADE: The European Journal for the Informatics Professional 3(5)*:10–25.

[5] Zilberstein, S. 2015. "Building Strong Semi-Autonomous Systems." In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, (pp. 4088-4092). 29-30 January, Austin, Texas, United States.

[6] Allen, J. F., C. I. Guinn, and E. Horvtz. 1999. "Mixed-initiative interaction." *IEEE Intelligent Systems and their Applications 14(5)*: 14-23.

[7] Ferguson, G., J. F. Allen and B. W. Miller. 1996. "TRAINS-95: Towards a Mixed-Initiative Planning Assistant." In *Proceedings of the Third International Conference on Artificial Intelligence Planning Systems (AIPS 1996)*, (pp. 70-77). AAAI, 29-31 May, Edinburgh, Scotland.

[8] Bresina, J. L., A. K. Jónsson, P. H. Morris and K. Rajan. 2005. "Mixed Initiative Planning in MAPGEN: Capabilities and Shortcomings," In *Workshop on Mixed-Initiative Planning and Scheduling at the International Conference on Automated Planning and Scheduling (ICAPS 2005)*, (pp. 54-61). 5-10 June, Monterey, California, United States.

[9] *Google Optimization Tools*. 10 July 2018. Google. https://developers.google.com/optimization/. Accessed 14 August 2018.

[10] Nau, D., T.-C. Au, O. Ilghami, U. Kuter, J. W. Murdock, D. Wu, and F. Yaman. 2003. "SHOP2: An

HTN Planning System," *Journal of Artificial Intelligence Research (JAIR) 20*:379–404.

[11] Molineaux, M., M. Klenk, and D. Aha. 2010. "Planning in Dynamic Environments: Extending HTNs with Nonlinear Continuous Effects." In *Proceeding of the Twenty-Fourth AAAI Conference on Artificial Intelligence* (pp. 1115–1120). July 11–15, Atlanta, Georgia, United States.

[12] Sohrabi, S., J. A. Baier, and S. A. McIlraith. 2009. "HTN Planning with Preferences." In *Proceedings of the Twenty-First International Joint Conference of Artificial Intelligence (IJCAI 2009)* (pp. 1790–1797). July 11–17, Pasadena, California, United States.

[13] Borrajo, D., A. Roubíčková, and I. Serina. 2015. "Progress in Case-Based Planning," *ACM Computing Surveys (CSUR) 47(2)*: Article No. 35.

[14] Boutilier, C., T. Dean, and S. Hanks. 1995. "Planning Under Uncertainty: Structural Assumptions and Computational Leverage." In *Proceedings of the Second European Workshop on Monte Carlo Treatment Planning (September)* (pp.157–171). October 19–21, Cardiff, United Kingdom.

[15] Kaelbling, L. P., M. L. Littman, and A. R. Cassandra. 1998. "Planning and Acting in Partially Observable Stochastic Domains." *Artificial Intelligence 101(1)*: 99–134.

[16] Meuleau, N., K. E. Kim, L. P. Kaelbling, and A. R. Cassandra. 1999. "Solving POMDPs by Searching the Space of Finite Policies." In *Proceedings of the Fifteenth conference on Uncertainty in Artificial Intelligence* (pp. 417–426). July 30–31, Stockholm, Sweden.

[17] Campbell, T., and J. P. How. 2015. "Bayesian Nonparametric Set Construction for Robust Optimization." In *Proceedings of the 2015 American Control Conference (ACC)* (pp. 4216–4221). July 1–3, Chicago, Illinois, United States. IEEE.

[18] Brown, N., and T. Sandholm. 2017. "Superhuman AI for heads-up no-limit poker: Libratus beats top professionals." *Science 359*: 418-424.

[19] Mnih, V., et al. 2015. "Human-level control through deep reinforcement learning." *Nature 518*: 529-533.

[20] Schulman, J., F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. 2017. "Proximal policy optimization algorithms." *arXiv preprint arXiv:1707.06347*.

[21] *OpenAI Five*. 25 June 2018. OpenAI. https://blog.openai.com/openai-five/. Accessed 20 August 2018.

[22] Ghallab, M., D. Nau and P. Traverso. 2004. *Automated Planning: theory and practice*. Morgan Kaufmann Publishers Inc., San Francisco, CA.

[23] Thrun, S., W. Burgard and D. Fox. 2005. *Probabilistic robotics*. MIT Press, Cambridge, Massachusetts.

[24] Kelley, L., M. Ouimet, B. Croft, E. Gustafson and L. Martinez. 2017. "An Approach to Integrating Human Knowledge into Agent-Based Planning." In *Workshop on Impedance Matching in Cognitive Partnerships* at the *2017 International Joint Conference on Artificial Intelligence (IJCAI-17)*. 21 August, Melbourne, Victoria, Australia.

[25] Behymer, K., C. Rothwell, H. Ruff, M. Patzek, G. Calhoun, M. Draper, S. Douglass, D. Kingston, and D. Lange. 2017. "Initial evaluation of the intelligent multi-uxv planner with adaptive

collaborative/control technologies (IMPACT).” Infoscitex Corp. Beavercreek, OH.

[26] Lange, D. and J. Reeder. 2017. “IMPACT machine learning efforts.” In *Proceedings of SPIE Unmanned Systems Technology XIX*. *10195*. 11 April, Anaheim, California, United States.

[27] Okabe, A., B. Boots, K. Sugihara and S. N. Chiu. 2009. *Spatial tessellations: concepts and applications of Voronoi diagrams*. Vol. 501. John Wiley & Sons, Ltd. Chichester, England.

[28] Kuhn, H.W. 1955. "The Hungarian method for the assignment problem." *Naval research logistics quarterly 2.1-2*: 83-97.

[29] Hart, S. G., and L. E. Staveland. 1988. “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.” *Advances in Psychology, 52*: 139–183.

[30] Hart, S. G. 2006. “NASA-task load index (NASA-TLX); 20 years later.” In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*: 904–908.

[31] Endsley, M. R., and D. J. Garland (Eds.). 2000. *Situation awareness analysis and measurement*. CRC Press. Boca Raton, FL.

Modeling Provenance of Decisions within the Human-Autonomy Team

Crisrael Lucero†, Dr Trung Dong Huynh‡, Dr Douglas S Lange†, Dr Luc Moreau‡
†Space and Naval Warfare Systems Center - Pacific
San Diego, California
UNITED STATES OF AMERICA

‡ King's College London
London, GBR
UNITED KINGDOM

crisrael.lucero@navy.mil, dong.huynh@kcl.ac.uk, doug.lange@navy.mil, luc.moreau@kcl.ac.uk

ABSTRACT

Dynamic mission environments create an always-changing battlespace with many actors. The use of artificial intelligence and autonomy can support the decision-making burden on human operators tasking unmanned vehicles. However, the burden of knowledge and memorization can become overwhelming to bear during complex operations. We introduce information provenance and the capabilities it provides when actors within the human-autonomy team have their actions and decisions modelled by the Provenance Data Model and recorded as provenance data. Issues that are addressed by enabling autonomous command and control systems with information provenance include situation awareness, transparency, and improved decision making aids.

1.0 INTRODUCTION

Provenance is information about entities, activities, agents, and the relationships between these concepts [1]. Provenance explains more than what happened, it also answers how data was manipulated, why, and who was involved in the process. This information should be quite important in military operations and Command and Control (C2), more so when we bring autonomous artificial agents into the decision-making process.

Provenance serves several purposes. Forensically, provenance can help us trace through decisions, understand who made them (human or autonomous agent and under what conditions). Fully recorded provenance can illuminate data dependencies, responsibility flow, and explain why certain actions were taken. During operations, these dependencies are important as well. Dependency tracking is a critical capability of C2, which can lead to explanations and reasoning as to why some decisions cannot be made or need to be updated as both the battlespace and the autonomous actors constantly change status.

As we bring autonomous agents into the C2 process, it is key to find ways of making their decisions transparent to human stakeholders. Some agents will contain decision models that are currently opaque to humans (e.g., artificial neural networks). Tracing the decisions made relative to the information available allows us to find ways of explaining their decisions even if their decision processes are not visible.

This paper describes how we've implemented a provenance recording system that tracks the actions between both humans and several autonomous systems. We also discuss how provenance capabilities can augment human-autonomy teaming and the necessity for provenance to enable explanations and transparency.

2.0 BACKGROUND AND MOTIVATION

The Provenance Data Model (PROV-DM) is a generic data model for provenance that was standardized by the World Wide Web Consortium in 2013. It allows domain and application-specific representations of provenance data to be translated and interchanged between systems [2].

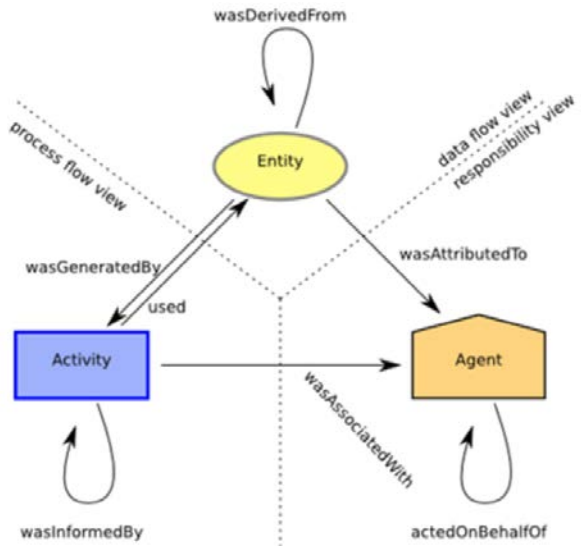


Figure 1: The main PROV concepts. Entities represented as yellow ellipses, activities are blue rectangles, and agents are orange pentagon-houses [2].

The ACM US Public Policy Council published a statement on the importance of algorithmic transparency and accountability, which discusses uncertainty in determining how biased an algorithm may be [3]. There is growing evidence that some algorithms can be opaque, making it impossible to determine whether system outputs are biased or erroneous. The usage of algorithms for automated decision making can potentially be harmful when a system begins to discriminate [4]. AI and autonomous systems have been seen as black boxes, yet these systems should be held at the same standard of ethics as human decision makers. Literature shows a plethora of use cases and methods for using provenance for accountability [5][6] and trust [7][6]. Some ethical aspects of Human-Autonomy Teaming (HAT) can be directly addressed by data provenance. Data provenance brings light on responsibility flow, how data was procured and processed, and explanations for how algorithms evolved data critical to the decision-making process.

Provenance analytics is another growing field of research that provide tools and techniques for provenance graphs.

Anomalies in a workflow execution can be detected by checking for “structural flaws” in its provenance graph, identify graphs whose characteristics are too different from the norms [8]. More recently, it has been shown that, in a number of applications, the topological characteristics of provenance graphs correlate with certain real-world properties of the data or events they describe. For example, predictive models were developed to classify the quality of data created in a crowdsourcing application [9] or to identify instructions from chat message in an alternate-reality game with high levels of accuracy [10].

3.0 PROVENANCE AND HUMAN-AUTONOMY TEAMING

When a system has been enabled with provenance tracking and modeling, different provenance analytics techniques can be used on the collected provenance data to build an autonomous decision support agent. Reference [11] claims that the following attributes represent successful systems: Interactivity, Event and Change Detection, Representation Aiding, Error Detection and Recovery, Information out of Data, and Predictive Capabilities. Additionally, [12] discusses that the following limitations for intelligent decision support systems: emphasis on technology, emphasis on modeling the world with little emphasis on modeling the user, insufficient trust, model incomprehensibility, model scope and model rigidity, and model vulnerability to adversarial attacks. The usage of analyzing and querying provenance for an autonomous agent can both supplement attributes of a successful decision support agent and support the shortcomings that limit their reliability and applicability.

In the context of emergency response, tracking the provenance of decisions was proposed to enable the capability of answering questions on the steps taken in a rescue task, why some target was not rescue, or why there was a delay, etc. [13]. In the same context, the HAC-ER disaster response system demonstrated that tracking provenance of data and decisions across human and agent teams allows for inconsistencies to be automatically detected [14]. Thereby, an autonomous agent monitoring incoming provenance from an active operation was able to alert commanders when new information conflicts with an existing plan sent out to first responders.

Provenance recording in clinical studies facilitates documenting evidences for regulatory submissions and provides auditability for decisions made by the decision support tools used during a study [15].

Our goals for an autonomous system that can query and analyse provenance graphs from systems is to incorporate strong HAT and intelligent decision support agent principles. Within the attributes that represent successful systems, provenance is capable of accomplishing several. The HAC-ER disaster response system showcases the ability to determine event and change detection by querying for and detecting anomalies [14]. Although the visual representations of provenance can be expressed as images, there are different summarization and analysis techniques to aid representation of the scenario. Provenance can also address some of the shortcomings of HAT. Because the PROV-DM is domain-agnostic and can be exchanged between different heterogeneous systems [2], the emphasis can be on both modeling the world and the user. Our specific implementation showcases how provenance is modeling both humans, autonomy, and how they interact with the scenario. Using provenance to build upon trust is an issue that has been thoroughly researched [6][7]. Different analytics techniques have been studied to approach the problem of model incomprehensibility, one such techniques treats provenance graphs as a networks and interprets activities modeled by provenance [9].

The Situation awareness-based agent transparency (SAT) model [16] has been used as a gold standard for modeling how autonomous agents should provide situation awareness and transparency to human operators in order to build trust and accountability. These additional attributes are important to HAT, and being able to convey the different levels of transparency will aid war fighters in developing trust for autonomy, reason over why certain predictions were made, and augment human-in-the loop and human-on-the-loop capabilities. The initial approaches taken to creating a provenance service that not only models provenance, but also queries the provenance generated and provides information about the actions of autonomy to the human operators.

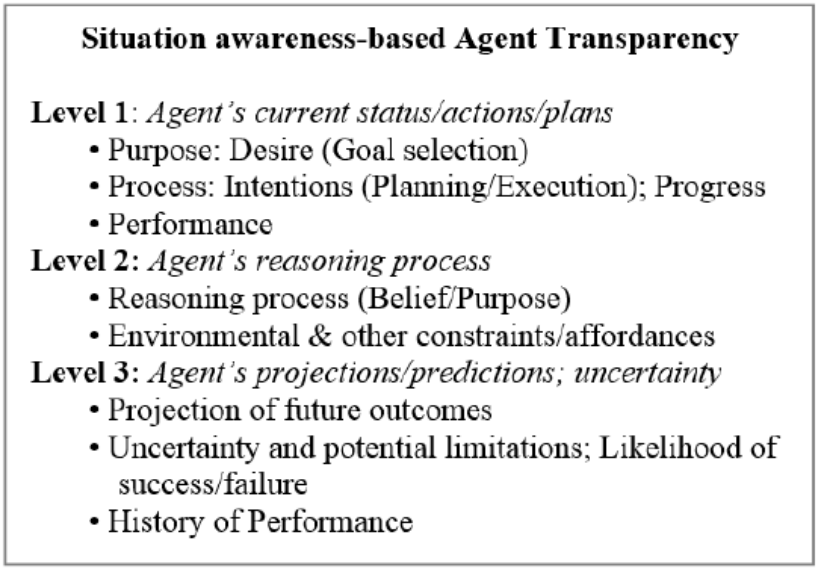


Figure 2: Situation awareness-based agent transparency (SAT) model [16].

In the following two sections, we discuss how we approach SAT Level 1 and SAT Level 2 by tracking the actions and decisions made by humans and autonomy. We discuss the approach of bringing information of the autonomous agents reasoning processes during violations and how autonomy has changed certain plans using a non-intrusive multimedia narrative.

4.0 ACTIONABLE PROVENANCE WITHIN ALLIED IMPACT

The Intelligent Multi-UxV Planner with Adaptive Collaborative Control Technologies (IMPACT) system is a prototype C2 system that practices different application and platform levels of autonomy [17]. The IMPACT project was an initiative to allow an inversion of the unmanned vehicle staffing ratio; as opposed to several operators controlling a single unmanned vehicle, IMPACT sought to grant a single operator control of several through the use of autonomy [18].

The Allied IMPACT (AIM) system is the joint international effort of several autonomous services working together within the centralized IMPACT system. In a previous study, the IMPACT system was instrumented with the PROV-DM using the Python Prov library [19]. Since then, the core technology has been switched to using the Provenance Templating System (PROV-Templates) to generate provenance [20]. This approach still treats the IMPACT system as a Gray-Box Decision Characterization system and generates expansions of those templates during runtime [21]. PROV-Templates is a declarative approach that enables designers to generate provenance compatible with the PROV standard. Different actions and interactions performed within the AIM system is modeled by individual templates that form the basic placeholder. As bindings are created during application runtime, the templates are instantiated and expanded with real scenario values. The templates are capable of merging together intelligently and provides significant benefits to software engineering [20].

The figure below showcases an example of provenance that has been tracked and modeled based on human interaction with the system and the different autonomous actors that augmented the decision making process. Going through the process flow, a chat event occurred that triggered an autonomous task management agent to discretize the goal and generate a task for the human operator to complete. Based on the different

constraints derived by the task manager, a different set of plays is made available by the planning agent to the operator to complete the task. The human operator can choose to engage these plays manually or use a different autonomous agent to carry out the task for them. After the play is selected, the planner will engage the play.

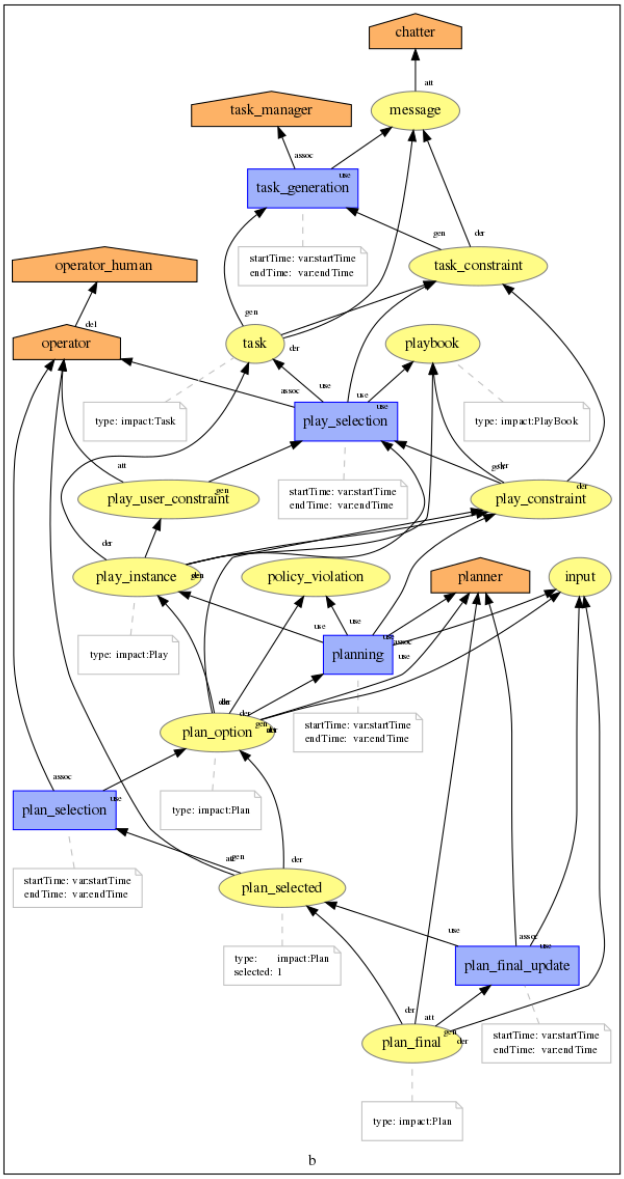


Figure 3: PROV-DM representation of a human play-call. Provenance graphs are to be read left-to-right and top-to-bottom. As provenance graphs grow, visual representations struggle to present information in an organized approach to human users.

Provenance data has also been tracked when autonomous agents interact with each other, whether the human was involved with the process or not. A policy management system autonomously provides warnings to IMPACT when air traffic control policies are violated [19]. In the figure below a policy violation occurs with an ongoing mission and the policy management system notifies IMPACT. The operator agent is not the

human, but instead either the Plan Monitor or Task Manager that selects and calls automated plays due to the urgency needed to avoid violating airspace restrictions.

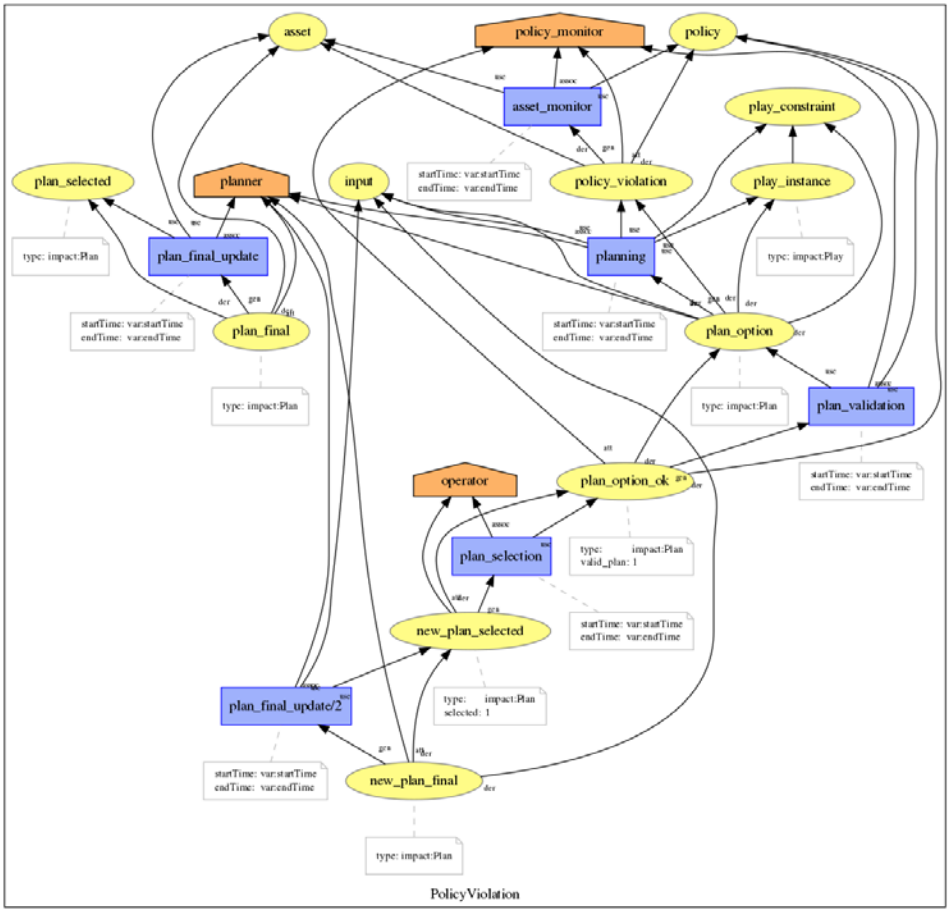


Figure 4: PROV-DM representation of a policy violation and the interactions that occur between the different autonomous systems.

A pilot, proof-of-concept agent has been created for initial testing and basic interfacing with a database of information provenance. The agent allows a human operator to query the provenance and determine which constraints and dependencies are blocking the planner from tasking autonomous vehicles with a play. The provenance agent replies to the human operator in chat a description of which vehicles are capable of accomplishing the play and why those vehicles, based on the application runtime provenance, are unable to accomplish the current task at hand. The next section discusses how we move a step up from using chat-based interactivity and how we've enhanced our agent to do more than dependency and constraint tracking.

5.0 THE NEED FOR EXPLAINABILITY AND TRANSPARENCY

Along with the US ACM statement [3] on the need for algorithmic transparency and accountability, the European Union has passed regulation restricting automated individual decision-making [22]. The regulation effectively creates a "right to explanation" for users so they can ask for explanations of algorithmic decisions that were made about them [23]. With the introduction of explanations generated by autonomous systems using provenance, human operators will not be blindly trusting autonomy. A proposed solution for

regulatory mechanisms was to increase the transparency in the data chain [24], which is what data provenance aims to do. The concept of using provenance for explainable artificial intelligence and autonomy has been explored [21], and steps forward can perfect the methodology.

Within the AIM system, the Multimedia Narrative system uses Rhetorical Structure Theory (RST) to provide automated presentations using a virtual human advisor [25]. Work has been done to convert provenance graphs into the RST graph structure to create on-the-fly, automated narratives based on actions autonomy has taken throughout the AIM scenario [26].

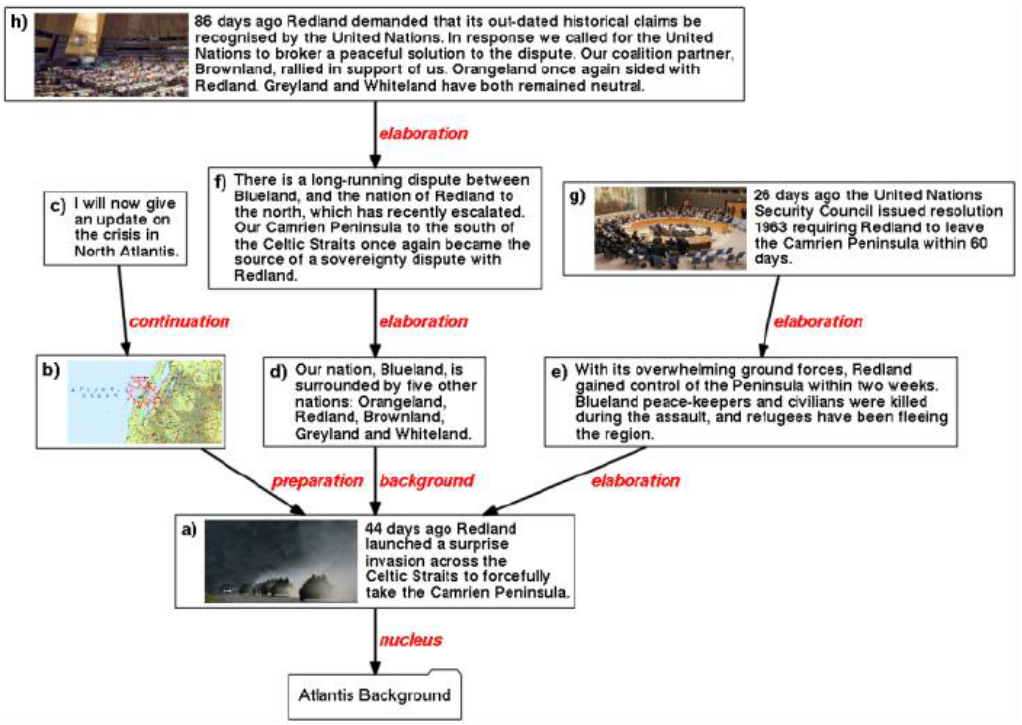


Figure 5: Rhetorical relations assigned to multimedia elements in the 'Military Strikes in Atlantis' scenario [25].

The above figure shows a detailed technical overview of the different rhetoric relations that can be assigned by multimedia elements. The six satellite relations perform a different goal: Preparation establishes narrative context, Elaboration provides more information, Background provides situational context, Conclusion provides a restatement and narrative context, Continuation represents a relation where multiple elements need to be presented together in sequence, and Initialization is a specialized case of preparation that must always be presented with the nucleus [25].

A PROV to RST conversion within the AIM system would allow for the narrative service to select a specific concept (an entity, activity, or agent) from an extracted, segmented piece of the entire application provenance. The specific concept selected acts as the basis of information and rhetoric relations are built based on the provenance relationships (e.g., elaborations and continuations can be built from wasDerivedFrom and wasAttributedTo relationships).

6.0 CONCLUSION AND FUTURE WORK

In this paper we discussed current and ongoing work with enabling and experimenting with provenance within the context of HAT in C2.

The Space and Naval Warfare Systems Center - Pacific along with King's College London and the University of Southern California have begun a research effort to take provenance of C2 information further. Utilizing data analytics and machine learning approaches, the following capabilities are envisioned:

- Learning to characterize and classify the quality of a planning evolution.
- Assigning responsibility for particular decisions to particular agents or teams to illuminate risk to decision makers.
- Summarizing complex composite decisions at appropriate levels of abstraction for informing decision makers.
- Expand upon UML2PROV [27] and attempt to automatically generate PROV-Templates with other modeling languages such as SysML.

With these future capabilities, we hope to further augment our autonomous PROV service with the principles and attributes of HAT systems.

7.0 REFERENCES

[1] Luc Moreau and Paul Groth. Provenance: An Introduction to PROV. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan and Claypool, 2013.

[2] Luc Moreau and Paolo Missier. PROV-DM: The PROV Data Model. *WC3 Recommendation,* 2013. http://www.w3.org/TR/prov-dm/.

[3] Statement on Algorithmic Transparency and Accountability. ACM US Public Policy Council, *USACM Press Releases*, no. January 12, pp. 1—2, 2017.

[4] David Danks and Alex John London. Algorithmic Bias in Autonomous Systems. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 4691—97, 2017.

[5] Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, and Gerald Jay Sussman. Information Accountability. In *Communications of the ACM*, 51, no. 6, pp. 82—87, 2008.

[6] Ryan K.L. Ko, Peter Jagadpramana, Miranda Mowbray, Siani Pearson, Markus Kirchberg, Qianhui Liang, and Bu Sung Lee. TrustCloud: A Framework for Accountability and Trust in Cloud Computing. In *IEEE World Congress on Services*, pp. 584—588, 2011.

[7] Jeremy J. Carroll, Christian Bizer, Pay Hayes, Patrick Strickler. Named Graphs, Provenance and Trust. In *Proceedings of the 14th International Conference on World Wide Web*, pp. 613—622, 2005.

[8] You-Wei Cheah and Beth Plale. Provenance Analysis: Towards Quality Provenance. In *2012 IEEE 8th International Conference on E-Science*, pp. 1—8, 2012.

[9] Trung Dong Huynh, Mark Ebden, Matteo Venanzi, Sarvapali Ramchurn, Stephen Roberts, and Luc Moreau. Interpretation of Crowdsourced Activities Using Provenance Network Analysis. In *The First*

AAAI Conference on Human Computation and Crowdsourcing, pp. 78—85, 2013.

[10] Trung Dong Huynh, Mark Ebden, Joel Fischer, Stephen Roberts, and Luc Moreau. Provenance Networks Analysis. *Data Mining and Knowledge Discovery*, 32(3). Springer US, 2018.

[11] Stephanie Guerlain, Donald E. Brown, and Christina Mastrangelo. Intelligent Decision Support Systems. *IEEE International Conference on Systems, Man, and Cybernetics*, 2000.

[12] Karel van den Bosch and Adelbert Bronkhorst. Human-AI Cooperation to Benefit Military Decision Making. In *Proceedings of the NATO IST-160 Specialist Meeting on Big Data and Artificial Intelligence for Military Decision Making*, 2018.

[13] Iman Naja, Luc Moreau, and Alex Rogers. Provenance of Decisions in Emergency Response Environments. In *Provenance and Annotation of Data and Processes*, pp. 221-30, Springer Berlin, 2010.

[14] Sarvapali D. Ramchurn, Trung Dong Huynh, Feng Wu, Yukki Ikuno, Jack Flann, Luc Moreau, Joel E. Fischer, et al. A Disaster Response System Based on Human-Agent Collectives. In *Journal of Artificial Intelligence Research*, 57, pp. 661-708, 2016.

[15] Vasa Curcin. Embeding Data Provenance into the Learning Health System to Facilitate Reproducible Research. *Learning Health Systems*, 1(2), 2017.

[16] Jessie Y.C. Chen, Shan G. Lakhmani, Kimberly Stowers, Anthony R. Selkowitz, Julia L. Write, and Michael Barnes. Situation Awareness-Based Agent Transparency and Human-Autonomy Teaming Effectiveness. In *Theoretical Issues in Ergonomics Science*, vol. 19, iss. 3, pp. 259—282, 2018.

[17] Robert S. Gutzwiller, Douglas S. Lange, John Reeder, Rob L. Morris, and Olinda Rodas. Human-Computer Collaboration in Adaptive Supervisory Control and Function Allocation. In *7th International Conference on Virtual, Augmented, and Mixed Reality*, 2015.

[18] Mary L. Cummings. Operator Interaction with Centralized versus Decentralized UAV Architectures. In *Handbook of Unmanned Aerial Vehicles*, pp. 977-992, Springer, 2015.

[19] Crisrael Lucero, Braulio Coronado, Eric Gustafson, Douglas S. Lange. Tracking Provenance in Decision Making Between the Human and Autonomy. In *15th International Conference on Engineering Psychology and Cognitive Ergonomics*, 2018.

[20] Luc Moreau, Belfrit V. Batlajery, Trung Dong Huynh, Danius Michaelides, and Heather Packer. A Templating System to Generate Provenance. *IEEE Transactions on Software Engineering*, 44(2), pp. 103—121, 2018.

[21] Crisrael Lucero, Braulio Coronado, Oliver Hui, and Douglas S. Lange. Exploring Explainable Artificial Intelligence and Autonomy through Provenance. *IJCAI-ECAI 2018 2nd Workshop on Explainable Artificial Intelligence*, 2018.

[22] Parliament and the Council of the European Union. General Data Protection Regulation. *Official Journal of the European Union*, 2016.

[23] Bryce Goodman and Seth Flaxman. European Union Regulations on Algorithmic Decision Making and the "Right to Explanation". In *AI Magazine*, vol. 38, no. 3, pp. 50—57, 2017.

[24] Eva Thelisson, Kirtan Padh, and L. Elisa Celis. Regulatory Mechanisms and Algorithms Towards Trust in AI/ML. In *IJCAI 2017 Workshop on Explainable Artificial Intelligence*, 2017.

[25] Steven Wark and Marcin Nowina-Krowicki. Generating Multimedia Narrative for Virtual Humans. In *Proceedings of the Australasian Computer Science Week Multiconference*, 2016.

[26] Marcin Nowina-Krowicki, Peter Argent, Crisrael Lucero, Douglas Lange, and Steven Wark. Now This Changes Everything: Managing Provenance using Multimedia Narrative. *IEEE IRCA Workshop on Robot Teammates Operating in Dynamic, Unstructured Environments*, 2018.

[27] Carlos Sáenz-Adán, Luc Moreau, Beatriz Pérez, Simon Miles, and Francisco J. García-Izquierdo. Automating Provenance Capture in Software Engineering with UML2PROV. In *IPAW'2018: 7th International Provenance and Annotation Workshop*, pp. 1—2, Springer-Verlag Berlin Heidelberg, 2018.

Two Types of Autonomous, Language-Endowed Learning Agents For Human Autonomy Teaming

Sergei Nirenburg and Marjorie McShane
Language-Endowed Intelligent Agents Laboratory
Rensselaer Polytechnic Institute
110 8th St. Troy, NY 12180-3590
USA

nirens@rpi.edu mcsham2@rpi.edu

ABSTRACT

This paper addresses the topics of a) human-machine interaction and communication; b) comprehension; and c) joint learning/training. The rationale behind our approach to autonomy is our belief that both true autonomy in non-trivial applications and high-quality human-machine interaction can be supported only if the machine has the capability of understanding the world in which it operates and can interact with its human collaborators in the manner that is natural to the them.

1.0 INTRODUCTION

Autonomous artificial agents will reach their full potential when human members of their teams will be able to rely on them not only to carry out their assigned tasks but also to interact with them in ways similar to human-human interaction. The agent's ability to communicate with humans using natural language is a core prerequisite for attaining this goal. Human-level language understanding and generation capabilities require much more than the ability to match input text strings with stored or computed stock responses. Humans can explain, learn, anticipate, revise and adapt behavior (including communication-oriented behavior) in response to unforeseen circumstances, particular interlocutors and changing operational needs. This means that human-level language processing is strongly interdependent with general reasoning, perception and action capabilities. In order to approach human-level communication capability, language-endowed intelligent agents (LEIAs) must, therefore, model not only language *per se*. They must also model the world of their application domain, their past experiences, the situation at hand, their team members (both human and artificial), goals and actions associated with their team roles. A very important property of LEIAs is their ability to learn by instruction in natural language.

In this paper, we first briefly outline the basics of the cognitive architecture environment, OntoAgent, that supports implementation of LEIAs. Then we describe two types of LEIAs built on the basis of OntoAgent: a) virtual patients used for training medical personnel and b) robotic members of human-robot teams engaged in the task of jointly assembling furniture.

2.0 ONTOAGENT: THE ENVIRONMENT FOR DEVELOPING LEIAS

Core capabilities of a simulated, embodied, language-endowed intelligent agent include:

- behaving in a physiologically realistic manner
- experiencing, interpreting and remembering its own mental, physical, and emotional states
- communicating with people in natural language
- learning through experience, reasoning, reading, and being told

- reasoning about its own plans and goals and those of its collaborators
- managing its memory – including forgetting and consolidating memories
- making decisions to further the plans that will fulfill its goals
- collaborating with human and intelligent-agent colleagues

In order to minimize development effort, maximize resource reuse, and avoid knowledge incompatibilities, all of these processes should be supported by an integrated knowledge substrate and a uniform knowledge representation language.

In the OntoAgent cognitive architecture, all perceptual, general cognitive, and language processing capabilities rely on the same ontological substrate, the same organization of episodic memory, and the same approach to ontologically-grounded knowledge representation. The approach to recording knowledge and computing meaning was originally developed for natural language processing pursued outside of a full agent architecture [1]. but it has required no fundamental extensions to be applied to more inclusive aspects of agent modeling.

Our computational cognitive modeling of LEIAs is a mature program of work that covers perception, reasoning and action – the typical pillars of cognitive architectures [2]. As Figure 1 shows, no matter which kinds of perceptual stimuli a LEIA receives, it must *interpret* them using its knowledge resources, which involves *translating* them from raw data streams into interpreted facts represented in its ontologically-grounded knowledge bases. Types of stimuli include language input, bodily signals generated through physiological simulations, visual stimuli, and, in principle, other stimuli.[1]

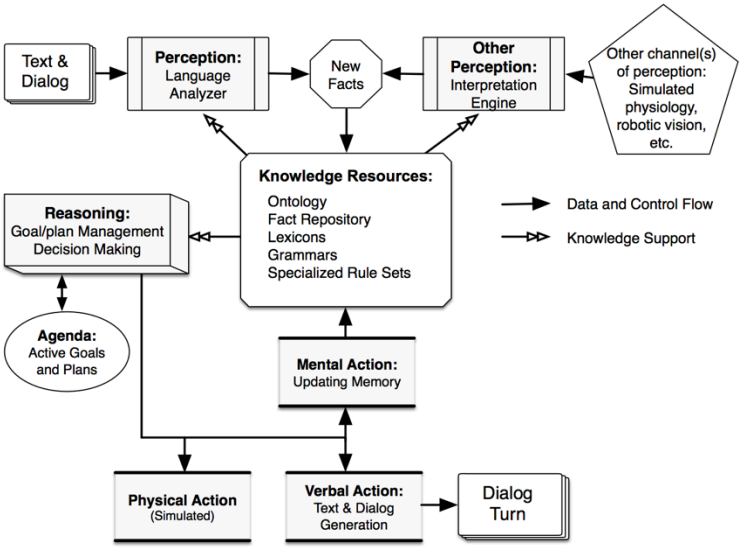


Figure 1. High-level view of OntoAgent architecture.

Since raw data are interpreted into ontologically-grounded facts, it does not matter whether a robotic LEIA knows that it must hit a given nail because its human teammate said, “Hit the nail” or because the latter said, “Hit this” while pointing at the nail – either way, the meaning representation will be the same, and the LEIA can use this new knowledge for subsequent reasoning about action.

[1] Language understanding and interoception were incorporated into LEIAs functioning as virtual patients in the Maryland Virtual Patient application [3]. Language understanding and robotic vision are currently being integrated into a robotic assistant [4].

The LEIA's knowledge representation language is a frame-based metalanguage in which concepts, or their instances, are described by properties. For example, an excerpt from the frame describing the event DRUG-DEALING in the ontology is as follows[2]:

```
DRUG-DEALING
        IS-A            VALUE           CRIMINAL-ACTIVITY
        AGENT           DEFAULT         CRIMINAL, DRUG-CARTEL
                        SEM             HUMAN
                        RELAXABLE-TO    SOCIAL-OBJECT
        THEME           DEFAULT         ILLEGAL-DRUG
        INSTRUMENT      SEM             MONEY
        …
```

When a LEIA learns about or remembers a particular *instance* of DRUG-DEALING, the concept types are replaced by numerically indexed concept instances, but the overall frame-based knowledge structures look largely the same. For example, the meaning of the language input **George Sanders sourced marijuana from a guy who lives in Phoenix** will be analyzed by a LEIA (that has specific memories in its belief repository) as follows.[3]

```
DRUG-DEALING-17
        AGENT           HUMAN-233
        THEME           MARIJUANA-21
        BENEFICIARY     HUMAN-14
        TIME            < FIND-ANCHOR-TIME    ; A FUNCTION CALL TO DETERMINE ACTUAL TIME
HUMAN-14
        GENDER          MALE
        HAS-NAME        GEORGE SANDERS
CITY-812
        HAS-NAME        PHOENIX
```

[2] For details about the OntoAgent representational substrate see [5]; for the particulars of the design and content of the OntoAgent ontology, see, e.g., [6-8].

[3] A large amount of material in this instance of text meaning representation (TMR) was omitted for presentation purposes.

HUMAN-233

HAS-GENDER	MALE
RESIDES-IN	CITY-812

While in this paper we cannot detail all the features of this representation or the process that generates it[4], we will comment on two particulars. First, when the verb *source* is used with a direct object indicating an illegal drug, the combination is compositionally understood to be an instance of the concept DRUG-DEALING. This disambiguation is supported by a combination of lexical and ontological knowledge. Second, when fillers of slots require further description, that description is provided in a co-referenced frame, meaning that there is no local reification – i.e., embedding of properties in slot fillers. E.g., HUMAN-233, who is the agent of DRUG-DEALING-17 (i.e., the seller of the drugs), has two additional known properties: HAS-GENDER MALE, and RESIDES-IN CITY-812 (i.e., the instance of the concept CITY corresponding to Phoenix in the LEIA's episodic memory), which are listed in the frame headed by HUMAN-233. This ontologically-grounded metalanguage has many advantages for agent reasoning. Most importantly, it is unambiguous and the concepts referred to are described extensively in the ontology, which means that more knowledge is available for reasoning about language and the world.

LEIAs overall – and their natural language understanding (NLU) capabilities in particular – are modeled according to principles of human-inspired cognitive modeling, e.g., [5]. The NLU modeling tenets particularly relevant to this discussion are:

1. NLU in our approach follows the theory of Ontological Semantics [1] as augmented in subsequent writings (e.g., [10-12]). Under this approach, language understanding consists of translating input language strings into unambiguous, context-sensitive, ontologically-grounded text meaning representations (TMRs) that are well-suited to automatic reasoning.
2. LEIAs analyze inputs using horizontal incrementality (processing words of input as they come in the language stream) and vertical incrementality (processing inputs to varying layers of depth, using a battery of text understanding methods).[5] Horizontal incrementality models human behavior and has practical utility as well – e.g., it is necessary for endowing LEIAs with the ability to interrupt (for clarification or correction) and to begin to act before a long utterance has been completed.
3. LEIAs operating in dialog contexts are not expected to arrive at a full and perfect interpretation of every input since even people do not do that. In real-life language use, many utterances do not make sense (e.g., thinking aloud while brainstorming), are irrelevant to the listener (e.g., a rant or a boring story at a cocktail party), or fall outside of the listener's scope of knowledge or interest (e.g., a technical discussion about nuclear physics for most readers of this text). Instead of striving for a full interpretation, LEIAs focus on achieving an *actionable* interpretation of each utterance, meaning that they are seeking grounds for performing an action in response. This action may be not only physical or verbal, but also mental, such as remembering a vague interpretation and deciding not to work harder to concretize its meaning because it is outside of the agent's scope of interest.
4. Some inputs have both more generic and more specific readings, which is due to what some refer to as the semantics vs. pragmatics distinction. LEIAs first generate the former and then, if they deem further specification worth the effort, they pursue the latter. This underscores that not only is NLU itself a reasoning-heavy enterprise, agents must also reason about how to carry out the process within the scope of their overall functioning.

[4] For details on the OntoAgent language processing module, OntoSem, see, e.g., [9].

[5] Of course, if an application is not time-sensitive, processing subsentential fragments can be skipped.

5. The LEIA's lexicon includes more kinds of knowledge to cover more linguistic phenomena (ellipsis, indirect speech acts, non-literal language, etc.) than do most lexicons. This is not only because we believe that people actually have and use such knowledge in processing language, but also because anchoring treatments of linguistic phenomena in the lexicon represents good practice in knowledge engineering as it helps to keep strict track of how declarative knowledge elements interact with processing modules.

Modeling NLU in LEIAs involves theory-based feature engineering, i.e., coming up with an inventory of properties (parameters) and their value sets to reflect aspects of context. So, the layers of context we will discuss involve different *types* of feature values accessed from different *sources* or *computed* in different ways. Note that the inventory of parameters is independently motivated by theoretical considerations – it is not developed specifically for any particular reasoning method; the reasoning methods that use the parameter inventory could include heuristic rules, logics, analogical reasoning on the basis of statistics, etc.

Since LEIAs are multifunctional intelligent agents, the features used to model them cover not just language-related material. LEIAs also routinely engage in many other aspects of cognition including, non-exhaustively, the agent's knowledge/beliefs about the world (ontology); its knowledge of remembered object, event and state instances (long-term episodic memory); the active object, event, and state instances comprising its situation model (working memory); its knowledge about language (including the lexicon and rule sets); its inventories of goals and plans; its personality traits; its cognitive biases; its physical, mental and emotional states; and its beliefs about all of these with respect to its interlocutors, which the agent must hypothesize its theory of mind, also called "mindreading" (e.g., [13]).

3.0 LEIAS AS VIRTUAL PATIENTS

Simulation-based training offer users the opportunity to gain extensive practical experience in a short time and without risk. For example, participants using the SHERLOCK II electronic troubleshooting system for F16 aircraft of the US Air Force were reported to have learned more in 20 hours of tutoring than in 4 years of field experience [14]. The Maryland Virtual Patient (MVP) system addresses the need for simulation-based training in the domain of clinical medicine, allowing medical trainees to develop clinical decision-making skills by managing a cohort of highly differentiated virtual patients in dynamic simulations, with the optional assistance of a virtual tutor.

The physician-educators who conceived of MVP set down the following requirements for the developers:

1. The system must expose students to virtual patients that demonstrate sophisticated, realistic behaviors, thus allowing the students to suspend their disbelief and interact naturally with them.
2. It must allow open-ended, trial-and-error investigation – i.e., learning through self-discovery – with the virtual patient's anatomy and physiology realistically adjusting to both expected and unexpected interventions.
3. It must offer a large population of patients suffering from each disease, with each patient displaying clinically relevant variations on the disease theme; these can involve the path and rate of disease progression, the profile and severity of symptoms, responses to treatments, and secondary diseases or disorders that affect treatment choices.
4. It must be built upon models that must:

    - be explanatory, both to provide transparency to the medical community who must endorse the system, and to provide tutorial support to students, explaining the *what* and *why* of what is happening.
    - integrate scientifically well-understood biological mechanisms with practical clinical knowledge

(population-level data, statistical evidence) to be used when causal explanations are not available.

- allow these non-explanatory "clinical bridges" to be replaced by biomechanical causal chains if they are discovered, without perturbation to the rest of the model.
- support automatic function and realism, without including *every* physiological mechanism known to medicine; in other words, create a useful application that does not impose the impossible precondition of creating a full-blown virtual human.
- cover diseases that are both chronic and acute, and both well and poorly understood by the medical community.
- allow students to have control of the clock – to advance the simulation to the next phase of patient management at will, thus simulating the doctor's choices about when a patient is to come for a follow-up visit.
- Optionally, offer tutoring support, parameterizable to suit student preferences.
- allow virtual patients to make all the kinds of decisions that real patients do, such as when to see the doctor, whether to agree to tests and interventions, and whether to comply with the treatment protocol.

The virtual patients in MVP are "double agents" in that they display both physiological and cognitive function, as shown by the high-level system architecture in Figure 2.[6] Physiologically, they undergo both normal and pathological processes in response to internal and external stimuli, and show realistic responses both to expected and to unexpected interventions. Cognitively, they experience symptoms, have lifestyle preferences, can communicate with the human user in natural language, have memories of language interactions and simulated experiences, and can make decisions based on their knowledge of the world, their physical, mental and emotional states, and their current goals and plans. An optional tutoring agent provides advice and feedback during the simulation. The "other medical personnel" include, at present, the agents that carry out tests and procedures and report their results.

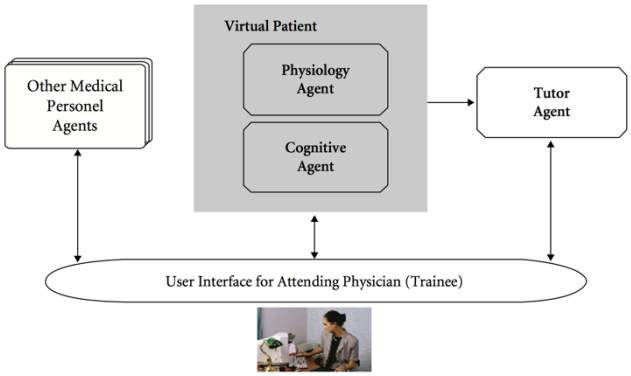


Figure 2. MVP architecture.

It is noteworthy that the MVP vision and modeling strategy anticipated the desiderata for virtual patient models detailed in the National Research Council's 2009 joint report [15], as they were developed *before* that report was published. A short excerpt from that report illustrates the overlap:

[6] In the realm of medical pedagogy, "virtual patients" have also been variously understood to refer to physical manikins, to live actors who roleplay with trainees, and to computer programs that rely on static, prefabricated decision trees.

"The clinician interacts with models and abstractions of the patient that place the raw data in context... These virtual patient models are the computational counterparts of the clinician's conceptual model of a patient... [The data] depict and simulate a theory about interactions going on in the patient and enable patient-specific parameterization... They build on submodels of biological and physiological systems..." [15]

MVP is a prototype system whose knowledge bases, software, and core theoretical and methodological foundations were developed from 2004 to 2013 (e.g., [16-23]).

The brief description of MVP below touches upon the modeling of the virtual patient's physiology and cognition and includes a sample system run, and a description of the "under the hood" traces of system functioning.

3.1 The Modeling of Physiology

Diseases are modeled as changes in the values of ontological properties over time. For each disease, a set number of conceptual stages is established, and typical values (or ranges of values) for each property are associated with each stage. Values at the start or end of each stage are recorded explicitly, with values between stages being interpolated. Disease models include a combination of fixed and variable features. For example, although the number of stages for a given disease is fixed, the duration of each stage is variable; similarly, although the values for *some* physiological properties undergo fixed changes across patients (to ensure that the disease manifests appropriately), the values for other physiological properties are variable within a specified range to allow for different instances of MVPs to differ in clinically relevant ways.

Roughly speaking, diseases fall into two classes: those for which the key causal chains are well understood and can drive the simulation, and those for which the key causal chains are not known. The models for the latter types of diseases rely on clinical observations about what happens and when, but not why. Most disease models integrate both kinds of modeling strategies in different proportions.

To develop computational cognitive models that are sufficient to support realistic patient simulations in MVP, a knowledge engineer leads physician-informants through the process of distilling their extensive and tightly coupled physiological and clinical knowledge into the most relevant subset and expressing it as causal chains with conditionals. Not infrequently, specialists are also called upon to formulate hypotheses about the unknowable, such as the preclinical (i.e., pre-symptomatic) stage of a disease and the behavior of the values of physiological properties between the times when tests are run to measure them. Such hypotheses are, by nature, imprecise. However, rather than permit this imprecision to grind agent development to a halt, we proceed in the same way as live clinicians do: by developing *a* model that is reasonable and useful, with no claims that it is the only model possible or that it precisely replicates human functioning.

The selection of parameters to be included in a disease model is guided by practical considerations. Parameters are included because (a) they can be measured by tests, (b) they can be affected by medications or treatments, and/or (c) they contribute to a physician's mental model of the disease. In addition to using parameters that directly reflect biologically attestable properties, we also include abstract parameters that foster the formulation of a compact, comprehensible model.[7] Such features reflect the creative aspect of computational modeling that is not directly extractable from print resources.

Once an approach to modeling a disease has been devised and all requisite details have been elicited from the experts, the disease-related events and their participants are encoded in ontologically-grounded scripts written in the metalanguage of the LEIA's ontology. MVP includes both domain scripts and workflow scripts. Domain scripts describe basic physiology, disease progression and responses to interventions,

[7] These features can be likened to the inclusion of intermediate categories in ontologies: although one does not typically talk about WHEELED-SURFACE-VEHICLES, this can be an appropriate node in an ontology.

whereas workflow scripts model the way an expert physician would handle a case, thus enabling automatic tutoring.

3.2 The Modeling of Cognition in Virtual Patients

Virtual patients need many cognitive capabilities. Their **language understanding** capabilities are described throughout this book. Their **language generation** involves two aspects, content and form. The content derives from reasoning and is encoded in ontologically grounded meaning representations. The form is constructed by templates, which proved sufficient for this application. Two other necessary cognitive capabilities of virtual patients are learning new words and concepts, and making decisions about action, which we consider in turn.

3.2.1 Learning New Words and Concepts

Learning is often a prerequisite to making a decision. Indeed, to give just one motivating example, no patient – real or virtual – should agree to things like medical procedures in full ignorance of their nature and risks. Table 1 shows an excerpt of a sample dialog in MVP between a virtual patient (P) and the human user/doctor (D) that illustrates the agent's learning of elements of ontology and lexicon as a prerequisite for informed decision-making.

Table 1. Learning lexicon and ontology through dialog in MVP.

Dialog	**Ontology learned**	**Lexicon learned**
D: You have achalasia.	The concept ACHALASIA is learned and made a child of DISEASE.	The noun "achalasia" is learned and mapped to the concept ACHALASIA.
P: Is it treatable? *D: Yes.*	The value for the property TREATABLE in the ontological concept ACHALASIA is set to 'yes'.	
D: I think you should have a Heller myotomy.	The concept HELLER-MYOTOMY is learned and made a child of medical-procedure. Its property treatment-option-for receives the filler heller-myotomy.	The noun "Heller myotomy" is learned and mapped to the concept HELLER-MYOTOMY.
P: What is that? *D: It is a type of esophageal surgery.*	The concept HELLER-MYOTOMY is moved in the ontology tree: it is made a child of surgical-procedure. Also, the theme of HELLER-MYOTOMY is specified as ESOPHAGUS.	
P: Are there any other options? *D: Yes, you could have a pneumatic dilation instead…*	The concept PNEUMATIC-DILATION is learned and made a child of MEDICAL-PROCEDURE.	The noun "pneumatic dilation" is learned and mapped to the concept PNEUMATIC-DILATION.

D: (cont.) which is an endoscopic procedure.	PNEUMATIC-DILATION is moved from being a child of MEDICAL-PROCEDURE to being a child of ENDOSCOPY.	
P: Does it hurt? *D: Not much.*	The value of the property PAIN-LEVEL in PNEUMATIC-DILATION is set to .2 (on a scale of {0-1}).	

When the virtual patient processes each of the doctor's utterances, it automatically creates text meaning representations that it then uses for reasoning and learning. The text meaning representation for the first sentence is ACHALASIA-1 (EXPERIENCER HUMAN-1). The patient knows to make the newly learned concept ACHALASIA a child of DISEASE in the ontology because the lexical sense it uses to process the input "You have X" asserts that X is a disease. This sense is prioritized over other transitive meanings of the verb 'have' because the discourse context is a doctor's appointment and the speaker is a doctor. A similar type of reasoning suggests that a Heller myotomy is a kind of MEDICAL-PROCEDURE. Our short dialog also shows two examples of belief revision: when the virtual patient learns more about the nature of the procedures HELLER-MYOTOMY and PNEUMATIC-DILATION, it selects more specific parents for them, thereby permitting the inheritance of more specific property values.

3.3 Making Decisions About Action

Virtual patients carry out dynamic decision-making in a style that approximates human decision-making, most notably because the agents' decisions are *explanatory*. Whenever a decision needs to be made, the virtual patient first determines whether it has sufficient information to make it – an assessment that is based on a combination of what it actually knows, what it believes to be necessary for making a good decision, and its personality traits. If it lacks some decision-making knowledge, it can posit the goal of obtaining this knowledge, which is a metacognitive behavior that leads to learning. Formally speaking, a goal is an ontological instance of a property, whose domain and range are specified. Goals can appear on the agenda in four ways:

- **Perception via interoception**. The moment the patient perceives a symptom, the symptom appears in its short-term memory. This triggers the addition of an instance of the goal BE-HEALTHY onto the agenda. We assume that achieving the highest possible value of BE-HEALTHY (1 on the abstract scale {0,1}) is a universal goal of all humans and, in cases in which it seems that a person is not fulfilling this goal, he is simply prioritizing another goal, such as experience-pleasure.
- **Perception via language**. Any user input that requires a response from the virtual patient (e.g., a direct or indirect question) puts the goal to respond to it on the agenda.
- **A precondition of an event inside a plan is unfulfilled**. For example, most patients will not agree to an intervention about which they know nothing. So, one of the events inside the plan of decision-making about an intervention is finding out whichever features of it are of interest to the individual.
- **A predetermined period of time has passed since the latest instances of the events BE-DIAGNOSED or BE-TREATED have been launched**. This models regular check-ups and scheduled follow-up visits for virtual patients.

The goal BE-HEALTHY is put on the agenda when a virtual patient begins experiencing a symptom (this is made manifest by a signal from the agent's physiology model via interoception). It remains on the agenda and is reevaluated when: (a) its intensity or frequency (depending on the symptom) reaches a certain level; (b) a new symptom arises; or (c) a certain amount of time has passed since the patient's last evaluation of its

current state of health, given that the patient has an ongoing or recurring symptom or set of symptoms: e.g., "I've had this mild symptom for too long; I should see a doctor."

When making decisions about its health care, the virtual patient incorporates the following types of features, which are used in the decision functions described below.

(a) its physiological state, as perceived via interoception and remembered in its memory – particularly the intensity and frequency of symptoms. It is important to note that neither the patient nor the virtual tutor in the MVP system are omniscient about the patient's physiological state; the *simulation system* knows this, but the intelligent agents, in accordance with the objective of making them human-like, do not.
(b) certain character traits: TRUST, SUGGESTIBILITY and COURAGE (the inventory can in principle be modified to suit a particular personality theory in use)
(c) certain physiological traits: PHYSIOLOGICAL-RESISTANCE, PAIN-THRESHOLD, and ABILITY-TO-TOLERATE-SYMPTOMS (i.e., how intense or frequent symptoms have to be before the patient feels the need to do something about them)
(d) certain properties of tests and procedures: PAIN, UNPLEASANTNESS, RISK and EFFECTIVENESS. PAIN and UNPLEASANTNESS are considered typical side effects when viewed at the population level; the patient's personal individual experience of them is described below.
(e) two time-related properties: the FOLLOW-UP-DATE, i.e., the time the doctor told the patient to come for a follow-up, and CURRENT-TIME, the time of a particular interaction.

Most of these properties are scalar attributes whose values are measured on the abstract scale {0,1}. The values for all subjective features, (a) – (c), are selected for each instance of virtual patient by the patient author. That is, at the same time as a patient author selects the physiological traits of the patient – like the pace of disease progression and the patient's response to treatments if they are administered – he selects certain traits specific to the cognitive agent as well as the amount of relevant world knowledge that the patient has in its ontology. Two decision functions, written in a simplified pseudocode, will suffice for illustration.

Decision Function 1: see-MD-or-do-nothing. This function decides when a patient goes to see the doctor, both initially and for follow-up visits.

IF FOLLOW-UP-DATE is not set

AND SYMPTOM-SEVERITY > ABILITY-TO-TOLERATE-SYMPTOMS

THEN SEE-MD *; this triggers the first visit to the doctor*

ELSE IF FOLLOW-UP-DATE is not set

AND SYMPTOM-SEVERITY < ABILITY-TO-TOLERATE-SYMPTOMS

AND the symptom has persisted > 6 months

THEN SEE-MD *; a tolerable symptom has been going on for too long*

ELSE IF there was a previous visit

AND at the time of visit symptom-severity <=.3

AND SYMPTOM-SEVERITY > .7
AND [SYMPTOM-SEVERITY – ABILITY-TO-TOLERATE-SYMPTOMS] > 0

THEN SEE-MD

ELSE IF there was a previous visit

AND at the time of visit SYMPTOM-SEVERITY ∈ □{.3 .7}

AND currently SYMPTOM-SEVERITY > .9

AND [SYMPTOM-SEVERITY – ABILITY-TO-TOLERATE-SYMPTOMS] > 0

THEN SEE-MD *; there was a big increase in symptom triggering an unplanned visit to MD*

ELSE IF there was a previous visit

AND at the time of visit SYMPTOM-SEVERITY > .7

AND currently SYMPTOM-SEVERITY > .8

THEN DO-NOTHING *; symptom severity was already high at last visit*

ELSE IF CURRENT-TIME ≥ FOLLOW-UP-VISIT.TIME

THEN SEE-MD *; go to the previously scheduled visit*

ELSE DO-NOTHING

As should be clear, given the same symptom level, patients with a lower ability to tolerate symptoms will see the doctor earlier in the disease progression than patients with a higher ability to tolerate symptoms. Of course, one could incorporate any number of other character traits and lifestyle factors into this function, such as the patient's eagerness to be fussed over by doctors, the patient's availability to see a doctor around its work schedule, and so on. But even this inventory allows for considerable variability across patients (which is an objective of the entire training set-up).

Evaluation Function 2: EVALUATE-INTERVENTION. Among the decisions a patient must make is whether or not to agree to a test or treatment procedure suggested by the doctor, since many interventions carry some degree of pain, risks, side-effects or general unpleasantness. Some patients are endowed with such high levels of trust, suggestibility and courage in their personality profile that they will agree to anything the doctor says without question. All other patients must decide if they have sufficient information about the intervention to make a decision and, once they have enough information, they must decide whether they want to (a) accept the doctor's advice, (b) ask about other options, or (c) reject the doctor's advice. A simplified version of the algorithm for making this decision (which suffices for our purposes) can be stated informally as follows:

IF a function of the patient's trust, suggestibility and courage is above a threshold

OR the risk associated with the intervention is below a threshold (e.g., in the case, for a blood test)

THEN the agent agrees to intervention right away.

ELSE [*] IF the patient feels it knows enough about the risks, side-effects and unpleasantness of the

intervention (as a result of evaluating the function DETERMINE-IF-ENOUGH-INFO-TO-EVALUATE)

AND a call to the function EVALUATE-INTERVENTION establishes that the above risks are acceptable

THEN the patient agrees to the intervention.

ELSE IF the patient feels it knows enough about the risks, side-effects and unpleasantness of the intervention

AND a call to the function evaluate-intervention establishes that the above risks are not acceptable

THEN IF there are other options
THEN the physician proposes them and control is switched to [*].
ELSE the patient refuses the intervention.

ELSE IF the patient does not feel it knows enough about the intervention (as a result of evaluating the function DETERMINE-IF-ENOUGH-INFO-TO-EVALUATE)

THEN the patient asks for information about the specific properties that interest it, based on its personality traits: e.g., a cowardly patient will ask about risks, side-effects and unpleasantness,
whereas a brave but sickly person might only ask about side-effects.

IF a call to the function evaluate-intervention establishes that the above risks are acceptable
THEN the patient agrees to the intervention
ELSE the patient asks about availability of other options

IF other options are available
THEN the physician proposes them and control is switched to [*].

ELSE the patient refuses the intervention.

This decision function makes use of two other decision functions which we do not detail here, EVALUATE-INTERVENTION and DETERMINE-IF-ENOUGH-INFO-TO-EVALUATE (see Nirenburg et al. 2008 for details), as the point was a) to show the kinds of decisions virtual patients make, b) their approach to knowledge-based decision-making, and c) the kinds of dialog that must be supported to simulate the necessary interactions.

3.4 An Example System Run

To illustrate system operation, we present a sample user interaction with a virtual patient named Michael Wu. "Sample" is the key word here, as there are several substantively different – and countless trivially different – paths that this simulation could take based on what the user chooses to do during the interaction session. For example, users could choose different points in the dialog to offer clinically appropriate (or inappropriate!) interventions, or do nothing at all; they could ask the patient to come back for frequent or infrequent follow-ups; they could order appropriate or inappropriate tests; and they could have the tutor set to intervene frequently or only in cases of imminent mistakes – or not at all. Since Mr. Wu is a particular *instance* of a virtual patient, he has an inventory of property values that define him, This put some constraints on the available outcomes of the simulation. At the start of the process, Mr. Wu's physiological, pathological, psychological, and cognitive profiles are specified as follows:[8]

- **Psychological Traits**: trust [.2], suggestibility [.3], courage [.4]
- **Physiological Traits**: physiological resistance [.9], pain threshold [.2], ability to tolerate symptoms [.4]
- **Patient's knowledge of medicine**: minimal
- **Disease(s) explicitly authored for this patient:**[9] Achalasia
- **Response to treatments if they are launched**: BoTox [effective, wearing off over 12 mos.], Pneumatic dilation [effective with regression], Heller myotomy [effective permanently]

The system user, who plays the role of attending physician, does not have direct access to any of this information and must learn everything about Mr. Wu through dialog, tests, and procedures. When the simulation clock is launched, the user must wait for Mr. Wu to present at the doctor's office. Mr. Wu makes this decision using the decision function of Section 3.3. The points below describe what happens in this simulation run.

1. Mr. Wu presents with the chief complaint "difficulty swallowing". This is Day 361 of the progression of his disease, which includes the preclinical stage and a portion of the first symptomatic stage. The user will not know this temporal information. Mr. Wu has had symptoms for some time but until now the evaluation function SEE-DOCTOR-OR-DO-NOTHING has returned the answer "do nothing."

2. When the virtual patient presents at the office, this initiates the first dialog with the user. The user types in unconstrained English text (note the extensive use of difficult to automatically analyze elliptical expressions). The virtual patient analyzes it into TMRs, makes a decision about how to answer, and generates a response. The interaction begins as follows:

User:	So, you have difficulty swallowing?
Mr. Wu:	Yes.
User:	Do you have difficulty swallowing solid food?
Mr. Wu:	Yes.
User:	Liquids?
Mr. Wu:	No.

[8] Both manual and automatic methods of creating virtual patient instances have been implemented in MVP, see (REFs).

[9] Diseases can also be automatically triggered if their preconditions are met.

User:	Do you have chest pain?
Mr. Wu:	Yes, but it's mild.
User:	Any heartburn?
Mr. Wu:	No.
User:	Do you ever regurgitate your food?
Mr. Wu:	No.
User:	How often do you have difficulty swallowing?
Mr. Wu:	Less than once a week.
User:	It is too early to take any action. Please come back in 9 months.
Mr. Wu:	OK.

As we see, the user decides to do nothing – an important kind of decision in clinical medicine, and one that is difficult to teach since a physician's natural response to a patient asking for help is to do *something*.

3. After 9 months (on Day 661 of the disease progression) Mr. Wu comes back for his follow-up. He has regularly been running the evaluation function SEE-DOCTOR-OR-DO-NOTHING (since he is still symptomatic), but it always returned "do nothing" – i.e., do not schedule a new appointment prior to the scheduled follow-up. The user again asks about his difficulty swallowing, chest pain and regurgitation, using paraphrases of the original formulations (for variety and, in system demonstrations, to show that this is handled well by the NLU component). Mr. Wu responds that he has moderate chest pain, experiences regurgitation a few times a week, has difficulty swallowing solids daily and difficulty swallowing liquids occasionally. Note that the progression of difficulty swallowing from solids to liquids is a key diagnostic point that the user should catch: this suggests a motility disorder rather than an obstructive disorder.

4. The user posits the hypothesis that Mr. Wu has a motility disorder and advises Mr. Wu to have a test called EGD (esophagogastroduodenoscopy). Mr. Wu evaluates whether he will accept this advice using the function EVALUATE-INTERVENTION, described in Section 3.3. Since he is concerned about the risks, he asks about them. When the user assures him that they are extremely minimal, he agrees to the procedure.

5. A specialist lab technician agent virtually runs the test and delivers the results. This involves querying the physiological model underlying the simulation at the given point in time. The specialist agent returns the results with the interpretation: "Narrowing of LES with a pop upon entering the stomach. No tumor in the distal esophagus. Normal esophageal mucosa." These results include both positive results and pertinent negatives.

6. The user reviews the test results, decides that it is still too early to intervene, and schedules Mr. Wu for another follow-up in 4 months.

7. When Mr. Wu presents in 4 months, he reports that the symptom that has changed the most is regurgitation, which he now experiences every day. Note that throughout the simulation the patient chart

is automatically populated with responses to questions, results of tests, etc., so the user can compare the patient's current state with previous states at a glance.

8. The user suggests having another EGD and Mr. Wu agrees immediately, not bothering to launch the evaluation function for EGD again since he agreed to it the last time.

9. Then the user suggests having two more tests: a barium swallow and esophageal manometry. Mr. Wu asks about their risks (that remains his only concern about medical testing), is satisfied that they are sufficiently low, and agrees to the procedures. The former returns "Narrowing of the lower esophageal sphincter with a bird's beak" and the latter returns "Incomplete relaxation of the LES, hypertensive LES, LES pressure: 53".

10. The user decides that these test results are sufficient to make the diagnosis of achalasia. The user records this diagnosis in Mr. Wu's chart.

11. The user suggests that Mr. Wu have a Heller myotomy. Mr. Wu asks about the risks and pain involved. The user responds that both are minimal. Mr. Wu agrees to have the procedure. The user tells him to come back for a follow-up a month after the procedure.

12. Mr. Wu has the procedure.

13. Mr. Wu returns in a month, the user asks questions about symptoms and there are none. The user tells Mr. Wu to return if any symptoms arise.

4.0 INTEGRATING LEIAS WITH ROBOTS

The main purpose of this section is to illustrate how an OntoAgent-based LEIA integrated within a robotic system can learn in a dialog with its human team member. Two characteristics set this system apart from the few recent systems that actually address aspects of extracting and representing linguistic meaning [5, 13]. First, it simultaneously addresses the challenges of a) learning-oriented language-based human-robotic interaction, b) symbol grounding, c) linguistic meaning extraction, and d) the enhancement and management of the episodic, semantic and procedural memory of a robot/agent. Second, the language processing component of the system and its associated knowledge resources address a broader set of meaning-related language phenomena, described at a finer grain-size of analysis than that used in other approaches. To implement language-based learning in a social robotics environment, we must address the co-dependence among three capabilities: language understanding, learning, and task-oriented physical, mental and verbal action. Language understanding and action require knowledge, while learning, achieved through language understanding and reasoning, automatically adds to that knowledge.

The system we describe is a social robot collaborating with a human user. The experimental domain is furniture assembly (e.g., [24]), widely accepted as useful for demonstrating human-robot collaboration on a joint activity. Roncone et al. [25] report on a Baxter robot supplied with high-level specifications, represented in the HTN formalism [26], of basic actions implementing chair-building tasks. Using a keyboard command or pressing a button, the user could trigger the execution of basic actions by triggering the operation of low-level task planners that the robot could directly execute. The robot could not reason about its actions, which were stored in its procedural memory as uninterpreted skills. The system described

here integrates the robotic architecture of [25] with the OntoAgent cognitive architecture [e.g., 5]. The integrated system allows the robot to a) learn the semantics of initially uninterpreted basic actions; b) learn the semantics of operations performed by the robot's human collaborator when they are described in natural language; c) learn, name and reason about meaningful groupings and sequences of actions and organize them hierarchically; and d) integrate the results of learning with knowledge stored in its semantic and episodic memory and establish their connections with the robot's procedural memory.

The core prerequisite for human-like learning is the ability to automatically extract, represent and use the meaning of natural language texts – utterances, dialog turns, etc. This task is notoriously difficult: to approach human-level capabilities, intelligent agents must account for both propositional and discourse meaning; interpret both literal and non-literal (e.g., metaphorical) meaning; resolve references; interpret implicatures; and, particularly in informal genres, deal with stops and starts, spurious repetitions, production errors, noisy communication channels and liberal (if unacknowledged) use of the least effort principle (e.g., [27]) by speakers and hearers. The language understanding module of OntoAgent, OntoSem [9], uses an ontology (semantic memory) of some 9,000 concepts with an average of 16 properties each; an episodic memory of remembered instances of ontological concepts; a semantic lexicon for English covering about 25,000 lexical senses; and a frame-oriented formalism suitable for representing the semantics of robotic actions, natural language utterances and results of the robot's processing of other perceptual modalities (e.g., interoception, [5]).

The Process. At the beginning of the learning process, the robot can a) visually *recognize* parts of the future chair (e.g., the seat and the tool (screwdriver) to be used, b) *generate* meaning representations (MRs) of user utterances and physical actions and c) *perform* basic programmed actions: GET(OBJECT) from storage area to workspace, HOLD(OBJECT) and RELEASE(OBJECT). The user teaches the robot three types of things (implemented in three learning modules of Figure 3): a) *concept grounding*: the connection between basic actions and MRs of utterances that describe them; b) *legal sequences* of basic actions forming complex actions; and c) *augmenting the robot's memory* with descriptions of newly learned complex actions and objects.

Suppose the user issues the call for the robot to execute get(screwdriver) *and* the utterance *Now you will fetch a screwdriver*. The physical-conceptual grounding learning module (LM) will link this procedure call with the representation of CHANGE-LOCATION-1 in the MR that OntoSem produces for the example utterance (Figure 4), thus linking the robotic and the cognitive architectures. This is done by adding the property PM-LINK with the filler GET(SCREWDRIVER) to the ontological concept instance CHANGE-LOCATION-1. The immediate purpose of this linking of the robotic and the cognitive architectures is to make the robot capable in its subsequent functioning to a) trigger basic actions autonomously on the basis of language input alone and b) learn complex event sequences by just being told, without having to actually perform the actions comprising the complex event.

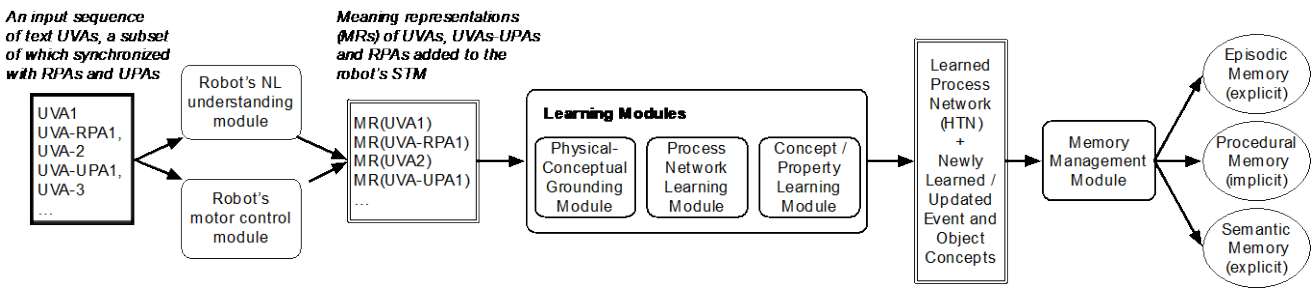


Figure 3. The core learning process. Input is a sequence of user verbal actions (UVAs) which explain user physical actions (UPAs) and issue commands to the robot, thus verbalizing robot's physical actions (RPAs), which facilitates grounding the former in the latter. UVAs are interpreted into uniform meaning representation and provide input to grounding, process

network and concept/property learning modules (LMs). The memory management module (MMM) incorporates the results of learning into the episodic and semantic memories of the robot and mutually grounds RPAs in the robot's procedural memory and corresponding concepts in its semantic memory.

SPEECH-ACT-1	
type	command
scope	CHANGE-LOCATION-1
producer	*speaker*
consumer	ROBOT-0
time	time-0 ; time of speech
CHANGE-LOCATION-1	
agent	ROBOT-0
theme	SCREWDRIVER-1
effect	BESIDE (AGENT.LOCATION THEME.LOCATION)
time	> time-0
token	*fetch*
from-sense	move-v2
HUMAN-1	
agent-of	CHANGE-LOCATION-1
token	*you*
from-sense	you-n1
SCREWDRIVER-1	
theme-of	CHANGE-LOCATION-1
token	*screwdriver*
from-sense	screwdriver-n1

Figure 4. Meaning representation for the utterance *Now you will fetch a screwdriver* (simplified).

The robot can learn legal sequences of basic actions by understanding the user's utterances in their context. The robot organizes action sequences hierarchically and makes sure that any non-terminal nodes in the resulting process network represent meaningful complex actions. If the robot does not have specifications for these complex actions in its stored knowledge, it *learns new concepts for them*, on the basis of the MRs obtained by processing the relevant user utterances. We treat joint tasks as complex tasks and require the system to decompose them into subtasks carried out by each of the team members. Basic individual tasks include Robot Physical Actions (RPAs), User Physical Actions (UPAs) and User Verbal Actions, UVA. The RPAs and UPAs appear as terminal nodes in the process network being learned. The robot's activity that includes all the kinds of learning it does as well as updating its memory structures comes under the rubric of Robot Mental Action, RMA. Due to space constraints we cannot illustrate a complete process of assembling a chair (even the shortest version of the process numbers over 150 steps). So, we present a small subset of this process – assembling the third of the four legs of the chair – accompanied by associated robotic learning, as illustrated in Table 2. All UVAs are first analyzed and their meanings are represented as MRs. UVA1 signals the beginning of the subsequence and, together with UVA7, marks the boundaries of the complex action. All the RPAs and the UPA occurring within this span, in the order of their occurrence, will form the set of the terminal nodes in the subset of the overall process network, becoming children of the non-terminal designating the complex action of building the right back leg. Once this (sub)hierarchy is constructed, the non-terminal node at its root must be named. As the robot assembles the back leg for the first time, it learns the composition of this complex action (RMA1) and labels the parent node of this small subhierarchy with the name of the concept ASSEMBLE-RIGHT-BACK-LEG. It also learns the new object-type concept RIGHT-BACK-LEG, whose existence is the effect of the above action (RMA2) and updates the concept chair by adding right-back-leg as a filler of that concept's HAS-OBJECT-AS-PART property (RMA3). The newly learned concepts are illustrated in Figure 5. The results of the operation of the process network LM are recorded in the HAS-EVENT-AS-PART property of a result of the operation of the concept LM. At this stage in the process, the fillers of some of the properties in the concepts are tentative and are expected to be modified/tightened at the memory management stage.

Memory management. Knowledge learned by the robot during each session with a human trainer (such as the sequence in Table 2) must be remembered so they can be used in subsequent functioning. Mutual grounding of basic actions and corresponding ontological events is recorded both in the robot's procedural memory (by augmenting the procedures implementing the robot's basic motor actions with links to their corresponding concepts in semantic memory) and in its semantic memory (by adding PM-LINKS, see above).

Table 2. Assembling the right back leg

UVA1	We will now build the right back leg
UVA2	Get another foot bracket
RPA1	GET(bracket-foot)
RPA2	RELEASE(bracket-foot)
UVA3	Get the right back bracket
RPA3	GET(bracket-back-right)
RPA4	RELEASE(bracket-back-right)
UVA4	Get and hold another dowel
RPA5	GET(dowel)
RPA6	HOLD(dowel)
UVA5	I am mounting the third set of brackets on a dowel
UPA1	The user affixes the foot and the right back brackets to the dowel
UVA6	Finished. Release the dowel
RPA7	RELEASE(dowel)
UVA7	OK, done assembling right back leg
RMA1	Learns action subsequence ASSEMBLE-RIGHT-BACK-LEG
RMA2	learns the object RIGHT-BACK-LEG with BRACKET-FOOT, BRACKET-BACK-RIGHT and DOWEL as fillers of HAS-OBJECT-AS-PART slot of RIGHT-BACK-LEG
RMA3	Adds RIGHT-BACK-LEG as a filler of HAS-OBJECT-AS-PART of CHAIR

Newly learned process sequences and objects (such as ASSEMBLE-RIGHT-BACK-LEG and RIGHT-BACK-LEG of Figure 5) must be incorporated in the robot's long-term semantic and episodic memories. Due to space constraints, in what follows we give an informal description of the process.

For each newly learned concept, the memory management module (MMM) first determines whether this concept should be a) added to the robot's semantic memory or b) merged with an existing concept. To make this choice, the MMM uses an extension of the concept matching algorithm of [28-29]. This algorithm is based on unification, with the added facility for naming concepts and determining their best position in the hierarchy of the ontological world model in the robot's long-term semantic memory.

ASSEMBLE-RIGHT-BACK-LEG

IS-A	PHYSICAL-EVENT
AGENT	HUMAN, ROBOT
THEME	RIGHT-BACK-LEG
INSTRUMENT	SCREWDRIVER
HAS-EVENT-AS-PART	GET(ROBOT, BRACKET-FOOT)
	RELEASE(USER, BRACKET-FOOT)
	GET(ROBOT, BRACKET-BACK-RIGHT)
	RELEASE(USER, BRACKET-BACK-RIGHT)
	GET(ROBOT, DOWEL)
	HOLD(ROBOT, DOWEL)
	MOUNT(USER, {BRACKET-FOOT, BRACKET-BACK-RIGHT}, DOWEL)
PART-OF-EVENT	ASSEMBLE-CHAIR
EFFECT	RIGHT-BACK-LEG //default effects are events; if filler of effect is an object, this means the effect is its existence

RIGHT-BACK-LEG

IS-A	CHAIR-PART
HAS-OBJECT-AS-PART	BRACKET-FOOT, BRACKET-BACK-RIGHT, DOWEL
PART-OF-OBJECT	CHAIR

Figure 5. Concepts learned as a result of processing the sequence in Table 2.

As an illustration, suppose, the concept matching algorithm has determined that the newly learned concept ASSEMBLE-RIGHT-BACK-LEG must be added to the robot's semantic memory. In this case, the algorithm also suggests the most appropriate position for the concept in the ontological hierarchy. This is determined by

comparing a) the inventory and b) sets of fillers of the properties defined for the new concepts and for the potential parents of the new concepts in the ontological hierarchy. In the example illustrated in Figure 5, the LM used the safest, though the least informative, filler (PHYSICAL-EVENT) for the IS-A property of ASSEMBLE-RIGHT-BACK-LEG. To determine the appropriate parent, the algorithm traverses the ontological hierarchy from PHYSICAL-EVENT down until it finds the closest match that does not violate recorded constraints (see REF for a description of the possible eventualities that this process can encounter).

```
RIGHT-BACK-LEG
    IS-A                    CHAIR-PART
    HAS-OBJECT-AS-PART      BRACKET-FOOT,
                            DOWEL
    PART-OF-OBJECT          CHAIR
```

FIGURE 6. A possible existing ontological concept RIGHT-BACK-LEG

```
RIGHT-BACK-LEG
    IS-A                    SOFA-PART
    HAS-OBJECT-AS-PART      BRACKET-FOOT,
                            DOWEL
    PART-OF-OBJECT          SOFA
```

FIGURE 7. Another possible existing ontological concept RIGHT-BACK-LEG

Suppose now that a concept RIGHT-BACK-LEG already exists in the ontology. If this concept is described as in Figure 6, then after comparing this concept with the newly learned concept (see Table 2), the MMM will, instead of adding a new (possibly renamed) concept to the robot's semantic memory, just add an *optional* filler BRACKET-BACK-RIGHT to the HAS-OBJECT-AS-PART property of the existing concept of Figure 6, thus merging the existing and the newly learned knowledge. If, however, the existing concept is as illustrated in Figure 7, then, because of the mismatch of the fillers of PART-OF properties between the newly learned and the standing concept, the MMM will yield two new concepts, as illustrated in Figure 8, and create a new common parent concept RIGHT-BACK-LEG for them.

```
RIGHT-BACK-LEG-CHAIR
  IS-A                  CHAIR-PART
  HAS-OBJECT-AS-PART    BRACKET-FOOT,
                        DOWEL,
                        BRACKET-BACK-RIGHT
  PART-OF-OBJECT        CHAIR

RIGHT-BACK-LEG-SOFA
  IS-A                  SOFA-PART
  HAS-OBJECT-AS-PART    BRACKET-FOOT,
                        DOWEL
  PART-OF-OBJECT        SOFA
```

FIGURE 8. Two new concepts RIGHT-BACK-LEG

Note the need of modifying the names of the concepts. An important case of merging several versions of a concept in one representation is the system's ability to represent the content of an action's HAS-EVENT-AS-PART property as an HTN, augmented with the means of expressing temporal ordering, optionality and valid alternative action sequences. Semantic memory stores the robot's knowledge of concept types. So, for example, it will contain a description of what the robot knows about chairs and chair legs. This knowledge will be used to feed the reasoning rules the robot will use while processing language, learning and making decisions. To make the robot more human-like, we also support reasoning by analogy. For this purpose, the MMM records sequences of RPAs, UPAs and UVAs that the robot represents and carries out during specific sessions of interacting with specific users in the robot's long-term episodic memory. The contents of the episodic memory will also support the robot's ability to "mind-read" its various users [e.g., [13, 28] and, as a result, to be able to anticipate their needs at various points during joint task execution as well as interpret their UVAs with higher confidence.

The system briefly described here concentrates on robotic learning through language understanding. This learning results in extensions to and modifications of the three kinds of robotic memory – the explicit semantic and episodic memory and the implicit (skill-oriented) procedural memory. The expected practical

impact of the ability to learn and reason will include the robot's ability to a) perform complex actions without the user having to spell out a complete sequence of basic and complex actions; b) reason about task allocation between itself and the human user; and c) test and verify its knowledge through dialog with the user, avoiding the need for large numbers of training examples required by learning by demonstration only. The inability of the state-of-the-art deep learning-based systems to provide human-level explanations is a well-known constraint on the utility of such systems. The cognitive robots we develop will still be capable of sophisticated reasoning by analogy but will be also capable of explaining their decisions and actions. Finally, our approach to learning does not depend on the availability of "big data" training materials. Instead, we model the way people learn since early childhood and throughout their lives – by being taught using natural language.

An important contribution of our work is overt modeling of the robot's memory components. These components include an implicit memory of skills and explicit memories of concepts (objects, events and their properties) and of instances of sequences of events (episodes, represented in our system as hierarchical transition networks, HTNs). The link established between the implicit and explicit layers of memory allows the robot to reason about its own actions. Scheutz et al. [30] discuss methodological options for integrating robotic and cognitive architectures and propose three "generic high-level interfaces" between them – the perceptual interface, the goal interface and the action interface. In our work, the basic interaction between the implicit robotic operation and explicit cognitive operation is supported by interactions among the three components of the memory system of the robot.

The first enhancement of the current learning system will consist in demonstrating how, after RPAs are mutually grounded in ontological concepts, the robot will be able to carry out commands or learn new action sequences by acting on UVAs, without any need for direct triggering through software function calls or hardware operations. Next, we intend to add text generation capabilities, both to allow the robot a more active role in the learning process (by asking questions) and to enrich interaction during joint task performance with a human user. Another novel direction of work will involve adapting to particular users – modeling robots' individuality and related phenomenological ("first-person" view) aspects of its internal organization and memory, developing and making use of mindreading capabilities that will in turn facilitate experimentation in collaboration among agents with different "theories of minds of others," similarly to the capability present in the MVP implementation of OntoAgent-based LEIAs.

5.0 FINAL WORDS

We have successfully used the infrastructure of OntoAgent in two very different applications of language-endowed intelligent agents. The MVP involves a disembodied virtual patient agent capable of interacting with the human user in a manner approaching the human level. The virtual patients are also supplied with a simulated physiology and pathology, formulated using the standard representation substrate of OntoAgent. This demonstrates that OntoAgent can successfully address issues beyond language processing, specifically, integrate different modes of perception. In the case of MVP, language inputs are integrated with inputs from interoception, made possible by the presence of the physiology model.

The knowledge structures in OntoAgent facilitate not only perception but also reasoning and decision-making, implemented mostly in the framework of goal- and plan-oriented functioning but also attempting to account for reactive actions. It is important to note that our environment can accept a variety of realizations of goal- and plan-based decision-making, which means that our general tendency is to be task-oriented rather than method-oriented. Thus, for MVP we developed both the classical rule-based decision procedures (such as those illustrated in Section 3.3 above) and decision procedures based on Bayesian networks.

In the process of integrating LEIAs with robots, we have demonstrated a way to integrate cognitive and robotic architectures and thus add physical (motor) actions to the repertoire of LEIAs that are already

capable of verbal and mental (remembering, learning, making decisions) actions.

The knowledge substrate that facilitates LEIAs is not trivial. However, we have demonstrated that once a core knowledge bases (ontology, lexicon, belief repository structures, goal and plan inventories, decision functions, etc.) have already been developed, adaptation to new domains can proceed with moderate resource allocation.

The overall approach we have briefly described here follows the premise that the only way to attain human-level functioning for autonomous entities operating in teams with people is to have these entities model both **what** people do and **how** they understand, reason, make decisions and operate. The vast majority of current AI follows the former goal. To fully succeed, the field must return to seriously taking into account the latter.

REFERENCES

[1] Nirenburg, S. and V. Raskin. 2004. Ontological Semantics. MIT Press.

[2] P. Langley, J. E. Laird, S. Rogers, Cognitive architectures: Research issues and challenges, Cognitive Systems Research 10 (2) (2009) 141–160.

[3] Nirenburg, S., McShane, M., & Beale, S. (2008). A simulated physiological/cognitive "double agent". In J. Beal, P. Bello, N. Cassimatis, M. Coen, & P. Winston (Eds.), *Papers from the AAAI Fall Symposium "Naturally Inspired Cognitive Architectures"*. AAAI Technical Report FS-08-06. Menlo Park, CA: AAAI Press.

[4] Nirenburg, S., M. McShane, S. Beale, P. Wood, B. Scassellati, O. Mangin and A. Roncone. 2018. Toward Human-Like Robot Learning. Proceedings of NLDB 2018, Paris, June.

[5] McShane, Marjorie and Sergei Nirenburg. 2012. A knowledge representation language for natural language processing, simulation and reasoning. *Intl. Journal of Semantic Computing* 6(1).

[6] Nirenburg, S., Raskin, V. and Onyshkevych, B. 1995. Apologiae Ontologiae. Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation. Leuven, Belgium. July.

[7] Nirenburg, S., M. McShane and S. Beale. 2009. A Unified Ontological-Semantic Substrate for Physiological Simulation and Cognitive Modeling. Proceedings of the International Conference on Biomedical Ontologies. Buffalo, July.

[8] McShane, M. and S. Nirenburg. 2013. Use of ontology, lexicon and fact repository for reference resolution in Ontological Semantics. In Oltramari, A., Vossen, P., Qin, L., Hovy, E. (Eds.), *New Trends of Research in Ontologies and Lexical Resources*. Springer. New Trends of Research in Ontologies and Lexical Resources, Theory and Applications of Natural Language Processing; pp 157-185.

[9] McShane, M., Nirenburg, S. and Beale, S. 2016. Language understanding with Ontological Semantics. *Advances in Cognitive Systems* 4: 35-55.

[10] McShane, M., Beale, S., and Babkin, P. 2014. **Nominal compound interpretation by intelligent agents**. Linguistic Issues in Language Technology (LiLT), 10(1): 1-34

[11] McShane, M., Nirenburg, S. and Beale, S. 2015. **The Ontological Semantic treatment of multiword expressions. Lingvisticæ Investigationes**, 38(1): 73-110. John Benjamins Publishing Company.

[12] McShane, M. and Babkin, P. 2016. **Detection and resolution of verb phrase ellipsis**. Linguistic Issues in Language Technology (LiLT).

[13] McShane, M. 2014. **Parameterizing mental model ascription across intelligent agents**. Interaction Studies, 15(3): 404-425.

[14] Evens, M., and J. Michael. 2006. *One-on-One Tutoring by Humans and Computers*. New Jersey and London: Lawrence Erlbaum and Associates, Publishers.

[15] Stead, W.W. and H.S. Lin (eds.) 2009. Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions. Committee on Engaging the Computer Science Research Community in Health Care Informatics; National Research Council.

[16] McShane, M., Fantry, G., Beale, S., Nirenburg, S. and Jarrell, B. 2007. **Disease interaction in cognitive simulations for medical training**. In Proceedings of the MODSIM World Conference &

Exposition, Virginia Beach
[17] McShane, M., Nirenburg, S., Beale, S., Jarrell, B. and Fantry, G. 2007. **Knowledge-based modeling and simulation of diseases with highly differentiated clinical manifestations**. In Bellazzi, R., Abu-Hanna, A., Hunter, J. (Eds.), Artificial Intelligence in Medicine: Proceedings of the 11th Conference on Artificial Intelligence in Medicine, AIME 2007. Volume 4594 of the series Lecture Notes in Computer Science, pp. 34-43. Berlin, Heidelberg: Springer-Verlag.
[18] McShane, M., Jarrell, B., Fantry, G., Nirenburg, S., Beale, S. and Johnson, B. 2008. **Revealing the conceptual substrate of biomedical cognitive models to the wider community**. In Westwood, J.D., Haluck, R.S., Hoffman, H.M., Mogel, G.T., Phillips, R., Robb, R.A. et al. (Eds.), Medicine Meets Virtual Reality 16: Parallel, combinatorial, convergent: NextMed by Design, pp. 281-286. Amsterdam; Berlin; Oxford; Tokyo; Washington, DC: IOS Press.
[19] McShane, M., Nirenburg, S., Beale, S., Jarrell, B., Fantry, G. and Mallott, D. 2013. **Mind-, body- and emotion-reading**. Proceedings of the Annual Meeting of the International Association for Computing and Philosophy (IACAP 2013).
[20] Nirenburg, S., McShane, M. and Beale, S. 2008. **A simulated physiological/cognitive "double agent"**. In Beal, J., Bello, P., Cassimatis, N., Coen, M. and Winston, P. (Eds.), Papers from the AAAI Fall Symposium "Naturally Inspired Cognitive Architectures". AAAI Technical Report FS-08-06. Menlo Park, CA: AAAI Press.
[21] Nirenburg, S., McShane, M., Beale, S. and Jarrell, B. 2008. **Adaptivity in a multi-agent clinical simulation system**. In Honkela, T., Pöllä, M., Paukkeri, M. And Simula, O. (Eds.), Proceedings of the 2nd International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning (AKRR'08), pp. 32-39.
[22] Nirenburg, S., McShane, M. and Beale, S. 2010. **Aspects of metacognitive self-awareness in Maryland Virtual Patient**. In Pirrone, R., Azevedo, R. and Biswas, G. (Eds.), Cognitive and Metacognitive Educational Systems: Papers from the AAAI Fall Symposium. AAAI Technical Report FS-10-01, pp. 69-74.
[23] McShane, M., Beale, S., Nirenburg, S., Jarrell, B. and Fantry, G. 2012. **Inconsistency as a diagnostic tool in a society of intelligent agents**. Artificial Intelligence in Medicine, 55(3): 137-48.
[24] Knepper, R.A., T. Layton, J. Romanishin, and D. Rus. 2013. IkeaBot: An autonomous multi-robot coordinated furniture assembly system. *IEEE International Conference on Robotics and Automation*.
[25] Roncone, A., O. Mangin and B. Scassellati. 2017. Transparent Role Assignment and Task Allocation in Human Robot Collaboration. Proceedings of International Conference on Robotics and Automation. Singapore.
[26] K. Erol, J. Hendler, and D. S. Nau 1994. HTN Planning: Complexity and Expressivity. Proceedings of AAAI-94.
[27] Piantadosi, S. T., Tily, H. & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition* 122, 280–291.
[28] English, J. and S. Nirenburg. 2007. Ontology Learning from Text Using Automatic Ontological-Semantic Text Annotation and the Web as the Corpus. Proceedings of the AAAI-07 Spring Symposium on Machine Reading.
[29] Nirenburg, S., T. Oates and J. English. 2007. Learning by Reading by Learning to Read. Proceedings of the International Conference on Semantic Computing.
[30] Scheutz, M. J. Harris and P. Schmermerhorn. 2013. Systematic Integration of Cognitive and Robotic Architectures. Advances in Cognitive Systems 2:277-296.

Experimental Evaluation of a Cooperative Automation Approach for Manned-Unmanned Teaming in Future Military Helicopter Missions

Fabian Schmitt

Yannick Brand

Georg Rudnick

Axel Schulte

GERMANY

fabian.schmitt@unibw.de, yannick.brand@unibw.de, georg.rudnick@unibw.de, axel.schulte@unibw.de

ABSTRACT

This paper describes a twofold approach of cognitive automation in the Manned Unmanned Teaming (MUM-T) domain and explains the concept, implementation and experimental evaluation with pilots of the Bundeswehr. The application includes the guidance of several Unmanned Aerial Vehicles (UAV) by the cockpit crew of a manned transport helicopter in future military scenarios. We introduce two types of artificial cognitive agents to compensate the resulting high and dynamic workload for the crew. Cognitive agents on board the UAVs are in a delegation relationship with the crew and carry out assigned tasks in a highly automated manner. In addition, an artificial cognitive agent in the form of an associate system supports the crew in a cooperative manner. The agent behaves similar to a human crew member and has the task of preventing pilot errors and reducing workload peaks. The support takes place in both mission planning and mission execution. The associate system dynamically adapts the extend of the support to the mental demands and the task context of the crew, as well as the criticality of the situation. We implemented this twofold approach as real-time-capable software modules in a helicopter mission simulator. To evaluate the overall system, an experimental campaign was conducted with crews of experienced pilots of the German Armed Forces. The results show the potential of the approach presented. The reconnaissance continuously ensured safety for the crew of the manned helicopter.

1.0 INTRODUCTION

The guidance of several Unmanned Aerial Vehicles (UAVs) from the cockpit of manned aircraft (Manned Unmanned Teaming, MUM-T) is a highly relevant field of research and technology [1], [2]. In this context, the Institute of Flight Systems conducts research on the guidance of unmanned flight systems by the cockpit crew of a two-seated helicopter. Here, the pilot flying is mainly responsible for the manual control of the helicopter, navigation, communication and system management. In addition to traditional tasks such as mission management, communication and system management, the commander is responsible for the UAVs. The new range of tasks includes the management and monitoring of UAVs, the evaluation of sensor results and the adaptation of the mission and flight plan based on the reconnaissance results obtained. This spectrum of tasks results in highly varying mental workload (MWL) conditions of the crew. Especially phases of high MWL can lead to overtaxing conditions. This can result in reduced crew performance, a significant reduction in situational awareness and an increased error rate. Our approach of *Human-Autonomy Teaming* (HAT) addresses these problems by introducing cognitive abilities on the part of automation. The goal is to increase

overall mission performance, balance workload and counteract human factors problems like out-of-the-loop problems [3], complacency and automation bias [4]. Therefore, we use two modes of cognitive automation, as depicted in Figure 1 [5][6][7]:

With the concept of *task-based guidance*, the crew is able to guide the UAVs by delegating tasks to the cognitive agent onboard each UAV (Figure 1, right side). The intelligent agents are able to understand and execute complex mission tasks in a highly automated manner. Thus, they use mission and system knowledge as well as the ability to plan and make decisions in complex scenarios. By delegating high-level tasks to automated systems, MWL is shifted from the crew to the agents. *Adaptable levels of automation* in the UAV guidance enable the crew to add additional knowledge at different levels of detail in order to compensate for the agent's lack of knowledge.

As second mode, a *workload-adaptive associate system* onboard the manned helicopter assists the crew. The associate system acts as artificial crew member and cooperates with the crew (Figure 1, right side). It supports the crew by modifying the task load by own initiative with the aim to correct, avoid, or even prevent pilot errors. To provide workload-adapted support, the agent generates a mental image of the crew and the environment (mission and tactical situation). The agent supports the planning and execution phase of the mission in dynamic environments based on *adaptive levels of automation*, enabling the manned-unmanned team to react quickly to dynamic tactical situations and unexpected events.

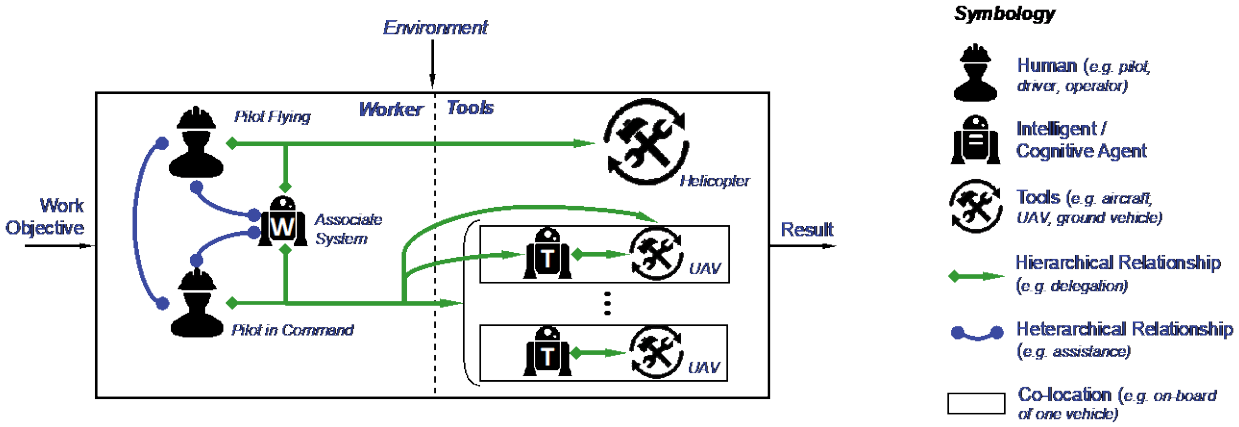


Figure 1: MUM-T work system

2.0 TASK-BASED GUIDANCE

The approach of task-based guidance for unmanned systems [8] offers the possibility to reduce the mental workload of the pilots by shifting cognitive load to the intelligent agent on board the UAVs. Such cognitive loads comprise the decomposition of UAV tasks into subtasks as well as the generation of flight plans and sensor commands. Thereby, the crew is able to manage several UAVs from the cockpit without changing the spectrum of use or usability. The relationship between commander and software agent is a delegation relationship similar to that of a human subordinate (Supervisory Control [9], green arrows in Figure 1). However, cognitive capabilities such as situation assessing, planning and problem solving are essential capabilities for such intelligent agents. Agents must be able to understand the tasks, break them down into subtasks, and execute them automatically. For returning the reconnaissance results obtained, the agent has to use the same, high level and a suitable format.

2.1 Levels of Automation in Task-based Guidance

We use the term *level of automation* in UAV guidance as a measure to describe the scope of necessary work steps of the pilot. A low level of automation (e.g. waypoint specification) can be compared with the direct manual control by the pilot, thus limiting the agent's scope for action. A high level of automation (e.g. task-based guidance) means that the pilot allows the agent more freedom in the implementation and is more interested in the result than in the sequence. Low levels of automation produce a higher workload, but they do allow better control, possibly a higher situation awareness, and the possibility to intervene in case of automation errors. This need for flexibility, especially in military missions, was also identified in earlier experiments on mission-based command and control. [10], [11]. The ability to compensate for automation errors and weaknesses and use human strengths plays a major role, especially in sensor and aerial image analysis. Above all, object recognition and friend-foe classification often push today's automation to its limits. [12], [13].

2.2 Scalable Autonomy in task-based guidance using variable levels of automation

The question now arises how to realize variable levels of automation in UAV management in general and from the cockpit in particular. Offering all available levels of automation at the same time and letting the pilot decide on their use brings some disadvantages. This includes possible unused cognitive resources of the agent, a poor structuring of the use of automation functions and problems if the agent runs in a low level mode and meanwhile losses its data link [14]. Another possibility is the hierarchical approach of scalable autonomy according to [15]. The pilot is able to access low levels of automation within a job in order to take advantage of the variable levels of automation described above. However, the concept ascertains that a task must be given first. The agent uses an HTN planner to break it down into subtasks which the pilot can access and modify in order to influence flight routes or sensor parameters, for example. Through this restriction, the agent knows the operator's intention expressed by the task and can provide appropriate support, for example through an associate system (chapter 3). The interventions are also possible during the plan execution and allow a specific and situation-dependent scaling of the used levels of automation. The agent was implemented as a knowledge-based system in the business rule engine Drools [16]. The system supports specific missions for transport helicopter missions, such as route and area reconnaissance, landing zone reconnaissance and tactical movement. Access to tasks and to low levels of automation are available for both pilots via the multifunction displays in the cockpit.

3.0 WORKLOAD-ADAPTIVE ASSOCIATE SYSTEM

The second mode of our human autonomy teaming concept is a workload-adaptive cognitive associate system that adapts its support to the current task situation and the mental state of the crew. The associate system cooperates with the pilots and thus is able to detect and correct human errors during the planning and execution phase of the mission, or to proactively prevent them.

3.1 Behaviour of the Associate System

The behavior of the associate system is derived from the behavior of a human copilot. The associate system pursues the mission objective independently and on its own initiative and supports the crew only if necessary. The extend of support is scaled according to the attention and mental workload of the crew. Therefore, the associate system follows behavior rules in the form of escalating basic requirements: The basic rule for this interaction is a defensive behavior of the associate system. If the crew is able to perform the task with the given automation tools, there is no reason to intervene in the functioning work process (basic requirement). Only if the pilot's attention is no longer focused on the most urgent task, it should be directed back to it (attention guiding, 1st requirement). If the pilot is overtaxed and therefore not able to solve a task situation, the associate system tries to simplify the task situation by transform it into a task situation, which can be handled by the crew again (task simplification, 2nd requirement). Only in the case of

a very high risk the associate system takes over tasks automatically (task adoption, 3rd requirement). In order to transfer the skills and capabilities of a supporting co-pilot to a technical associate system, the system needs knowledge of the domain, cognitive capabilities to interpret and plan based on the knowledge and an interaction model to communicate with the crew. The basis for the interaction between pilot and agent and also between all associate system components is a common task model (subchapter 3.2). The situation interpretation based on this task model is explained in subchapters 3.3 and 3.4. The adaptive support for the crew is described in chapter 3.5.

3.2 A task model for task-centered communication

The basic idea of a task-centered communication approach is that only the name of a task is sufficient for the communicating partners to generate a mental picture of the task, which includes scope, prerequisites and demands [17]. This task-centered communication approach is used for communication between pilots and the associate system on the one hand and between individual modules of the associate system on the other hand. The developed task model contains all mission and pilot tasks in a hierarchical structure. Mission tasks are vehicle-specific, domain relevant tasks. For each mission task, all pilot tasks which are required to carry out the mission task, are also stored in the model. Each pilot task contains the mental resources required to execute the task.

3.3 Mixed-Initiative Mission Planning

The goal of the mission planning process is to generate a plan, containing the necessary and optional plan elements, with which the mission goal can then be achieved. The contained tasks are based on the mission tasks from the task model. The associate system itself must be capable of planning and scheduling in order to derive the necessary action steps with which the pilot shall be supported in (re)planning if necessary. While there is probably sufficient time for detailed manual planning during mission preparation on the ground, unforeseen changes of either the tactical situation or the mission objective/mission constraints during the flight require time-critical replanning. Especially in situations of excessive workload, problems of human performance and errors can occur [18], [19]. For this reason, a manual mission planning is not suitable in most situations. Automated planning, on the other hand, reduces the demands placed on pilots, but can lead to problems such as loss of competence [20], [21]. Additionally, automated systems lack of transparency [22], create comprehension problems [23] and might yield to a loss of plan and situation awareness [24]. In extreme cases, the hierarchy may be inverted during plan execution, resulting in an operator executing a machine generated plan. A mixed-initiative approach enables cooperative planning where both human and automation bring their strength on own initiative [25]. The mission planner can actively propose plan elements to the pilot.

3.3.1 Modes of Automation in Mixed-Initiative Planning

The advantages and disadvantages of various automation approaches described above show that different situations require different characteristics of support by the mission planner. Therefore, our mixed initiative concept provides three modes of automation (Figure 2), varying in type and extend to support in planning phases adaptively. The type can vary between forward planning, flaw correction and optimization, the extend can vary from single plan elements to complete mission plans. The smaller the scope of support, the better the pilot can follow the proposal for plan adjustment. From this point of view, incremental support is suggested. It is difficult to interpret complex segment plans in a short time. On the other hand, several incremental support proposals take more time. For time-critical and complex planning tasks, this process may simply take too long. These can then be adapted to the pilot with the knowledge of the complexity of the planning tasks involved and their time criticality. The adaption process is described in chapter 3.5.

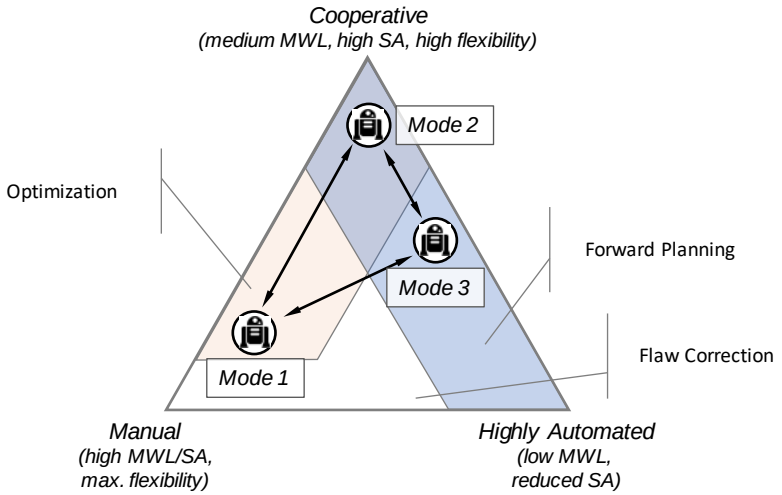


Figure 2: Three modes of automation provided by the mission planner

3.3.2 Mission Planner

To enable planning support, the pilot must inform the system about the desired mission objective. The mission planner then plans on the level of the mission tasks, i.e. links the individual mission tasks logically, spatially and temporally to a mission plan. Therefore, the planner uses knowledge about the mission objective and the task model. For the technical implementation, a PDDL planner [26] and the Constraints Optimization Solver CPLEX [27] are used as planning tools. We modelled the MUM-T domain comprising own, unknown and hostile forces, air space regulations, infrastructures, as well as possible actions in PDDL. Subsequently, the planning result is optimized with regard to various constraints such as threat exposure, time requirements, the resource requirements of the UAVs or the consideration of geospecific features such as the terrain. Pilot planning inputs on the tactical map displays are considered as logical constraints by the planner. Constraint relaxation is used to identify weaknesses in the implemented partial plan.

3.4 Activity and Mental Workload Determination

An activity determination serves as the basis for the determination of the mental resources currently needed. Since many, especially cognitive tasks, cannot be observed directly, the indirect approach of evidence-based activity recognition is pursued. [28], [29]. During the execution of the necessary tasks, the pilots interact with the system in a variety of ways. Interactions include manual interactions via helicopter controls, buttons and multi-touch screens (MFDs), auditive communication via aircraft radio and intercom, and visual attention to individual cockpit instruments. During these interactions, the system observes the pilots with the help of various measurement sensors. It uses these observations to combine the related evidences from the task model, which oppose or reject the execution of a specific task. Thereby, the system deduces the current activity of a pilot. Using the activity and knowledge of the task-specific demand on mental resources from the task model, the associate system estimates the total resource demand, which is necessary to execute the activity [30]. With this resource estimate, acute demand peaks can be detected and the level of support from the associate system can be adjusted accordingly.

3.5 Workload-adaptive Intervention Generation

The goal of the associate system is to avoid mistakes of the helicopter crew through deliberate interventions. For a proactive error avoidance, the estimation of the current demands is not sufficient. Therefore, the associate system uses the mission plan to predict future task situations and demand peaks [31].

3.5.1 Pilot Plan and Projection of Future Workload Peaks

Firstly, mission tasks are broken down into pilot tasks in order to identify which tasks the crew must perform in order to fulfil the mission. These planned pilot tasks are compared with the pilot's actual activities and all completed plan elements are checked. (green tasks in Figure 3). The remaining tasks not completed on time are triggers for intervention. (red tasks in Figure 3). The associate system uses the Constraints Optimization Solver CPLEX [27] to logically arrange all pilot tasks that have not (yet) been completed. The system considers dependencies between tasks and execution times of the individual tasks and simulates the resulting sequence of pilot tasks using the information of demand on mental resources. The system determines when areas of high workload may cause performance losses. These predicted workload peaks are a second trigger for intervention. In situations that are not foreseeable by the plan, such as a sudden change of the tactical situation, the associate system supports firstly to minimize the acute danger of the crew and secondly to create a new valid plan in cooperation with the crew as quickly as possible.

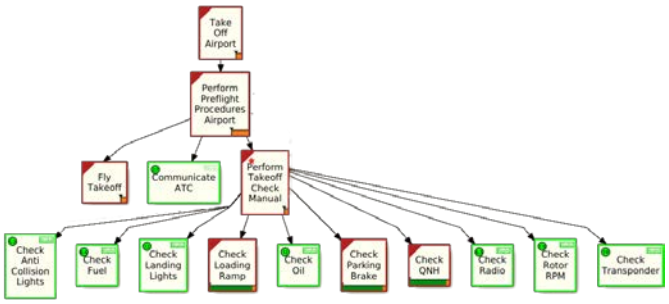


Figure 3: Detailed pilot plan of the mission task TakeOffAirport and corresponding required pilot tasks

3.5.2 Intervention Planning

If the associate system detects need for action in the form of intervention triggers (neglected tasks, acute and future workload peaks, as well as critical changes in the situation), it plans the intervention. The aim is to find an intervention strategy which, depending on the attention and demands on the crew and the criticality, leads minimally intrusively to the solution of the problem. With the help of activity recognition, the system first determines whether the pilots themselves are already working on a solution to the problem, i.e. whether their attention is directed correctly. If this is not the case, the consideration of mental resources determines whether the crew is able to solve the problem on their own. If so, only attention must be drawn to the most urgent task, otherwise support through task simplification or task assumption is necessary. It is only necessary to take over tasks if the crew cannot solve a mission-critical problem themselves. The described decision-making process, from the identification of the triggers, through the selection of an appropriate intervention strategy, to its realization, is implemented in the cognitive framework Soar [32].

3.5.2 Intervention Strategies

The associate system uses warning sounds, voice output and the multifunctional display of the helicopter to direct attention (Figure 4). The task simplification and adaptation to the task situation is achieved on the one hand by pointing out solution strategies and on the other hand by changing levels of automation (subchapters 2.1 and 3.3). By selecting a higher level of automation, the associate system transfers task components from the pilot to the associate system, thus reducing the overload on the pilot. The cognitive agents on board the UAVs, the mission planner and the sensor and perception management provide these levels of automation. In the event of a high risk of serious consequences, the associate system can take over partial tasks or entire tasks on its own initiative as the highest escalation level. Extreme examples are the intervention in flight control to avoid collisions or the adoption of an uncompleted landing checklist. The associate system adopts

planner proposals if otherwise the successful achievement of the mission objective is highly unlikely. The strategy for simplifying and adopting tasks is illustrated using mission planning process as an example: If the system determines that the mental demands of the planning task are high, it adapts the extend of support to the situation taking the time criticality of the planning problem into account. If the pilot does not react despite a high degree of criticality, the system decides to take on a task at the latest possible time. The proposed solution is then automatically accepted.

Figure 4: Detailed Dialog to direct the pilot's attention including the highlighting of the interaction element (here: button) that solves the problem (pink) and the overlayed gaze tracking data (green dot at the top edge)4

4.0 EXPERIMENTAL EVALUATION

To evaluate the overall system, an experimental campaign was conducted with pilots of the Bundeswehr in the helicopter mission simulator of the Institute of Flight Systems at the University of the Bundeswehr in Munich. Seven trained military pilots at an age between 31 and 59 years (M = 50.4, SD = 9.2) and a flying experience between 535 and 6850 (M = 3933, SD = 1807) total hours served as test persons. The pilots were divided into four crews. One of the pilots took the role of commander, while the other pilot took the role of pilot flying.

4.1 Missions

We designed five realistic MUM-T missions, which are based on current missions of the German military (see Table 1). These missions included the transport of own forces into or from military operation areas in hostile territory, each conducted by a single transport helicopter accompanied by three sensor-equipped, unarmed UAVs. The crew of the manned helicopter managed the UAVs under their own responsibility. The commander used the sensors of the UAVs for area and route reconnaissance. Unknown and clearly hostile forces appeared within the operating area, which had to be located and identified by the UAVs. In addition to such sudden changes on the tactical situation, 4 out of 5 missions involved at least one change of the mission objective in order to create alternating workload conditions. Each mission has individually defined airspace regulations in order to avoid accustoming effects.

Table 1: Mission overview.

Mission name	Content	Events	Duration [min]
1 Special Observation	EOD-Transport	-	45
2 Golden Hour	MedEvac	Change of Destinataion	40
3 Desert Shield	Troop Transport	CasEvac	50
4 Rocket Raid	Troop Transport	Change of Target	65
5 Mosahi Konvoi	Troop Transport	Pers. Recovery	45

4.2 Experimental Procedure

Due to the complexity of the experiment, the sequence of missions was maintained throughout the entire experiment. Each crew completed a two-day training course with tutorial and two complex training missions. All crews then flow the five missions in the same order. The mission time varied between 31 and 71 minutes (M = 45.77 min SD = 10.25 min). The pilots evaluated the scenarios using seven-stage Likert scales ranging from 1/negative to 7/positive. They rate the realism of the scenarios as very well (M = 5.6, SD = 0.7) and rate the course of the missions also very positively (M = 5.3, SD = 0.83). The pilots evaluate the simulation environment also positively (M = 5.0, SD = 1.0). After each mission, we conducted a detailed debriefing of selected situations, in which use cases of the UAVs or the associate system occurred.

4.3 Results of the Overall Approach

During the missions studied, the helicopter covered an average flight distance of 92 km (MW=92.15 km, SD=28.77 km in a time of 46 minutes (MW=45.82 min, SD=9.97min). The UAVs investigated an average of 172 km^2 per mission (MW=171.96 km^2, SD=70.33 km^2). This results in an average reconnaissance performance of 3.75 km^2 per minute. This high reconnaissance performance also justifies the average time and distance during which the helicopter was over reconnaissance area (Figure 5). The helicopter was over 90% of the time and distance over reconnaissance area. Since enemy forces could be detected over a wide area, the helicopter stayed over 90% of the time and distance over confirmed enemy-free areas, which shows a very large increase of mission security compared to today's missions.

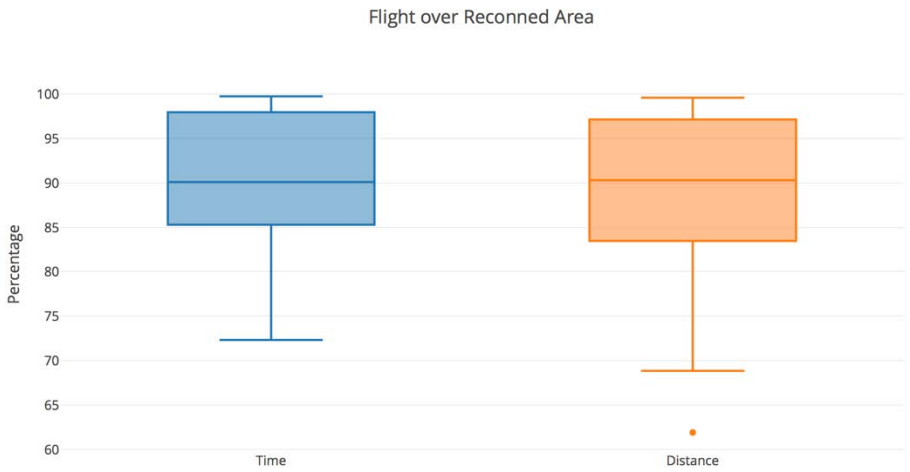


Figure 5: Flight over confirmed cleared area

4.4 Results Task-Based UAV Guidance

The high level of reconnaissance services listed under the overall results is primarily due to the use of UAVs. A recording of the subjective evaluation of the UAV system with the aid of an adjusted Cooper Harper rating according to [33] brought the result shown in Figure 6. After each individual mission, the same questionnaire was presented to the subjects and the values averaged (individual evaluation). After completion of all missions, the questionnaire was presented to the subjects again, with the request to incorporate their experience over the entire week (overall assessment). The subjects evaluated mission aptitude between 2 and 4, with 1 being the best and 10 the worst. This means that a sufficient mission performance (rating 3-6) is guaranteed in any case and in most cases a satisfactory mission performance (rating 1-3) has been achieved. It is also interesting that the overall evaluation, which took place after the completion of all missions, is slightly better than the average of the individual evaluations. This could be due to a learning effect. With greater familiarity with the UAV system, the mission performance and thus the evaluation increases.

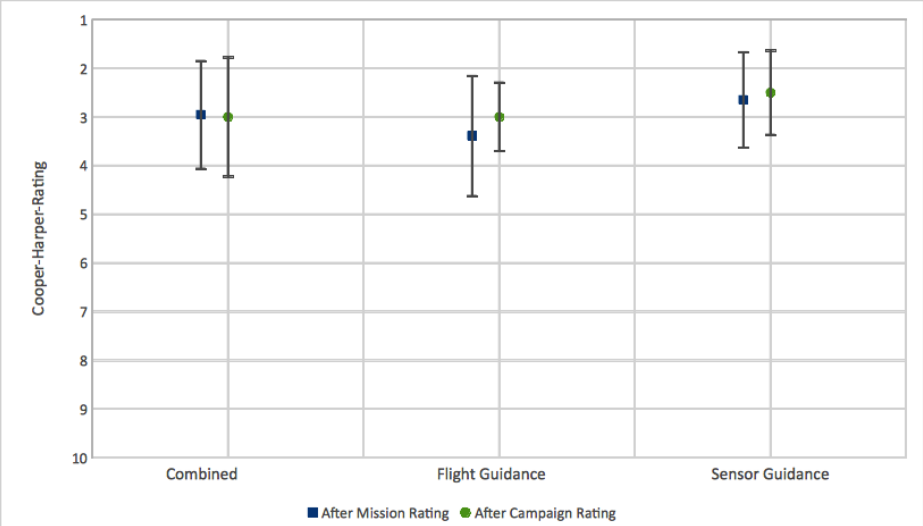


Figure 6: Cooper-Harper Rating for UAV Guidance

4.5 Results Mixed-Initiative Planning

Three use cases with different planning complexity and time criticality were investigated for each mission: At the beginning of the mission, a complete planning has to be be carried out with little time criticality. A pop-up threat makes it necessary to replan the manned helicopters and UAVs locally with high time criticality. A change of the overall mission objective during the existing mission requires a complete time-critical replanning. The results of the intervention behaviour for the described use cases are shown in Figure 7. During initial planning, in 14 out of 20 cases, the system remained silent. A total of 25 pop-up threats appeared, which required an active reaction from the pilot. In 18 cases, a proposal generated by the agent was accepted by the pilot. In 4 cases the pilot switched to a pre-planned alternative route and in 3 cases the pilot carried out a replanning himself. In all these cases, the system offered support on the second level of automation. For the third use case, in 19 out of 20 cases the system supported on the highest level of automation. In 18 cases the system's proposal was accepted. This can be explained by the fact that the planning effort for manual replanning by the pilot would have been very high. After each mission, the pilots evaluated the different extend of intervention for each of the three use cases (Figure 8). The extend was assessed as very suitable by the test persons. Overall, the results show that the mixed initiative approach with adaptive extend of support was very positively received.

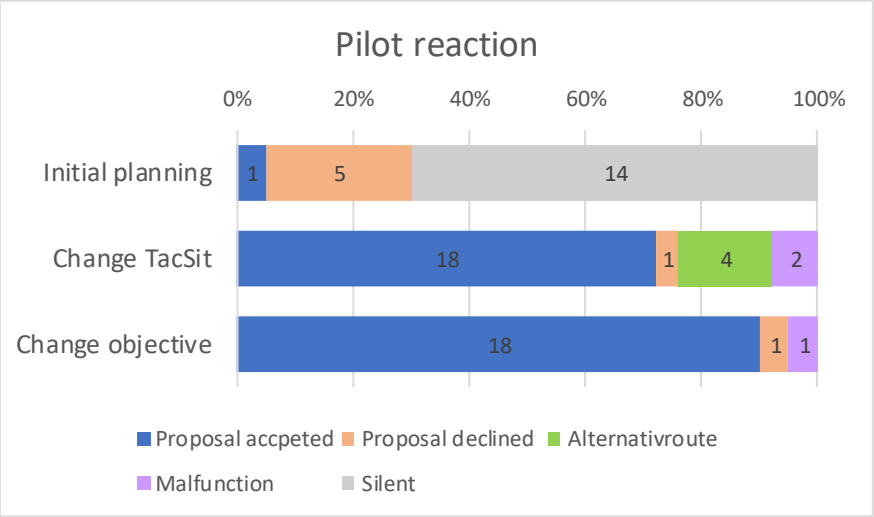


Figure 7: Agent action for different use cases and corresponding pilot reaction

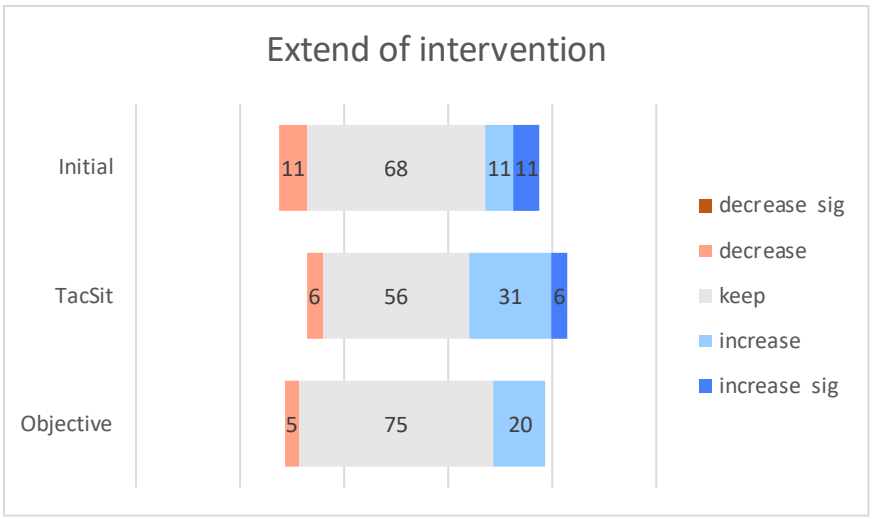


Figure 8: Pilot's subjective ratings for extend of intervention

4.6 Results Adaptive Associate System

An example of an associate system intervention is explained below. In the specific situation the commander fails to assign a reconnaissance task to a UAV on time (pink route segments in Figure 9 and pink highlighted pilot tasks *Plan Mission* in Figure 10). In addition, he is currently involved in an ongoing route reconnaissance task (task blocks highlighted in green in Figure 10). The parallelism of these two task situations leads to high workload (orange plot in Figure 10). The associate system detects a workload peak for the immediate future. Due to the fact that a planning task is jointly responsible for the workload peak, the mission planner is commanded to solve these planning tasks on a high level of automation in order to reduce the task load on the crew for the identified high-workload period. The planner then generates a task proposal for the forgotten route reconnaissance (the associate system dialog with a yellow border in Figure 9). Because of the high criticality and urgency of the problem and the high demands on the crew, the associate system decides to accept this proposal without further notice. The commander assessed this intervention of

the associate system as very helpful and appropriate for the high-risk and time-critical situation. Overall, the pilots assessed the associate system interventions as expedient and mostly helpful. The type of intervention selected – attention guiding, task simplification and task adoption - was assessed as mostly situation-adaptive and appropriate. The interventions were considered justified, which is why it can be assumed that there was an understandable reason for support by the associate system in most cases. The scope of the intervention was mostly correct, only one subject found that the interventions took too much work off. Two subjects tended that the interventions must support more (Figure 11).

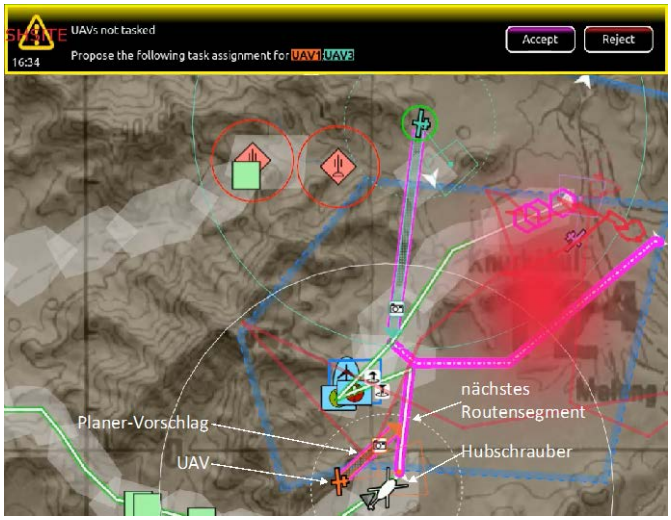


Figure 9: Tactical map of helicopter mission simulator and dialog of associate system informing about missing tasks

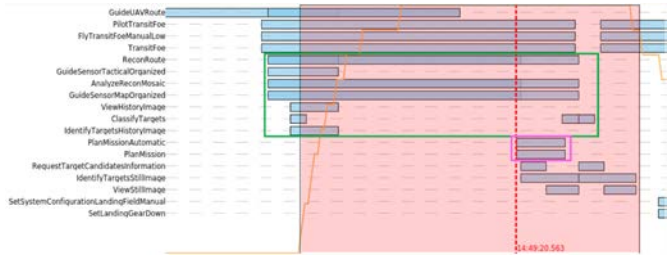


Figure 10: Projected pilot workload (orange line) und detected workload peaks (red area) calculated based on future task situation (blue blocks)

	überhaupt nicht					voll und ganz	
Die Art der Interventionen ist situationsangepasst.		●		●		●●●●	●●
Die Art der Interventionen ist angemessen.			●	●●		●●●●	●
Die Interventionen waren hilfreich.			●●		●●●●	●	●
Die Interventionen waren sinnvoll.				●	●●●	●●●	●
Die Interventionen waren NICHT gerechtfertigt.	●	●●●	●●●●				
Die Interventionen nahmen mir zu viel Arbeit ab.		●●●●	●	●●		●	
Die Interventionen hätten mich mehr unterstützen müssen.	●	●●●	●●		●●		

Figure 11: Pilot's subjective ratings of associate system interventions (n=8).

5.0 CONCLUSION

In this paper, a concept of adaptive assistance with cognitive automation was presented. The proposed concept was demonstrated using the MUM-T as an example containing The presented implementation was successfully evaluated with pilots of the German armed forces, the results show the potential of the two-tier approach. Both flight time and route over reconnoitered area show that the developed system can significantly increase crew safety compared to today's systems. The support provided by the associate system was also very well received. Through its interventions, the crew's scope of action was increased to the extent that cognitively demanding tasks were temporarily simplified or taken over by the associate system. Especially in situations of high workload, the system was able to reduce human erroneous performance. Further experiments to evaluate specific research questions are currently being conducted. The concept presented here can also be applied to other domains, such as civil defense, autonomous driving and maritime use cases.

REFERENCES

[1] [1] M. Lundy, M. Taylor, S. K. Frazee, J. Meinders, C. Milstead, D. C. Engelhardt, F. Capri, M. J. Humble, J. Q. Bolton, and T. C. Walsh, “Manned-Unmanned Teaming,” *Aviation Digest, United States Army, July - September 2014*, vol. 2, no. 3, pp. 1–31, 2014.

[2] R. Whittle, “MUM-T is the Word for AH-64E: Helos Fly, Use Drones,” *Breaking Defense*, 2015. .

[3] M. R. Endsley and E. O. Kiris, “The out-of-the-loop performance problem and level of control in automation,” *Human factors*, vol. 37, no. 2, pp. 381–394, 1995.

[4] R. Parasuraman and D. H. Manzey, “Complacency and bias in human use of automation: an attentional integration.,” *Human factors*, vol. 52, pp. 381–410, 2010.

[5] R. Onken and A. Schulte, *System-Ergonomic Design of Cognitive Automation: Dual-Mode Cognitive Design of Vehicle Guidance and Control Work Systems*. Heidelberg, 2010.

[6] A. Schulte, D. Donath, and D. S. Lange, “Design Patterns for Human-Cognitive Agent Teaming,” *Lecture Notes in Computer Science*, vol. 9736, pp. 231–243, 2016.

[7] A. Schulte and D. Donath, “A design and description method for human-autonomy teaming systems,” *Advances in Intelligent Systems and Computing*, vol. 722, pp. 3–9, 2018.

[8] J. Uhrmann and A. Schulte, “Concept, Design and Evaluation of Cognitive Task-based UAV Guidance,” *International Journal on Advances in Intelligent Systems*, vol. 5, pp. 145–158, 2012.

[9] T. B. Sheridan and W. L. Verplank, “Human and Computer Control of Undersea Teleoperators,” *ManMachine Systems Lab Department of Mechanical Engineering MIT Grant N0001477C0256*, no. DECEMBER 1978, p. 343, 1978.

[10] S. Gangl, “Kooperative Führung mehrerer unbemannter Luftfahrzeuge aus einem einsitzigen Kampfflugzeug,” 2014.

[11] J. Uhrmann, “Auftragsbasierte Multi-UAV-Führung aus dem Helikoptercockpit durch kognitive Automation,” Universität der Bundeswehr München, 2012.

[12] C. Ruf and P. Stütz, “Model-driven Sensor Operation Assistance for a Transport Helicopter Crew in Manned-Unmanned Teaming Missions: Selecting the Automation Level by Machine Decision-

making," in *7th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences, AHFE 2016*, 2016.

[13] C. Ruf, S. von der Mülbe, and P. Stütz, "Assistenz beim Einsatz von Aufklärungssensorik für die Crew eines Transporthelikopters in MUM-T-Missionen: Gestaltung eines Automationssystems zur gezielten Reduktion der Mental Workload," in *6. Interdisziplinärer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten*, 2017.

[14] G. Rudnick, A. Schulte, and A. Rudnick Georg; Schulte, "Experimental Evaluation of a Scalable Autonomy Concept for Cognitive Agents aboard Reconnaissance UAVs," in *19th International Symposium on Aviation Psychology (ISAP)*, 2017, pp. 1–6.

[15] G. Rudnick and A. Schulte, "Scalable autonomy concept for reconnaissance UAVs on the basis of an HTN agent architecture," in *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2016, pp. 40–46.

[16] Drools, "Drools Business Rule Engine," *Red Hat Inc. https://www.drools.org/*, 2016. .

[17] F. Honecker, Y. Brand, and A. Schulte, "A Task-centered Approach for Workload-adaptive Pilot Associate Systems," in *Proceedings of the 32rd Conference of the European Association for Aviation Psychology – Thinking High AND Low: Cognition and Decision Making in Aviation*, 2016.

[18] E. L. Wiener, "Beyond the Sterile Cockpit," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 27, no. 1, pp. 75–90, 1985.

[19] E. L. Wiener, "Human Factors of Advanced Technology (" Glass Cockpit ") Transport Aircraft," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, no. June, 1989.

[20] M. L. Cummings, "Automation Bias in Intelligent Time Critical Decision Support Systems," in *AIAA 3rd Intelligent Systems Conference*, 2004.

[21] E. L. Wiener and R. E. Curry, "Flight-deck automation: promises and problems," *Ergonomics*, vol. 23, no. 10, pp. 995–1011, 1980.

[22] C. Billings, *Aviation automation: The search for a human-centered approach*. Mahwah: Lawrence Erlbaum Associates, 1997.

[23] E. Svensson, M. Angelborg-Thanderz, L. Sjöberg, and S. Olsson, "Information complexity - mental workload and performance in combat aircraft," *Ergonomics*, vol. 40, no. 3, pp. 362–380, 1997.

[24] M. R. Endsley, "Automation and Situation Awareness," in *Automation and Human Performance: Theory and Applications*, R. Parasuraman and M. Mouloua, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 1996, pp. 163–181.

[25] M. H. Burstein and D. V. McDermott, "Issues in the development of human-computer mixed-initiative planning," *Advances in Psychology*, vol. 113, pp. 285–303, 1996.

[26] D. McDermott, M. Ghallab, A. Howe, and C, "PDDL-The Planning Domain Definition Language,"

The AIPS-98 Planning, 1998.

[27] IBM Corporation, “IBM ILOG CPLEX Optimization Studio,” 2017. .

[28] F. Honecker and A. Schulte, “Evidenzbasierte Pilotentätigkeitserkennung unter Berücksichtigung unterschiedlich zuverlässiger Beobachtungen,” in *57. Fachausschusssitzung Anthropotechnik der DGLR: Kooperation und kooperative Systeme in der Fahrzeug- und Prozessführung*, 2015, pp. 115–130.

[29] F. Honecker and A. Schulte, “Konzept für eine automatische evidenzbasierte Online-Pilotenbeobachtung in bemannt-unbemannten Hubschraubermissionen,” in *4. Interdisziplinärer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten – Verstehen, Beschreiben und Gestalten kognitiver technischer Systeme*, 2015.

[30] C. Wickens, “Multiple resources and mental workload,” in *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 2008, vol. 50, no. 3, pp. 449–455.

[31] Y. Brand and A. Schulte, “Model-based Prediction of Workload for Adaptive Associate Systems,” in *Proceedings - 2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, 2017, pp. 1722–1727.

[32] J. E. Laird, “The Soar Cognitive Architecture,” *AISB Quarterly*, vol. 171, no. 134, pp. 224–235, 2012.

[33] M. C. Cotting, “UAV Performance Rating Scale Based on the Cooper-Harper Piloted Rating Scale,” *49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, no. January, 2011.

Pluggable Social Artificial Intelligence for Enabling Human-Agent Teaming

J. van Diggelen (PhD), J. S. Barnhoorn (PhD), M. M. M. Peeters (PhD), W. van Staal (MSc), M. van Stolk (MSc), B. van der Vecht (PhD), J. van der Waa (MSc), J.M. Schraagen (Prof., PhD)
TNO
PO Box 23
3769 ZG Soesterberg
THE NETHERLANDS

{jurriaan.vandiggelen, jonathan.barnhoorn, marieke.peeters, martin.stolk, bob.vandervecht, jasper.vanderwaa, jan_maarten.schraagen}@tno.nl

ABSTRACT

As intelligent systems are increasingly capable of performing their tasks without the need for continuous human input, direction, or supervision, new human-machine interaction concepts are needed. A promising approach to this end is human-agent teaming, which envisions a novel interaction form where humans and machines behave as equal team partners. This paper presents an overview of the current state of the art in human-agent teaming, including the analysis of human-agent teams on five dimensions; a framework describing important teaming functionalities; a technical architecture, called SAIL, supporting social human-agent teaming through the modular implementation of the human-agent teaming functionalities; a technical implementation of the architecture; and a proof-of-concept prototype created with the framework and architecture. We conclude this paper with a reflection on where we stand and a glance into the future showing the way forward.

1.0 INTRODUCTION

Recent developments in Artificial Intelligence (AI) technology, computational processing power, and the availability of data have given rise to increasingly intelligent systems, i.e. entities capable of engaging in dynamic and goal-directed interaction with their environment [48]. Intelligent systems are often described in terms of their behaviour and capabilities: intelligent systems can *sense* their environment, *reason* about their observations and goals in order to make *decisions*, and *act* upon their environment [62]. Due to their processing speed and their vast and almost infallible memory, intelligent systems outperform humans in handling large amounts of (heterogeneous) data, dealing with complex problems, and rapid decision-making [59].

A more recent development is that, for certain tasks, intelligent systems are capable of operating at high performance levels for extended periods of time without the constant need of human support, guidance, or intervention [55][4][36][41]. Concerns about the proliferation of intelligent systems in defence, healthcare, aviation, and other high-risk domains - the military domain being one of the most prominent domains under debate [32][8] - has ignited heated debates around the globe [48], [49], [6], [14].

Central to these debates about intelligent systems is the term ‘meaningful human control’ [19][1][3][32]: How can intelligent systems be developed in such a way that humans remain in control of the behaviour and effects of said systems. Ultimately, such control is necessary to allow for human responsibility and accountability for potential outcomes of the deployment of intelligent systems.

One way of approaching the realization of meaningful human control is human-agent teaming (HAT). HAT aims to fully benefit from a system’s autonomous capabilities while still maintaining meaningful human control. This is accomplished by endowing an intelligent system with a variety of team behaviours, making

the system anticipatory, sensitive, and responsive to the needs, wishes, intentions, control, and/or influence of other team members. Examples are: pro-actively sharing bits of information to maintain a minimally required extent of shared situation awareness; being (re-)directable at a higher level of abstraction (e.g. strategic or tactical); anticipating (human) team members' actions; or foreseeing potential problems, sending human teammates timely warnings, and asking them for assistance. Most common architectures for autonomous systems (such as 4D/RCS [1]) ignore teaming functionalities and place the focus on task-oriented Artificial Intelligence (TAI), such as planning and sensing. The purpose of this paper is to describe a method for pluggable Social Artificial Intelligence (SAI) which allows developers to complement an existing autonomous (TAI) system with the capability to team up with humans.

The framework, called SAIL (Social Artificial Intelligence Layer) can be added to an autonomous system any time during commissioning or, at a later stage, while the system is in use. SAIL provides an infrastructure and a library of common HAT behaviours, promoting reusability, problem decomposition, and adaptability. This paper discusses the application of SAIL from functional analysis to functional design, to system architecture design to technical implementation.

We start our discussion in Section 2 on functional analysis by distinguishing five dimensions that can be used to characterize a Human-Agent Team, such as spatial dispersion of the team members, time criticality, and communication characteristics. Using three illustrative examples from the defence domain (mine hunting, aerial surveillance, and robot-assisted house search), we will demonstrate the application of the framework. These scenarios impose different requirements on a HAT and illustrate the scope of the problem space representing military HAT applications.

We have extracted a number of common high-level functions needed to enable team collaboration between humans and autonomous systems. These functions form the main HAT-functions provided by SAIL. For example, we identify a *proactive communication* function that decides whether a particular piece of information is relevant for a human given the current task context, user state, and system capabilities. Another common function within HAT is an *explainable AI* function that allows an autonomous system to explain why it has chosen a certain course of action and select useful explanations to offer to its human user to increase its predictability. Seven common HAT functions are described in Section 3.

The HAT functions are realized in the SAIL software architecture in a modular way using dedicated SAI-components which can be plugged into the autonomous system. Communication between these components is facilitated by our newly developed Human-Agent Teaming Communication Language (HATCL). This language provides the constructs needed for all components (i.e. TAI, SAI, and human) to coordinate actions among one another, ultimately leading to a coherent team consisting of humans and autonomous (TAI) systems glued together by SAI components. Of particular concern is to enable a mapping from the concepts in the internal control logic in the autonomous system to HATCL and back. This is realized in SAIL by so-called *semantic anchors*. A description of the SAIL architecture, HATCL and semantic anchors is provided in Section 4.

Using the SAIL framework, we have implemented a prototype application in which a swarm of military surveillance drones running in the Gazebo simulation environment can be controlled in a meaningful way using HAT techniques. The prototype is described in Section 5.

Section 6 presents a conclusion and future activities.

2.0 ANALYSING HAT SCENARIOS IN VARIOUS DIMENSIONS

Teamwork and collaboration requires interaction and tuning, especially in a HAT. What type of interaction, coordination, and alignment is needed, strongly depends on the type of HAT and the context in which it

operates. Therefore, we have identified five important dimensions that describe and define a specific HAT, and that can be used as a guidance when determining specific requirements for a given HAT.

2.1 HAT dimensions

We describe a set of five dimensions that affect the requirements of a specific HAT: environment, mission/task, team organization, team dynamics, and communication (Figure 2-1 provides an overview). The dimensions can be used in multiple ways: to evaluate the generic applicability of HAT solutions for a wide variety of tasks, teams, contexts, and situations; to support rapid prototyping and analysis of specific HAT cases and identify challenges and requirements; or to compare multiple scenarios or cases to one another in terms of complexity and/or required SAI functionality.

2.1.1 Environment

The environment dimension describes which part of the environment is (a) dynamic, as opposed to static, and (b) predictable, and to what extent. This dimension affects, among others, HAT requirements aimed at the establishment of (shared) situation awareness (SA). Situation Awareness is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [20]. In teams, SA needs to be distributed optimally across team members. Acquiring and maintaining shared SA becomes increasingly difficult as the environment becomes more dynamic and unpredictable [21]. Furthermore, complex environments also make it more challenging to determine when or in which manner the autonomous system may reach the boundaries of its capability envelope [57], posing additional challenges for the human to quickly switch to manual control if needed.

2.1.2 Mission / task

The mission / task dimension describes four factors. First of all, the (a) duration of the work cycles - which can be short, long, repeated, continuous, or team lifespan [54] - affects training and evaluation requirements. For instance, when teams collaborate for brief durations, extensive training and / or preparation may be needed because there is little time for tuning or adjustments during actual task performance. Secondly, the (b) interdependency of the team (e.g. [47]) can affect the need for interpredictability (also see Section 3.2) and shared mental models of team members' capabilities and status as members depend on each other's performance and actions. The other two factors of the mission / task dimension are (c) time criticality, affecting communication requirements, and (d) risk. From a social perspective, increased risk may require artificial teammates to be more aware of human team members' emotional status, which can potentially affect task performance.

2.1.3 Team organization

The team organization dimension describes (a) the team's physical proximity, (b) the number of team members, (c) the team's adaptability [47], and (d) their skill and authority differentiation (e.g., the extent to which the team members have different specialisms and ranks, [54]) and (e) network structure. Different organizational set-ups require different behaviour from teammates. For instance, we know from the human-human teaming literature that in distributed teams, team trust is an important mediator of success, and can be increased through effective knowledge sharing and exchange behaviours [24][58]. Proactive communication in HATs, a requirement we elaborate on later, contributes to this. The team adaptability factor is described as the ability to alter a course of action or team repertoire in response to changing conditions and is thus especially important in dynamic and / or unpredictable environments.

2.1.4 Team dynamics

The team dynamics dimension contains (a) the temporal scope - a team may be standing, ad-hoc, to be formed in the future or having ceased to exist [54]. The team's temporal scope may have implications for the extent

to which teams are capable of forming longitudinal reciprocal and personalized team relationships. For example, in ad-hoc teams, the artificial teammate may have little or no time to develop a user model, and needs to be able to quickly identify roles, capabilities, and responsibilities. This is related to human-awareness, a requirement we elaborate on later in Section 3. The second aspect of the team dynamics dimension is (b) the team's current development phase - which cycles trough commissioning, preparation, action, and debriefing / learning. The development phase of the team, and especially the extent to which they actively cycle through these phases, may affect the extent to which teams engage in after action review and reflection to incrementally improve their processes and procedures. For instance, teams that cycle through preparation, action and learning phases may benefit from the ability to make work agreements among the team members, e.g. about constraints imposed on artificial agents, a functionality that we have implemented a solution for in SAIL (see Sections 3 and 4).

2.1.5 Communication

The communication dimension describes (a) the communication streams, which may vary between many-to-many, one-to-many, or one-to-one, (b) the information richness in communication [12], and (c) the quality of the infrastructure, which may vary in reliability, bandwidth, and range. During the design phase of a HAT, the communication dimension needs to be taken into account in order to optimize how, for instance, (shared) SA will be maintained [50]. For instance, when working in an unreliable network, relevant information needs to be pushed whenever possible, whereas with a reliable network, the artificial teammate may take other teammates' workload into account when timing communication. Furthermore, reduced communication abilities may increase the need to extensive training and clear work agreements so that teammates know what to expect from one another during periods of limited communication.

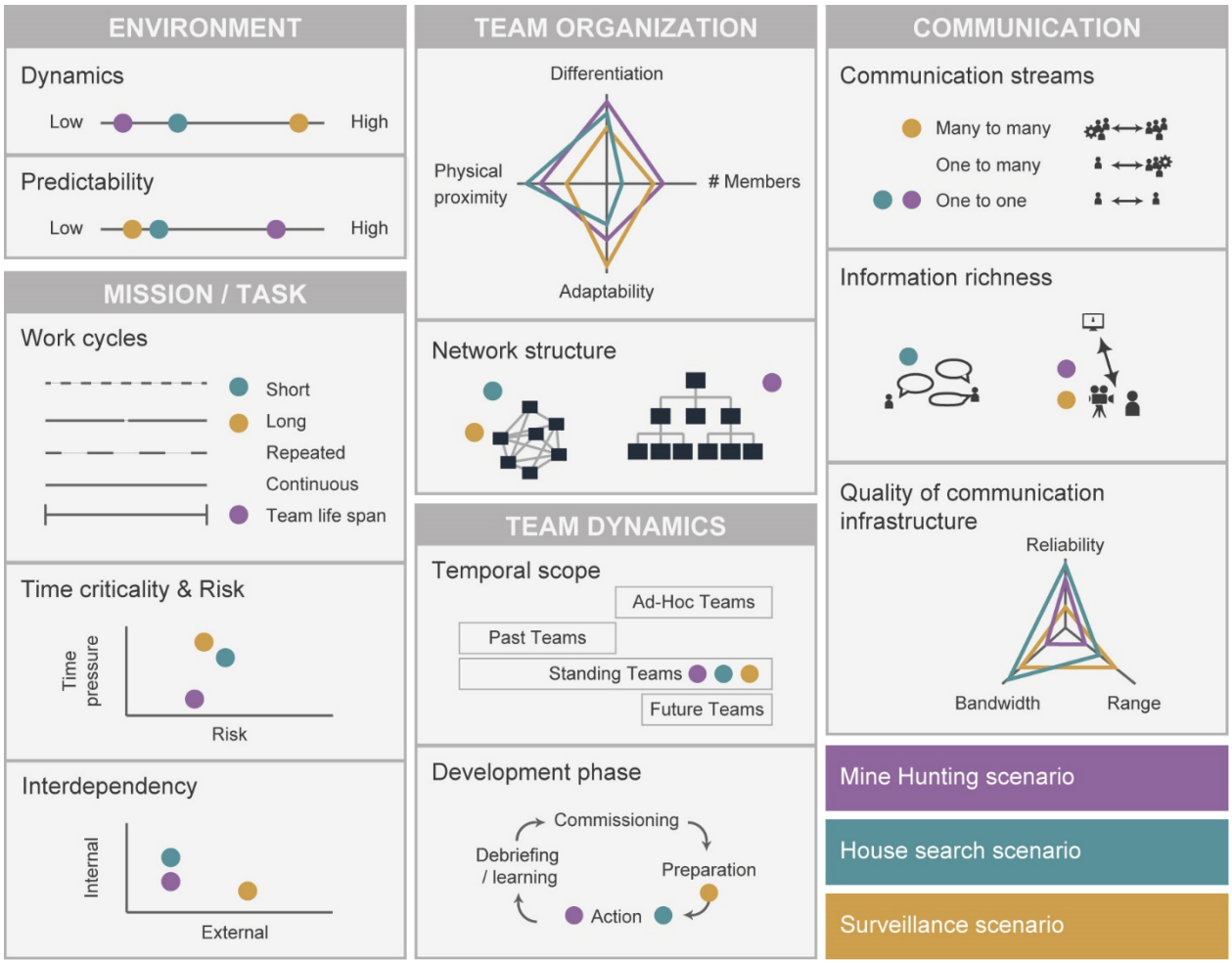


Figure 2-1 Overview of the HAT dimensions, and the way each of the scenarios maps onto these factors. Scenarios are indicated by colour-coding (yellow, green, and purple).

2.2 Example scenarios illustrating the dimensions

To illustrate the five dimensions presented in the above, we will discuss and analyse three typical defence scenarios. Figure 2-1 provides a visual overview comparing each of the scenarios with respect to the various team characteristics.

2.2.1 Surveillance scenario

In the Surveillance scenario, a swarm of unmanned aerial vehicles (UAVs) collaborates with a base commander to conduct a surveillance task around a small, temporary military base. The team has a low physical proximity, meaning that it requires the ability to efficiently exchange (shared) situation awareness. Differentiation of skill among the UAVs is low, allowing for quick back-up behaviour within the team. Furthermore, as the task concerns mainly surveillance, differentiation of authority between UAVs and the base commander is relatively low. In normal conditions, the UAVs can autonomously select sensors; plan and coordinate their flight paths; and choose what information to send to the base commander. However, when anomalies are detected or when the UAVs are unable to cope with a situation autonomously, the base commander quickly needs to (re)gain SA and decide on subsequent actions. This poses interesting requirements to the HAT: in normal conditions the marine officer has a low workload and SA, but in case of

potential (often time critical) anomalies or unexpected situations he or she needs to be able to effectively transfer from 'out-of-the-loop' to 'in-the-loop' as quickly as possible (i.e. management by exception).

2.2.2 House search scenario

In the House search scenario, a single soldier teams up with a single unmanned ground vehicle (UGV) to perform a house search. Short work cycles are used as multiple houses may be searched subsequently. This means that there is room for (longitudinal) team development as the team repeatedly cycles from preparation to action and evaluation. Furthermore, the team has high internal, but usually low external dependency. The high internal dependency means that consistent behaviour, potentially enforced trough work agreements, and trust calibration are important. For instance, under what circumstances does the UGV perform below average in terms of situation assessment? And when trust is damaged, how can the artificial teammate repair it (e.g. [57][46][58])? The risk is high while time pressure is average, this means that it is probably most effective to proactively provide updates and request confirmation regularly while still considering the other teammate's current mental state. The house search team often needs to jump right into action and the environment is highly unpredictable, dynamic, and variable. This means that the team needs to be highly trained so that the team members work together seamlessly, even when they have little time to prepare for action while facing unpredictable situations.

2.2.3 Mine hunting scenario

In the Mine hunting scenario, personnel on a naval mine hunter teams up with a swarm of unmanned underwater vehicles (UUVs) to search for and dismantle sea mines. In this scenario, ample time is available for commissioning and preparation as there is usually relatively little time pressure. However, the availability of communication infrastructure during action is restricted due to limitations of underwater communication. This means that common ground is essential and work agreements are required that take unexpected situations into account, since agents may need to solve problems without being able to communicate. The mine hunting case is a typical example of a predictable, static environment with a team that is long-standing.

2.2.4 Applicability to other cases

The dimensions discussed above can also be used to describe and analyse other examples of military HAT. For instance, the goal keeper used to defend navy ships against incoming missiles is characterized by repeated work cycles with extreme time pressure and high risk. The wide range of potential HAT applications, situations, and settings makes it clear that HAT functions and solutions that aim to be generic need to be flexible, adaptable, and expandable. Together, SAIL and HATCL provide a platform to facilitate this. For the remainder of the paper, the surveillance scenario will be used as an illustrative running example.

2.2.5 Added value of HAT analysis

The HAT analysis can be used in many ways: to evaluate whether concept HAT solutions are as generic as imagined; to analyse specific HAT situations or cases and identify which capabilities are required for the artificial teammate; or to compare a set of scenarios or cases. In our experience, analysing cases, scenarios, and HAT solutions in this way can be very helpful in designing better human-agent interaction. Furthermore, it helped identifying the most common HAT functions, which we discuss in the next section.

3.0 COMMON HAT FUNCTIONS

This section describes the most common functions of agents participating in Human-Agent Teams. Our starting point are the three basic requirements for effective coordination [22][36]:

- *Common ground*: Team members must have shared beliefs about the world state, the goals, the conventions associated with the task at hand, etc.
- *Interpredictability*: team members must be capable of predicting each other's actions with a reasonable degree of accuracy.
- *Directability*: Team members must be capable of redirecting each other's behaviour.

By creating functionalities and behaviour capabilities for artificial team members that support each of these requirements, artificial agents can contribute to team processes in service of optimizing team performance. The following subsections offer insight in the way this can be achieved.

3.1 Common ground

In order for communication to be accurate and effective, participants in a conversation must establish proper common ground. Common ground refers to information which is mutually believed by all parties involved in a conversation [27]. Conversational efficiency is greatly enhanced through common ground as speakers need not explain all assumptions underlying their statements. Yet conversely, conversational effectiveness is greatly limited when speakers assume common ground, where it is lacking. Relying as a speaker on common ground with the audience is an invitation to cooperate [27] as the speaker appeals to the audience "to base their inferences not on just any knowledge or beliefs they may have, but only on the mutual knowledge or beliefs [shared by the conversational participants]" [10].

According to Stalnaker (2002), common ground relies on the speaker's presuppositions about the audience's common beliefs [53]. The common beliefs of the parties to a conversation are the beliefs they share, and that they recognize they share. By presupposing certain beliefs, the speaker takes the audience's understanding of such beliefs for granted as background information. By identifying the common ground of a conversation with the common belief of the participants, the presuppositions of an individual speaker can be identified with what the speaker believes to be common belief.

For HATs to establish and maintain common ground and thereby speed up the communication process, the following functionalities can be introduced in artificial team members.

3.1.1 Shared situation awareness

An important challenge in developing HATs is to endow artificial team members with functionalities allowing them to effectively and efficiently share situation awareness (SA) within the team [19]. SA primarily refers to knowledge about the current state of the task environment, as well as team activities, team performance, and overall progression with respect to the team task. Such knowledge facilitates coordination and reallocation of tasks within the team, but can also be used for effective and efficient communication among the team members. SA can be used to explain dynamic goal selection, attention to appropriate critical cues, and future state predictions, but this requires for the parties involved to have access to the information needed to assess –and, through that, become aware of– the situation. Establishing shared situation awareness, then, means for team members to reason about the necessity of sharing certain information with the other members in the team, to develop a shared team "theory of the situation" [8]. To realize this, team members should decide to share relevant information, while withholding irrelevant information, so as to prevent their team members from becoming overloaded with information [46]. Deciding on the necessity of information for the other team members, requires an additional type of awareness, dubbed "intention awareness" by Howard and Cambria (2013), by which they mean "the process of integrating actors' intentions into a unified view of the surrounding environment" [31]. Intention awareness can also be used to reason about potential adversaries and other actors

3.1.2 Explanations

At times throughout a conversation, team members may fail to understand one another's actions or utterances. On such occasions, team members may require for additional explanations to gain understanding of the other team member's meaning, motivation, intention, or assumption. When creating artificial team members, developing functionalities enabling them to explain their behaviour, recommendation, or expression is often referred to as "explainable artificial intelligence (XAI)". XAI facilitates the disclosure of information by artificial team members to human team members. Compared to the sharing of situation awareness, XAI typically concerns more advanced functionalities, as it requires for the intelligent system to trace its internal line of reasoning, inference, classification, or input-output mapping, depending on the type of technology used. When done appropriately, XAI helps human team members understand the system's rationale underlying its behaviour and/or decisions. For example, an analyst who receives recommendations from a smart decision support system needs to understand why the algorithm has recommended a certain course of action.

As of recent, the research area of explainable AI (XAI) has exploded (e.g. [28]). In addition to the ability to offer a meaningful explanation for a specific human actor when needed, XAI also refers to the ability to ask for an interpretable explanation from a specific human actor when needed. So far, research centred primarily on the first type of ability. However, more integrative methods are under development, including bottom-up data-driven (perceptual) processes and top-down model-based (cognitive) processes [42]. Such methods could help assess the trustworthiness of the autonomous system's task performance and, subsequently, explain the foundation and reasons of this performance to establish trust calibration [58].

3.2 Interpredictability

Smooth collaboration can be enhanced by team members anticipating the interdependencies within the team. By predicting team members' task performance, team members can shorten waiting times, expedite task performance by aiding a team member in need of assistance, provide important just-in-time information to those who need it, and plan for contingencies if a team member may be unavailable or incapable of performing a task. To enhance interpredictability between team members, endowing an artificial team member with the following functionalities can be useful.

3.2.1 HAT training and longitudinal teaming

Team performance requires coordination between activities of each of the team members, under routine conditions as well as under novel conditions. Procedures, protocols, and doctrines are all artefacts created to foster team performance, as they increase interpredictability and support coordination. Procedural team training therefore focuses on the internalisation of procedures. However, procedures are often insufficient when the actual task deviates from the training task. On such occasions, teams have to improvise new ways of working together. One way of dealing with this challenge, is through team cross-training: team members switch roles with one another, so as to understand one another's roles and responsibilities [10]. Nikolaidis et al. (2015) investigated the use of cross-training in human-robot teams for assembly in the manufacturing process [43]. By iteratively switching roles between the robot and the human worker, the robot learned a model of human behaviour, describing the sequence of actions necessary for task completion and matching the preferences of the human worker.

Team training (both procedural and cross-training) aims to support the development of *shared mental models*: an overlapping understanding of one another's objectives, roles, tasks, activities, whereabouts, team structure, and so on [63]. Shared mental models enable team members to reason not only about their own situation, but also about that of their team members in the pursuit of their joint goal. Shared mental models, in other words, enable team members to predict a team member's performance on a particular task, potential need for help or information, risks of failure within the team, etc.

HAT training supports the development of shared mental models, by learning about one another's capabilities and limitations through experience, and optimizing the team processes needed to mitigate risks and limitations within the team. By endowing artificial team members with the functionalities required to engage in HAT training, artificial team members can learn to optimize their behaviour with respect to a team, i.e. the constellation of specific team members and their particular dynamics. In other words, team training facilitates the development of increasingly accurate mental models of team members, with the objective of increasing the team's ability to flexibly coordinate their activities and increase team performance. This, however, requires for artificial team members to be able to learn from team interactions, updating their models to better interpret the tasks, behaviour, and corresponding needs of their team members. This is also referred to in the literature as "interactive shaping" [38][43]. For example, Nikolaidis et al. (2015) found that cross-training improved mental model similarity, as well as the human worker's perceived robot performance and trust in the robot.

3.2.2 Proactive communication

Teamwork often entails the processing, interpretation, and analysis of large amounts of information. Based on their roles and/or expertise, responsibility for handling certain information sources and types is distributed across the various team members. Oftentimes, the results of the information processed by one of the team members are relevant to the activities of another team member, requiring the team members to communicate with one another. Communication in teams often aims to contribute to one of the following: (1) problem-solving, (2) structuring and coordination, (3) socio-emotional alignment, or (4) proactive communication [33].

An important challenge in teams is how team members decide to *proactively communicate*: when should an actor communicate what with whom. Such decisions are often based on shared mental models [63]. Interpredictability (facilitated by mental models) enables an agent to infer that its team member is currently working on a task that requires certain information, or that newly retrieved information affects the decisions and tasks of a team member. Based on such reasoning, the agent may decide to proactively share its knowledge with the respective team mate. Proactive communication entails team members providing one another with information on their own accord, i.e. without the need for a team member to explicitly request for that information to be shared. We distinguish between proactive communication to accommodate a team member's information need based on: (1) that team member's preferences (e.g. as learned from prior information requests), (2) knowledge about that team member's situation (i.e. based on a shared mental model), and (3) one's own potential need for assistance in the near future, requiring the envisioned assistant to be up to date with the situation at hand. There might be additional reasons and situations where proactive communication advances team performance, that we currently haven't thought of yet.

For an artificial agent to be able to reason about its *human* team members, requires for that agent to be *human-aware*. Human-awareness entails that intelligent actor(s) have access to information about human team members and their characteristics (e.g. preferences, tasks, capabilities and limitations, etc.). In addition to this information, the agent(s) also employ various functionalities aiding the maintenance of, reasoning about, and learning from such information. For example, functionalities related to human-aware computing include location monitoring, attention tracking, and trust calibration. Ultimately, human-awareness enables intelligent systems to predict the behaviour of human team members, fostering better shared mental models and interpredictability.

3.3 Directability

Directability entails the ability of team members to influence and/or control one another's behaviour, to accommodate adaptations in the team's activities, coordination, behaviour, and overall performance. Traditionally, in human-robot collaboration, directability entailed tasking of a robot by a human operator. However, as robots (and agents) become more capable of determining their own plans and activities to try and accomplish the team goals, human-robot collaboration gradually moves away from traditional tasking, and towards e.g. dynamic task allocation, shared initiative, and work agreements.

3.3.1 HAT communication

Being able to influence the behaviour of another actor first requires for the team members to be able to express themselves about more than simple information sharing, as was the case for common ground and interpredictability. We therefore need to extend communication to include speech acts that express the desire for another agent to perform a certain activity. In most agent communication languages, these speech acts are derived from "request" and "query" [23]. The request speech act asks another agent to perform a certain task or action, whereas the query speech act asks another agent to provide particular information. We will go deeper into the use of speech acts in the next section.

3.3.2 Work agreements

Work agreements are used to impose constraints on the autonomous behaviour of agents [16]. This forms one of the core building blocks of human-agent teams as they enable external directability on an agent's behaviour without compromising the agent's autonomy (also known as internal autonomy requirement [18][56]). Furthermore, they contribute to maintaining common ground and interpredictability, by enabling an explicit way to specify shared conventions on the agent's behaviour.

Work agreements are very similar to policies which have been applied as a teamwork coordination mechanism in [8]. We distinguish between two types of work agreements: obligations which describe which actions must be performed by an agent in a given context; prohibitions which describe which actions are not permitted to be performed by an agent in a given context. Examples of work agreements are

- UGV 2 has an obligation to notify the human worker when it detects a potentially hostile target
- UAV 3 has a prohibition to fly above the village

Work agreements can be applied for various purposes on different time scales. For example, one might specify a work agreement that specifies the current plan that is followed by the human and agent. On a longer timescale, a work agreement can be used to specify the Rules of Engagement which the system has to adhere to, or even military doctrine which lasts for the entire lifecycle of the system.

Work agreements can also be applied when designing the teamwork process itself by stating which communicative acts must be performed under which circumstances. For example, to specify the task division between the human and the agent, or to specify the level of human involvement in the process.

A more detailed (technical) description of work agreements is provided in the next section.

3.3.3 Dynamic task allocation and fit-for-purpose collaboration

As artificial team mates become more self-sufficient, it is at times unnecessary for a human to control, or even monitor, artificial agents at all times. The challenge is, however, that most artificial agents are capable of performing their task without the need for assistance or control under specific circumstances, whereas at other times, their performance degrades, or they malfunction altogether. As a result, the human-agent collaborative work relationship may vary across situations. To deal with this phenomenon, the team should be able to engage in dynamic task allocation and fit-for-purpose collaboration [25]. For example, the team should be able to shift between the following work relationships depending on the situation at hand:

- Parallel task performance: team members perform their tasks in coordination with their team members. They are capable of identifying, organising, and performing their own tasks and responsibilities without the need for assistance.

- Management by exception: team members can perform independently, yet when help or directions are required, they either ask for help from their supervisors or colleagues, or their team members notice a break-down and offer help proactively.
- Training / educating: team members are still in training and require constant feedback and monitoring so as to strengthen their understanding of the team task or goal, their team members, the roles within the team, and their own activities as part of the entire team performance.
- Tasking: team members are capable of performing a task once it is provided to them, but are incapable of determining their next task, as they are unaware of the encompassing team task or goal, their surroundings, role within the team, and/or team members.

4.0 SYSTEM ARCHITECTURE DESIGN FOR HUMAN-AGENT TEAMING

This section describes a functional architecture, called SAIL, in which the common HAT functions can be combined and configured to turn a set of autonomous agents and humans into a coherently working team. The next subsection describes the SAIL architecture

4.1 Social Artificial Intelligence Layer

SAIL (Social Artificial Intelligence Layer) is an environment in which HAT functionalities can be implemented in a modular way, i.e. using HAT modules. We distinguish between three types of components in a SAIL system:

- **Humans** in their ambient environment. For example, these may be professionals working in crisis management wearing mobile interaction devices such as smart watches and head-mounted displays supporting augmented reality; or these may be operators working in a control station with large information displays.

- **TAI** (task-oriented AI) components, i.e. technical AI components designed to optimally perform a certain task, but which may not be optimized for human interaction. These may be robots conducting surveillance in a certain area, cyber agents protecting vital ICT infrastructure against cyber threats, etc.

- **SAI** (Social AI) modules which serve as intelligent middleware aiming to transform task-oriented AI components and humans into a coherent human-agent team. Such components include machine learning technology that can decide how to exchange the right information at the right moment among the right actors [17], or AI message interpretation that can translate a high level command such as “secure the area” into commands that can be processed by the TAI component.

An overview of a SAIL configuration is depicted below.

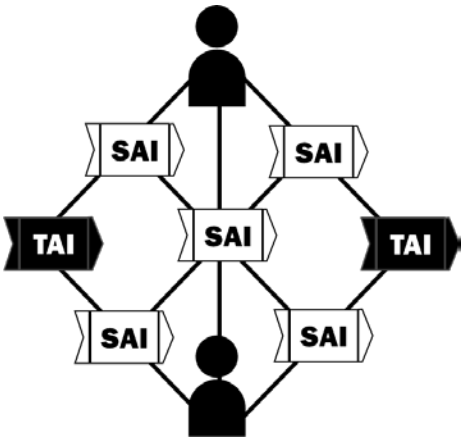


Figure 4-1 SAIL system architecture, where Task-oriented AI components (TAI) are mediated by Social AI Components (SAI) to allow interaction with humans.

Because the different components in SAIL can take on many forms depending on the requirements of the system, SAIL does not impose any constraints on their internal workings. The focus of SAIL therefore is on the interactions between the different components. These interactions are specified in a dedicated language called HATCL (Human-Agent Team Communication Language), which provides an abstract specification language of the possible information and control flows between the different components of a HAT. HATCL is abstract in the sense that it specifies the *information* in the message and its *illocutionary force* [50]. Illocutionary force refers to the intention of a speaker behind an utterance, e.g. obtaining information, directing, etc. HATCL messages are neutral with respect to interaction modality or graphical representation. A specification of the language is described in the next subsection.

4.2 HATCL

HATCL is inspired by FIPA-ACL, which has been developed as a language for communication between intelligent software agents [23]. To function as an effective means of communication between the different components in Figure 4-1, HATCL should satisfy a number of additional requirements which makes it suitable for Human-Agent Teaming.

R1: HATCL should be capable of capturing the different abstraction levels at which machines and humans interpret and process information within a certain problem domain.

HATCL should mediate between the different levels of abstraction of the various SAIL components. In the interaction between humans and SAI modules, a HATCL message should be capable of representing the information that stems from human input or that is outputted to humans. In other words, the language should (at least partly) be understandable by humans. This means that the language should align with human mental models [42]. One way to interpret this is to adopt Daniel Dennett's *intentional stance* [12]. This theory states that humans should comprehend the behaviour of artificial agents by attributing *goals*, *beliefs*, and *intentions* to them. The communication of such concepts should be facilitated by the HATCL language. In the interaction between TAI and SAI modules, the language should (at least partly) be understandable for machines. Therefore, the conceptualization should be alignable with the internal logic of the system. For example, when the system is based on internal logic that chooses its *actions* based on maximizing the expected *reward*, the language should be capable of communicating these concepts to be understandable for the system.

R2: HATCL should allow for sending soft directives.

To implement directability in a team of equal humans and agents, the rule *orders are orders* does not always apply. For example, consider the situation where the human commands an autonomous drone to fly back to the command post (CP). Various reasons exist in which the drone's responsive behaviour might deviate from simply following this command:

- **Incapability:** The drone might know that it is currently incapable of completing the action due to a reason unknown to the human (e.g. low battery level). In this case, we might want the drone to point out this problem to the human, rather than trying and failing.
- **Conflicting with team goal:** The drone might possess additional situation awareness that the given command would lead to an outcome which conflicts with the team goal. For example, that flying back to the CP would cause the drone to be subjected to hostile fire, which conflicts with the team goal to remain safe. If the drone can think of a better alternative, we might want the drone to propose this alternative instead.
- **Dealing with multiple orders:** In practical situations, the drone must deal with multiple orders which apply to different timescales and may even be in conflict with each other. For example, when the drone receives the order to fly to the CP, the drone may have to decide to finish its previous order first (e.g. to take a picture of an area of interest) or fly back immediately. Most likely, this depends on the time it takes to finish the previous order. In case of uncertainty, we might want the drone to discuss this problem with the human.

As the examples above illustrates, a directive used for communication between autonomous agents is fundamentally different than a directive used between objects in object-oriented programming language (e.g. remote method invocation). Jennings et al. have famously phrased this as: *Objects do it for free, agents do it for money* [36]. Nevertheless, defining the precise meaning of these soft directives remains a challenge, and is one of the major objectives of HATCL specifications.

R3: HATCL should allow for specifying unambiguous work agreements.

As argued in Section 3.3.1, work agreements form an essential part of HAT technology. Therefore, the specification of these work agreements in an unambiguous way is one of the main purposes of HATCL (also see Section 4.2.2).

R4: HATCL should allow symbol grounding in various system architectures.

The symbol grounding problem [15] refers to the problem of relating symbolic messages (such as HATCL messages) to internal structures that are processed by the agent (such as perceptions and plans). We use semantic anchors to formalize this relation. Different system architectures typically require different semantic anchors for the same HATCL message. For example, opaque deep learning networks require very different types of semantic anchors than rule-based systems.

To satisfy each of the requirements described above, we define the message structure (syntax) and its semantics (in terms of work agreements, ontologies, and semantic anchors).

4.2.1 The HATCL Message syntax

A HATCL message has the following structure: <Performative, Sender, Receiver, In-reply-to, Content, Protocol, Ontology, Message-ID, Conversation-ID>. Most of these fields, such as Sender, Receiver, Message-ID, contain meta-information used for routing the message. The fields Performative, Content, and Ontology are worthy of further explanation and are discussed below.

The *Performative* is used to denote the illocutionary force of a message, which could be:

- *Inform*: Provide another actor with information

- *Query*: Ask another actor for information
- *Subscribe*: Subscribe to information updates on a specific topic from another actor
- *Request*: Ask another actor to perform a certain task (acts as a single purpose work agreement)
- *Propose*: Propose a work agreement to another actor
- *Accept*: Accept the proposed work agreement
- *Reject*: Reject the proposed work agreement
- *Understood*: Acknowledge reception, and correct interpretation, of an inform message
- *Not understood*: Acknowledge reception, yet misinterpretation, of an inform message
- *Cancel*: Cancel a previously instantiated work agreement

This is the set of performatives currently included in HATCL. New performatives may be added as required by future applications. Each of these performatives has been defined both syntactically and semantically in our HATCL specification document.

The *content* of a message specifies what is actually communicated and can be specified in a query language, working agreement language, or assertion language. The content

```
{ “Performative” :      “Query” ,
  “Sender” :            “Hum1”,
  “Receiver” :          “UGV1”,
  “In-reply-to” :       “”,
  “Content” :           “$.vehicles.*”
  “Protocol” :          “”,
  “Ontology” :          “military_ont”,
  “Message-ID” :        “msg13”,
  “Conversation-ID :    “cnv-2” }
```

4.2.2 Work agreements

As work agreements between humans and autonomous systems impose well specified constraints on autonomous behaviour, they form the core building blocks of HAT technology; many message types in HATCL can be interpreted in terms of work agreements. For example, a HATCL message of the type *inform* is translated into the work agreement: <$Actor_1$, $Actor_2$, *upon receiving this message, O(send information of type x to* $Actor_1$*)*>, which states that by accepting this work agreement, $Actor_2$ commits to an obligation to immediately send information of type x to $Actor_1$.

Scientific research on the formalization of work agreements (also referred to as “social commitments”) comes from the field of normative multi-agent systems [40][51]. A work agreement is an explicit agreement between two actors, specifying that one actor, denoted as *debtor*, owes it to another actor, denoted as *creditor*, to effectuate some *consequent* (e.g., refrain from or see to it that some action is performed or some objective is achieved) if the *antecedent* (e.g., some precondition) is valid [51]. Work agreements, in short, aim to specify permissions and obligations on agent behaviours. And so allow for the voluntary restriction of an actor’s autonomy as proposed by another actor. Work agreements hold explicitly between two actors. Therefore, work agreements are sometimes compared to contracts. An example of a work agreement is: “Lawrence is obligated to notify Lisa about his change in intent, if he decides to pause his current task to switch to a more pressing task encountered along the way”.

A work agreement is first and foremost a *voluntary* restriction on an actor's autonomy (also see Figure 4-2), as the actor receiving the *proposed* work agreement is also allowed to reject the work agreement. The *acceptance* of the work agreement, and hence the restriction on its autonomy, is completely voluntary. As soon as the debtor has accepted the work agreement, though, the debtor's autonomy is *conditionally* restricted: the debtor must satisfy the work agreement once the *antecedent* becomes valid. If the debtor fails to provide the consequent of an activated agreement before the deadline, this implies that the agreement has been violated. If the debtor succeeds to do so, the work agreement is satisfied. Furthermore, commitments can, in general, be cancelled by the actors involved (although this may be subdue to overarching rules).

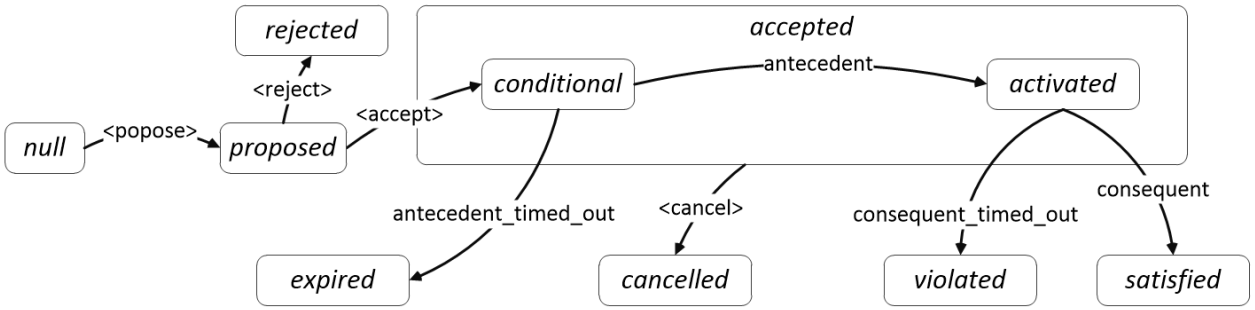


Figure 4-2: The lifecycle of a work agreement as affected by the dynamics between the actors and events taking place in the environment

A similar approach to work agreements are *policies*: “enforceable, well-specified constraints on the performance of a machine-executable action by a subject in a given situation” [16]. Policies come in two flavours: authorization policies, stating what is permitted, and obligation policies, stating what is obligatory in a given situation [16]. Policies do not hold between two actors, but instead are generally applicable to a set of actors. Therefore, policies are sometimes compared to laws. An example of a policy is: “Team members are obligated to notify their team leader about a change in intent if they decide to switch tasks after encountering a more pressing task”.

The foundation of both policies and work agreements are normative rules in the form of deontic logic [59]. Deontic logic is used to reason about obligations and permissions. Therefore, deontic logic still forms the core of work agreements and policies, as it enables reasoning and verification.

4.2.3 Ontologies

Ontologies offer explicit, structured, and semantically rich representations of declarative knowledge. They consist of concepts (`classes'), and relations between them, to describe certain parts of the world [45]. HATCL uses a domain-independent “top ontology” and a domain-specific ontology to enable actors in the HAT to parse the messages they receive. Figure 4-3 shows the top ontology, consisting of relatively generic concepts, such as Actor, Plan, Goal, and Action. The domain-specific lower-level ontology would provide specific instances of tasks, actors, and plans particular to that domain.

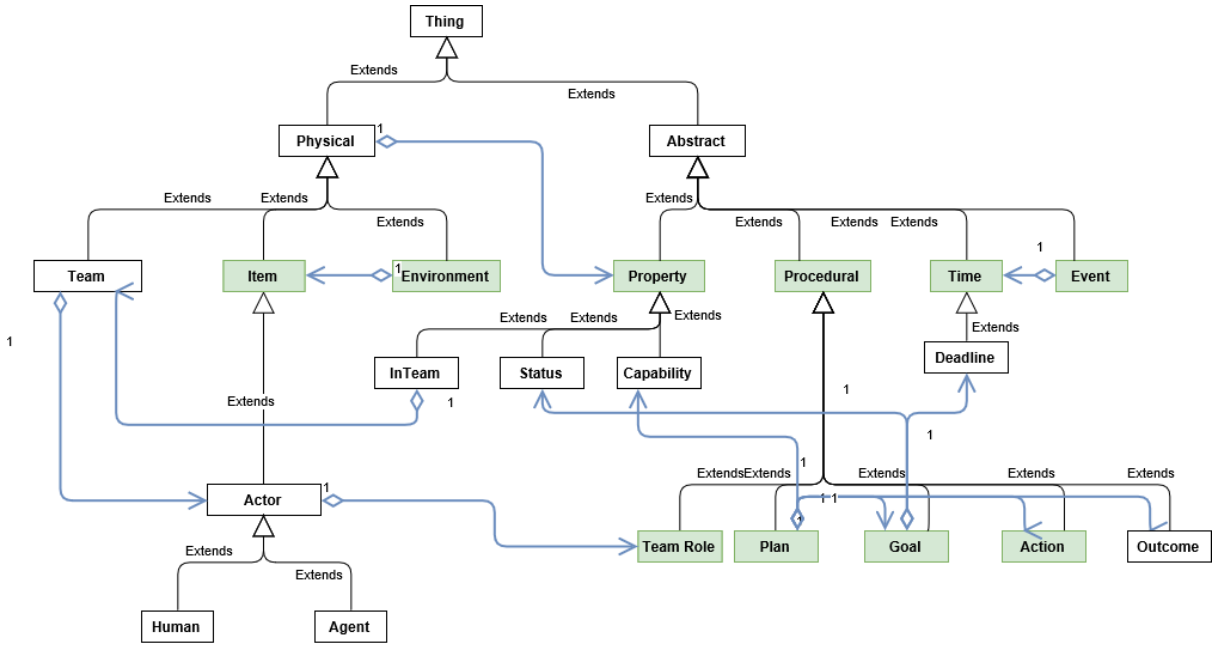


Figure 4-3: Top-level domain-independent ontology used to specify work agreements and HATCL messages, facilitating coordination and communication between actors in the HAT

4.2.4 Semantic Anchors

Semantic anchors relate concepts and variables in HATCL to concepts and variables in the autonomous system. This translation operation is implemented inside the autonomous system. To illustrate this important principle in human machine teaming, we will start with a simple example.

Example 1: Suppose we have a human team-member controlling an autonomous system which runs the following code:

```
Repeat
       turn_left,
       turn_right,
       move_straight
Until false
```

Suppose the human teammember expresses the desire to prohibit left turns via work agreements. Right now, the autonomous system code is not suitable for that. Therefore, the developer makes the following code-update:

```
Repeat
       If tl_permitted then turn_left,
       turn_right,
       move_straight
Until false
```

Furthermore, a semantic anchor is created that maps the WA (specified in HATCL) to code that can be interpreted by the autonomous system:

```
prohibited(turn_left) -> tl_permitted=false
```

Example 2: Assume now that the autonomous system's actions do not have the attribute `tl_permitted`. Instead, the system has a variable map indicating the desirability of each position on the ground, which is used to determine where to move. In this case, the anchor would access the variable map and set a low desirability on the coordinates on the system's left side.

As this example shows, multiple ways exist in which a semantic anchor can be realized. In general, the human interfaces work at a higher abstraction level than the internal control logic of the autonomous system. This means that semantic anchors whose information flows from the teaming software towards the system perform a translation into a lower abstraction level. Whereas the translation into a higher abstraction level occurs when the information flow is reversed. Note that this does not need to be the case nor are semantic anchors limited to a one direction flow of information. See the figure below for two examples of the functionality of semantic anchors. The top anchor simply acts as a gateway between an abstract variable in the teaming software and the system, as the abstract meaning is appropriate in both (see the example 1). Whereas the bottom anchor performs an actual translate operation and whose information flow is bidirectional (see example 2). Also note the presence of the API as a distinct entity.

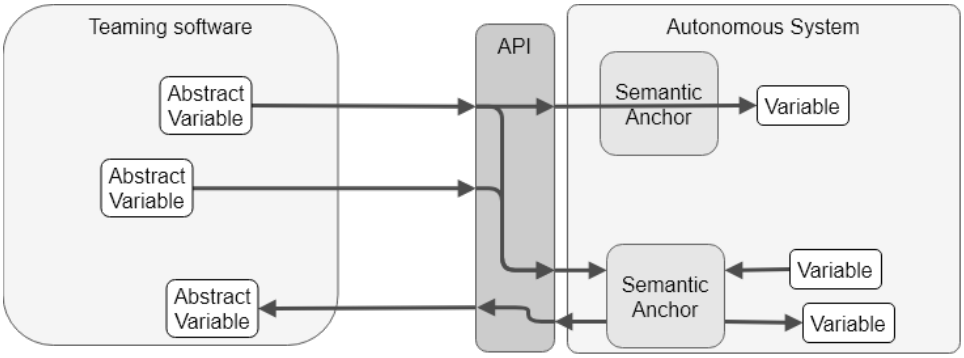


Figure 4-4: Semantic anchors describe how abstract variables in the HATCL ontology are grounded in the variables available in the autonomous system

Note that semantic anchoring can become very difficult or even impossible depending on the representations used. HATCL is based on the symbolic representation paradigm where each element corresponds to one entity. Neural Networks, which are widely used in AI applications such as autonomous driving and image classification, are based on distributed representations [29]. In these representations, each entity is represented by a pattern of activity distributed over many elements. How to map HATCL to distributed representations in a neural net remains an open question.

4.2.5 Software implementation of SAIL

SAIL is implemented using an open source distributed application framework, called Akka, which is used as a software wrapper around the various pieces of code, making it a coherent human-agent system. The SAIL components can be programmed in any language, and may run on any type of hardware (e.g. robots, head-mounted displays, mobile devices, sensors).

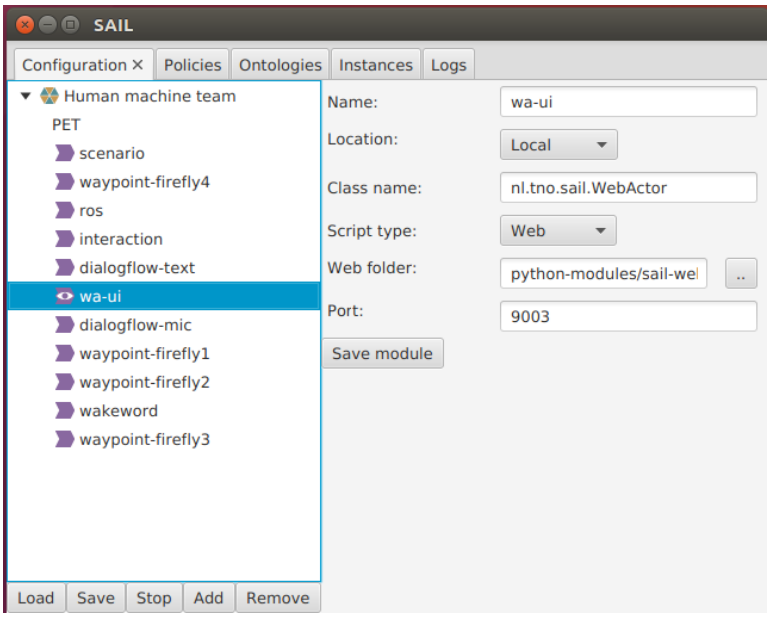


Figure 4-5 SAIL development environment allowing the HAT engineer to add extra SAI modules and configure the system's set of fixed working agreements and ontologies.

The figure above shows the Sail development environment. The different SAIL components can be added to the tree on the left-hand side of the window. The configuration of these components is done on the right-hand side. Other tabs can be used to specify the policies (working agreements), and ontologies that are shared within this system.

5.0 PROTOTYPE OF A HAT APPLICATION WITHIN SAIL

To demonstrate the application of the techniques discussed in the previous section, we used a case for aerial surveillance of a compound as described in Section 2.2.1. The proof of concept implements an initial subset of the proposed common HAT functions and shows that the underlying concepts and SAIL architecture translate into a viable HAT set-up. The surveillance scenario is recreated as a virtual environment. This simulation (implemented in Gazebo [35]) includes a 3D modelled military compound, a variety of potential threats in the vicinity of this compound as well as a swarm of UAVs that survey the surrounding area.

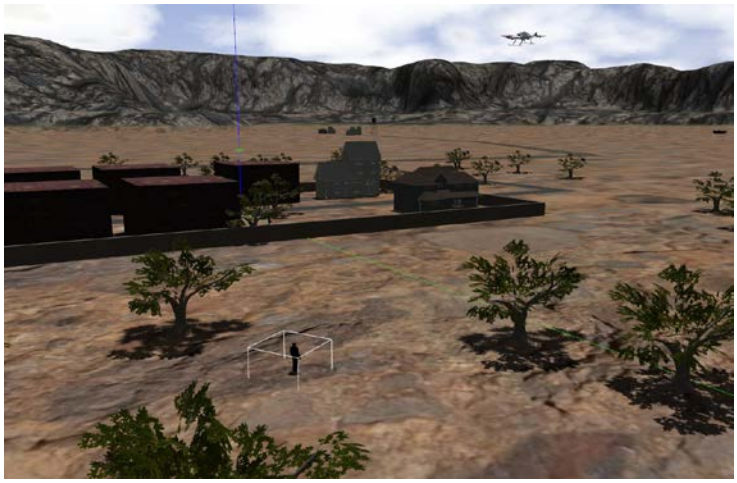


Figure 5-1 the Gazebo simulation environment, which provides the backbone of the prototype but which is typically not visible as such to the base commander.

Task-related functionality of the UAVs is implemented in TAI modules that cover capabilities such as waypoint navigation, path planning, object detection, video streaming, and threat identification. Using its set of TAI modules, each UAV is capable of autonomously scanning the surroundings of the military compound for suspicious activities.

To turn the UAV's into teammates, we added a Social Artificial Intelligence Layer (SAIL). This layer includes a set of SAI modules, including a multi-modal user interface, and offers the infrastructure to provide the middleware between the task-oriented modules and the human team members. In the proof of concept, we have implemented and combined a number of core HAT functions (as discussed in Section 3), namely those for shared situational awareness, proactive communication (ProCom), human-aware computing, HAT communication, and the work agreement mechanism.

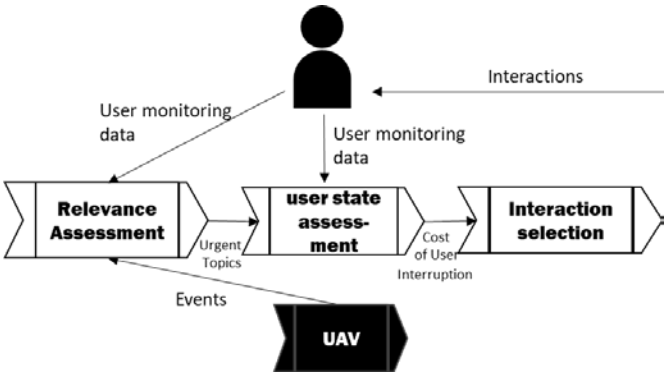


Figure 5-2 An overview of the main SAI modules within the base protection HAT.

The figure above shows the SAI modules that are related to proactive communication (ProCom). ProCom aims to establish a balance between the value of sharing information and the costs of imposing cognitive workload on the human team member. The autonomous system (the UAV) performs the task of aerial surveillance. Using a semantic anchor, events are published using HATCL and made available to the SAI module relevance assessment. This module regards each event as a topic and assesses its relevance to the human team member.

For example, topics about the detection of a potential hostile contact are more relevant than topics about a friendly civilian contact. Also, when the human team member has explicitly asked for a certain topic, this topic has high relevance. The SAI module "user state assessment" builds up a value that indicates the cognitive task load of the human team member (how busy he/she is) and the situational awareness (what the human team member currently does and does not know). Based on this information, the SAI module interaction selection decides whether that information should be communicated to the human team member or not and in which way (e.g. using a textual message or using a voice-message). More information on content-based modality selection can be found in earlier work [17].

The combined SAI functionality manifests itself to the human team member in the form of an avatar. The avatar is able to maintain a dialogue with the human team member and act as an intuitive interface between him/her and the SAI components of the autonomous agents (e.g. UAVs). In addition the avatar acts as an intelligent information retrieval system, capable of accessing various information resources within the available SAIL modules. The avatar responds to speech, typed chat messages, and touch input from the human team member. Its actions can vary from information retrieval, engaging conversational dialogues with the UAVs, establishing work agreements with these SAI-plugged agents and displaying (task-related) content tiles on one of the available computer screens.

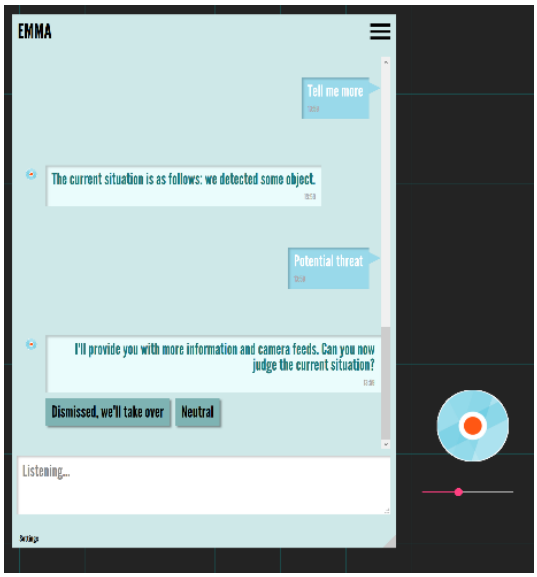


Figure 5-3 The avatar representing the UAV swarm and the dialogue window enabling interaction.

The figure above shows the avatar (the round circle on the right), and the dialogue window which interacts in a similar way as Google Allo (https://allo.google.com/). Natural language input can be inputted via speech or command buttons, and a history of the conversation is shown using text balloons. Besides replying to the human team member's input via text messages, the system can also open up additional windows to visualise information and provide new ways of interaction, such as maps, camera feeds of the robots, etc. Note that this can only be done if sufficient screen space is available.

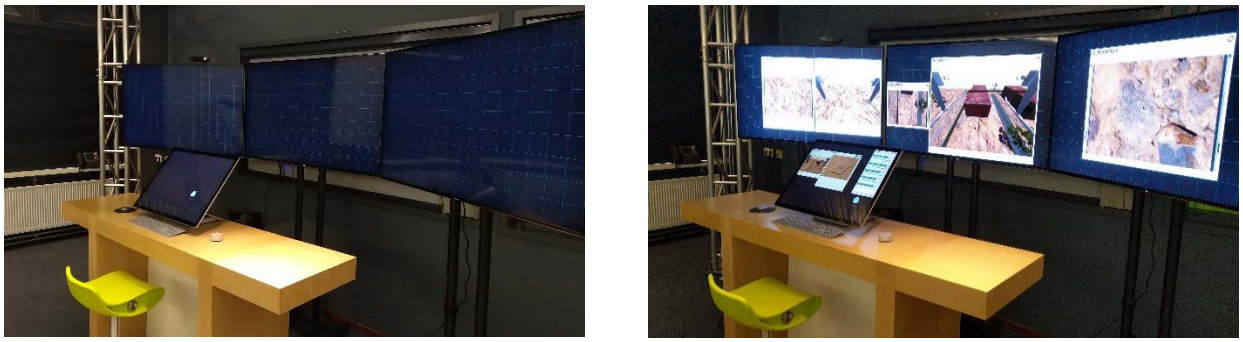


Figure 5-4 The demonstration environment of the base protection HAT. When the autonomous system is not encountering problems, the base commander is not distracted with any information (as on the left hand side). When problems occur, the base commander can drill down on every piece of relevant task information.

The figure above shows the set-up of the demo, consisting of three large information displays, a MS surface pro touch interaction device, microphone, and speakers. When the UAVs are functioning normally, the screens show as little information as possible, and only the avatar is visible (see the right hand side of the Figure). When an important topic arises, such as a suspicious contact, more and more information is exchanged between the UAV swarm and the base commander. The right hand side of the figure shows the situation in which the individual camera-feeds of the UAVs are shown to the base commander allowing the UAV and base commander to make a joint decision on how to classify the object. After the issue is solved, the topic becomes irrelevant again which implies that all camera-feeds close, and the interface returns to its calm state again.

The application described above illustrates a management-by-exception type of HAT. This means that the base commander is not bothered with superfluous information when this is not needed. However, upon request, or by system initiative, a rich interaction can be set up which drills down to the details of the matter at hand.

6.0 CONCLUSION

Human agent teaming is a problem with many faces. This paper is an attempt to combine a human factors, engineering, and military perspective on the issue. Our solution is based on the idea of a Social Artificial Intelligence Layer (SAIL), which is a framework for the development of HMT-concepts. The starting point of SAIL is that a HAT can be developed without changing the internal capabilities of the autonomous system. We have argued that these social capabilities are to some extent generic. Examples are functions for situation awareness, human awareness, explainable AI, working agreements and tasking. Within SAIL, HAT-modules are developed that construct these social capabilities. The modules are reusable in multiple domains.

We have demonstrated the use of SAIL by building an application for military compound protection using surveillance drones. The surveillance drones where simulated using the robot simulation environment Gazebo, and SAIL was used to build a teaming layer on top of it. The approach resulted in a system which embraces a management by exception type of HAT: no information reaches the human teammember, unless a problem arises which requires negotiation.

We believe that our approach is promising and an important step towards developing and prototyping human agent teaming applications in the defence domain. We identify three directions for future work. Firstly, we intend to develop more complex types of semantic anchors, and explore which autonomous system architectures enable which types of anchors. This allows us to discover the boundaries of HAT applications, as only those types of information can be communicated which can actually be anchored in the agent. Secondly, we intend to explore validation methods for HAT applications, which also take the long term aspects

of teaming into account. Thirdly, we intend to explore combinations of HAT interaction with more immersive interaction techniques such as tele-presence.

7.0 REFERENCES

[1] Albus, J. S., Huang, H. M., Messina, E. R., Murphy, K., Juberts, M., Lacaze, A., ... & Proctor, F. M. (2002). 4D/RCS Version 2.0: A reference model architecture for unmanned vehicle systems (No. NIST Interagency/Internal Report (NISTIR)-6910).

[2] Article36. (2015). Killing by machine - Key issues for understanding meaningful human control. Retrieved on July 30, 2018 from: http://www.article36.org/weapons-review/killing-by-machine-key-issues-for-understanding-meaningful-human-control/

[3] Article36. (2014). Key areas for debate on autonomous weapons systems. Retrieved on July 30, 2018 from http://www.article36.org/wp-content/uploads/2014/05/A36-CCW-May-2014.pdf

[4] Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. Journal of Human-Robot Interaction, 3(2), 74–99.

[5] Bethel, C. L., Carruth, D., & Garrison, T. (2012, November). Discoveries from integrating robots into SWAT team training exercises. In Safety, Security, and Rescue Robotics (SSRR), 2012 IEEE International Symposium on (pp. 1-8). IEEE.

[6] Bolman, L. (1979). Aviation accidents and the "theory of the situation." In G. E. Cooper, M. D. White, and J. K. Lauber (Eds.), Resource management on the flight deck: Proceedings of a NASA/industry workshop (pp. 31-58). Moffett Field, CA: NASA Ames Research Center.

[7] Boulanin, V., & Verbruggen, M. (2017). Mapping the development of autonomy in weapon systems. Stockholm International Peace Research Institute (SIPRI). Retrieved on July 30, 2018 from: https://www.sipri.org/publications/2017/other-publications/mapping-development-autonomy-weapon-systems

[8] Bradshaw, J. M., Feltovich, P. J., Johnson, M. J., Bunch, L., Breedy, M. R., Eskridge, T., ... & Uszok, A. (2008, May). Coordination in human-agent-robot teamwork. In Collaborative Technologies and Systems, 2008. CTS 2008. International Symposium on (pp. 467-476). IEEE.

[9] Campaign to Stop Killer Robots, "The Problem," retrieved on July 24, 2018 from: http://www.stopkillerrobots.org/the-problem/

[10] Cannon-Bowers JA, Salas E, Blickensderfer E, et al. (1998) The impact of cross-training and workload on team functioning: A replication and extension of initial findings. Human Factors: The Journal of the Human Factors and Ergonomics Society 40: 92–101.

[11] Clark, H.H. (1992). Arenas of language use. Chicago: University of Chicago Press.
Coradeschi, S., Loutfi, A., & Wrede, B. (2013). A short review of symbol grounding in robotic and intelligent systems. KI-Künstliche Intelligenz, 27(2), 129-136.

[12] Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. Management science, 32(5), 554-571.

[13] Danzig, R. (2018). Technology roulette – managing loss of control as many militaries pursue technological superiority. Center for a New American Security. Retrieved on August 7, 2018 from: https://www.cnas.org/publications/reports/technology-roulette

[14] David, R. A., & Nielsen, P. (2016). Defense science board summer study on autonomy. Defense Science Board Washington United States. Retrieved on July 30, 2018 from: https://www.hsdl.org/?view&did=794641

[15] Dennett, D. C. (1989). The intentional stance. MIT press.

[16] van Diggelen, J., Bradshaw, J. M., Johnson, M., Uszok, A., & Feltovich, P. J. (2010). Implementing collective obligations in human-agent teams using KAoS policies. In Coordination, Organizations, Institutions and Norms in Agent Systems V (pp. 36-52). Springer, Berlin, Heidelberg.

[17] van Diggelen, J., Grootjen, M., Ubink, E. M., van Zomeren, M., & Smets, N. J. (2013). Content-based design and implementation of ambient intelligence applications. In Ambient Intelligence-Software and Applications (pp. 1-8). Springer, Heidelberg.

[18] Dignum, M. V. (2004). A model for organizational interaction: based on agents, founded in logic. SIKS.
[19] Docherty, B. L., Human Rights Watch (Organization), Harvard Law School, & International Human Rights Clinic. (2016). Making the case: the dangers of killer robots and the need for a preemptive ban. Retrieved on July 30, 2018 from: https://www.hrw.org/report/2016/12/09/making-case/dangers-killer-robots-and-need-preemptive-ban
[20] Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). In Aerospace and Electronics Conference, 1988. NAECON 1988., Proceedings of the IEEE 1988 National (pp. 789–795). IEEE.
[21] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1), 32–64.
[22] de Greef, T., Arciszewski, H., & van Delft, J. (2006, September). Adaptive Automation using an object-orientated task model. In: Proceedings MAST conference (pp. 4-6).
[23] Fipa, (2002). Fipa acl message structure specification. Foundation for Intelligent Physical Agents, http://www. fipa. org/specs/fipa00061/SC00061G. html (30.6. 2004).
[24] Gilson, L. L., Maynard, M. T., Jones Young, N. C., Vartiainen, M., & Hakonen, M. (2015). Virtual teams research: 10 years, 10 themes, and 10 opportunities. Journal of Management, 41(5), 1313-1337.
[25] Goodrich, M. A., McLain, T. W., Anderson, J. D., Sun, J., & Crandall, J. W. (n.d.). Managing Autonomy in Robot Teams: Observations from Four Experiments, 8.
[26] Gray, B. (1989). Collaborating: Finding common ground for multiparty problems.
[27] Grice, H.P. (1975). Logic and conversation. In E Cole & J. Morgan (Eds.), Syntax and semantics 3: Speech acts. New York: Academic Press.
[28] Guidotti, R., Monreale, A., Turini, F., Pedreschi, D., & Giannotti, F. (2018). A survey of methods for explaining black box models. arXiv preprint arXiv:1802.01933.
[29] Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1984). Distributed representations (pp. 1-127). Pittsburgh, PA: Carnegie-Mellon University.
[30] Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground?. Cognition, 59(1), 91-117.
[31] Howard, N., & Cambria, E. (2013). Intention awareness: improving upon situation awareness in human-centric environments. Human-Centric Computing and Information Sciences, 3(1), 9.
[32] Human Rights Watch (2016). Killer Robots and the Concept of Meaningful Human Control. Memorandum to Convention on Conventional Weapons (CCW) Delegates. Retrieved on July 30, 2018 from: https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control
[33] Kauffeld, S., & Lehmann-Willenbrock, N. (2012). Meetings matter: Effects of team meetings on team and organizational success. Small Group Research, 43(2), 130-158.
[34] Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground and coordination in joint activity. Organizational simulation, 53, 139-184.
[35] Koenig, N. P., & Howard, A. (2004, September). Design and use paradigms for Gazebo, an open-source multi-robot simulator. In IROS (Vol. 4, pp. 2149-2154).
[36] Jennings, N.R., Sycara, K., Wooldridge, M.: A roadmap of agent research and development. Autonomous Agents and Multi-Agent Systems 1 (1998) 275–306.
[37] Johnson, M., Bradshaw, J. M., Feltovich, P. J., Hoffman, R. R., Jonker, C., Van Riemsdijk, B., & Sierhuis, M. (2011). Beyond cooperative robotics: The central role of interdependence in coactive design. Intelligent Systems, 26(3), 81–88.
[38] Knox WB and Stone P (2013) Learning non-myopically from human-generated reward. In: Proceedings of the 2013 international conference on intelligent user interfaces, pp. 191–202.
[39] Mioch, T., Kroon, L., & Neerincx, M. A. (2017). Driver Readiness Model for Regulating the Transfer from Automation to Human Control (pp. 205–213). ACM Press.
[40] Mioch, T., Peeters, M. M. M., and Neerincx M. A. (2018) Improving Adaptive Human-Robot Cooperation through Work Agreements. In: International Conference on Robot and Human Interactive Communication (RO-MAN), Nanjing, China. IEEE.

[41] Murphy, R. & Shields, J. (2012). The Role of Autonomy in DoD Systems (No. 20301–3140). Washington D.C.: Defense Science Board. Retrieved on July 30, 2018 from: https://fas.org/irp/agency/dod/dsb/autonomy.pdf
[42] Neerincx, M. A., van der Waa, J., Kaptein, F., & van Diggelen, J. (2018, July). Using Perceptual and Cognitive Explanations for Enhanced Human-Agent Team Performance. In International Conference on Engineering Psychology and Cognitive Ergonomics (pp. 204-214). Springer, Cham.
[43] Nikolaidis, S., Lasota, P., Ramakrishnan, R., & Shah, J. (2015). Improved human–robot team performance through cross-training, an approach inspired by human team training practices. The International Journal of Robotics Research, 34(14), 1711-1730.
[44] Norman, D. A. (2014). Some observations on mental models. In Mental models (pp. 15-22). Psychology Press.
[45] Peeters, M. M. M., van den Bosch, K., Neerincx, M. A., & Meyer, J.-J. C. (2014). An ontology for automated scenario-based training. International Journal of Technology Enhanced Learning, 6(3), 195–211.
[46] Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation Awareness in Team Performance: Implications for Measurement and Training. Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1), 123–136.
[47] Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork? Small group research, 36(5), 555-599.
[48] Scharre, P., & Horowitz, M. C. (2018). Artificial Intelligence: What Every Policymaker Needs to Know. Center for a New American Security. Retrieved on July 30, 2018 from: https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know
[49] Schaub, G., & Kristoffersen, J. W. (2017). In On or Out of the Loop - Denmark and Autonomous Weapon Systems. Centre for military studies, University of Copenhagen. Retrieved on July 30, 2018 from: https://cms.polsci.ku.dk/english/publications/in-on-or-out-of-the-loop/In_On_or_Out_of_the_Loop.pdf
[50] Searle, J. R., (1969). Speech acts: An essay in the philosophy of language (Vol. 626). Cambridge university press.
[51] Singh, M. P. (1999). An Ontology for Commitments in Multiagent Systems: Toward. Artificial Intelligence and Law, 7, 97--113.
[52] Stanton, N. A., Salmon, P. M., Walker, G. H., Salas, E., & Hancock, P. A. (2017). State-of-science: situation awareness in individuals, teams and systems. Ergonomics, 60(4), 449-466.
[53] Stalnaker, R. (2002). Common ground. Linguistics and philosophy, 25(5-6), 701-721.
[54] Sundstrom, E., De Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. American psychologist, 45(2), 120.
[55] Vagia, M., Transeth, A., and Fjerdingen, S. (2016) 'A Literature Review on the Levels of Automation During the Years. What are the Different Taxonomies that have been Proposed?', Applied Ergonomics, 53, pp. 190-202.
[56] van der Vecht, B., Dignum, F., Meyer, J.-J. C., & Neef, M. (2008). A dynamic coordination mechanism using adjustable autonomy. In Coordination, Organizations, Institutions, and Norms in Agent Systems III (pp. 83–96). Springer.
[57] de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation' to 'autonomy': the importance of trust repair in human–machine interaction. Ergonomics, 1-19.
[58] de Visser, E. J., Peeters, M. M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., and Neerincx, M. A. (under review) Longitudinal Trust Development in Human-Robot Teams: Models, Methods and a Research Agenda. IEEE Transactions on Human-Machine Systems, special issue.
[59] Wieringa, R. J., & Meyer, J.-J. C. (1993). Applications of deontic logic in computer science: A concise overview. Deontic Logic in Computer Science, 17–40.
[60] Williams, A. P., & Scharre, P. D. (2015) 'Autonomous Systems – Issues for Defence Policymakers'. Retrieved on July 30, 2018 from: www.dtic.mil/docs/citations/AD1010077. NATO ACT.

[61] Woods, D. D., & Cook, R. I. (2002). Nine steps to move forward from error. Cognition, Technology & Work, 4(2), 137-144.
[62] Wooldridge, M. (2008). An introduction to multi agent systems. ISBN: 9780470519462. Wiley.
[63] Yen, J., Fan, X., Sun, S., Hanratty, T., & Dumer, J. (2006). Agents with shared mental models for enhancing team decision makings. Decision Support Systems, 41(3), 634-653.

Policy Management and Negotiation: Enabling Effective Human Autonomy Teaming

Michael Thomas
QinetiQ
Farnborough, GU14 0LX
UNITED KINGDOM
mthomas1@qinetiq.com

Rob Cottrell
QinetiQ
Farnborough, GU14 0LX
UNITED KINGDOM
rjcottrell@qinetiq.com

Antony Grabham
Dstl
Portsmouth, PO17 6AD
UNITED KINGDOM
apgrabham@mail.dstl.gov.uk

ABSTRACT

Autonomy technology is being developed and deployed in support of military unmanned systems. However issues persist, particularly in relation to future aspirations for unmanned aircraft operations, on how to balance requirements for flight certification and attendant needs for transparency and the trust of machine autonomy by human operators and team members, with the need for increasingly complex autonomy to achieve mission objectives. With autonomy being essential to unmanned aircraft operations during certain mission phases or under specific circumstances, human autonomy teaming cannot be effectively achieved by merely managing levels of autonomy. Effective human autonomy teaming needs to include paradigms for ensuring that autonomy technologies are employed that are appropriate to the prevailing mission and operational context and which satisfy the applicable policies. Furthermore, human autonomy teaming needs to include provision for autonomy technology to electronically negotiate with controlling authorities where such negotiations are expected such that routine adaptations can be accommodated. In such cases human intervention is only required (by policy) when human authority or decision complexity demands it. This paper discusses research into a Configurable Operating Model for Policy Automation and Control of Tasks (COMPACT) which seeks to employ higher-integrity policy management technology to enable lower integrity with more complex autonomy technology invoked according to the prevailing context and with provision for electronic negotiation between operating tiers of command. The paper also summarises Live Virtual Constructive trials conducted in 2017 to test elements of the research with military participants in militarily representative scenarios.

1.0 INTRODUCTION

This paper provides an overview of a conceptual adaptable autonomy architecture that is currently the subject of a UK MOD research programme undertaken by QinetiQ on behalf of Dstl. The paper also describes a recent trial during which the architecture was tested using military personnel against the backdrop of a militarily representative scenario.

The Configurable Operating Model for Policy Automation and Control of Tasks (COMPACT) was originally conceived [1] as a means of addressing the problems associated with the employment of advanced and novel computing techniques that can deliver the complex, and often ‘intelligent’, software that underpins automated and autonomous systems.

The role of automation and autonomy software in aircraft, particularly unmanned aircraft, has significantly increased and has led to more complex sets of requirements to provide the necessary functionality to enable human operators and pilots’ greater scope and capacity for undertaking missions.

In many cases Tthese complex requirements can only be delivered through provision of novel and non-traditional software development concepts, such as Artificial Intelligence (AI) and Machine Learning (ML). This can lead to software components that are difficult to certify and therefore exploit on operational

platforms [2]. In addition, the increased software complexity can sometimes be difficult to interpret by the operators and pilots who use them, leading to issues of trust and comprehension.

For a number of decades, this potential division of roles and responsibilities between the human and the system has been subject to much research. Early efforts focussed on the role of a system providing pilots of manned systems some level of assisted, associate or coach support [3]. Examples include the US PA [4]and RPA [5], and the UK MMA [6] & [7], FOAEW TDSS [8] and Cognitive Cockpit [9].

In many of these cases, the final responsibility for decision making was retained by the human operator as they were ultimately in control of their platform. But today, with operators of unmanned systems potentially expected to control multiple platforms, and multiple platform types, the human component has much less cognitive capacity to effectively manage their responsibilities without the application of novel software able to provide the required levels of decision support, automation and autonomy. Furthermore, it is expected that there will be mission phases or specific circumstances when autonomy is routinely required as an essential component of future unmanned aircraft operations, rather than being required as a crew aid as was often the case for the research focussed on manned platforms. The COMPACT concept provides a potential means to bridge many of these issues and provide a robust mechanism for ensuring complex software exploitation and trusted decision making.

2.0 THE COMPACT CONCEPT

The COMPACT concept is intended to provide a means for monitoring and controlling Unmanned Air Systems (UAS) that necessarily employ various autonomy and automation technologies, based on a set of pre-determined (and configurable) rules.

When characterising the COMPACT concept, it is worth looking back to the human machine teaming concepts reported in [3]; the assistant, associate and coach. Within these definitions, the COMPACT concept has a role similar to that of a coach, in that its purpose is to monitor both the performance of the system and the human (in terms of how the human interacts with the system), and determines whether this performance is consistent and compliant with applicable rules and constraints.

In many respects, the COMPACT concept architecture (see Figure 1) shares many of the capabilities of the Situational Assessment Support System (SASS) and certain elements of the Task Interface Manager (TIM) developed under the Cognitive Cockpit Programme [9]. The SASS comprised a rule based system which, by understanding the pilot's cognitive state, the platform status in relation to the mission plan (and associated task model), and the environment, could determine the appropriate course of action in response to dynamic events and most appropriate means of conveying information to the pilot for the current level of autonomy (using the Pilot Authority and Control of Tasks taxonomy – PACT). In relation to COMPACT, the UAS operator's state could be an analogue of pilot's cognitive state, however the initial research focus has been on the assessment of the UAS in relation to its mission goals, mission plan and the prevailing context.

In order to ensure that COMPACT can achieve the desired aims, it requires information and knowledge that relates to the platform/system state, mission and operational context, mission plans and goals, and the task models that describe tactics, techniques and procedures (TTPs). Using this information COMPACT is will be able to monitor the platform state against the platform model, task model, mission and task plans and identify discrepancies. COMPACT will have knowledge of courses of action when discrepancies in the platform state against the platform or task model, or the sortie, mission and task plan occur. COMPACT will expedite defined actions to ensure consistency is maintained in the platform state against the platform and task models and the sortie, mission and task plans unless this conflicts with the control state.

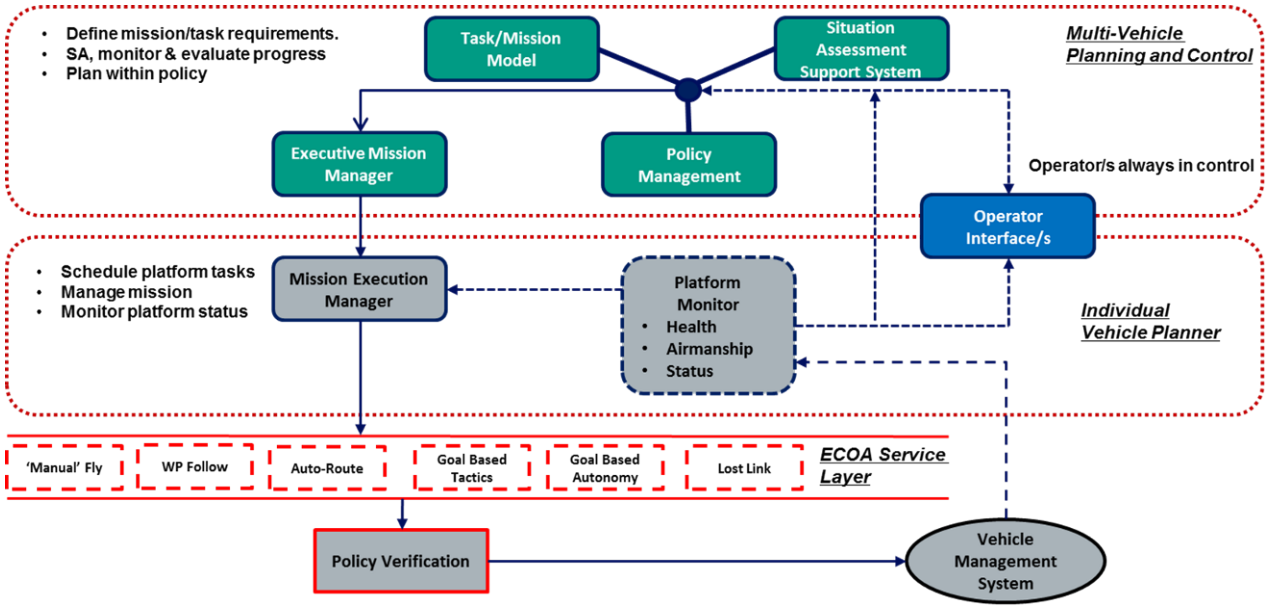


Figure 1 COMPACT conceptual architecture

COMPACT is predicated on the idea that there are a collection of software agents that can invoke various types of software components (including novel and AI techniques) able to provide platform or system decision support, automation and autonomy functions (which may be implemented using the European Component Orientated Architecture – ECOA, see [10]). The invocation of these agents is controlled by a Task or Mission Management function in response to evolving mission requirement and situation assessment.

The Task Management component continuously monitors which agents can be used through Situation Assessment (SASS) and policy management. The SASS Module will continuously monitor current status. To include:

- Environmental: terrain, landscape, weather
- Platform: position, altitude, speed, direction, fuel,
- Mission Phase– current and future (as advised by Task Manager)
- Policy Manager - Airspace / Mission Compliance; rules, constraints and Rules Of Engagement (ROE) as defined by Policy data

The Task Manager Module has knowledge of how the mission should be carried out (based on a priori knowledge) and will 'decide' which behaviours are to be conducted in order to achieve the task. These behaviours may include those delivered by the agents, the operator, or combinations thereof.

2.1 Policy Management and Negotiation

Policy Management is the means by which the performance and configuration of a system is monitored by COMPACT to ensure its operation is compliant with current policies. The term policy refers to the constraints and guidelines within which military operations are to be conducted. At the highest level, different nations apply laws and regulations that must be adhered to, such as (in the UK) the Law of Armed Conflict. At the operational level, the ROE provide a set of constraints that determine how military force should be applied. Tactical operations are managed and directed using a multitude of policies and in the case of aircraft; airspace management and the Airspace Control Order (ACO) are of particular relevance as is the

Air Tasking Order (ATO). At the platform level, Tactics, Techniques and Procedures (TTP) shape behaviours that maximise platform performance and the mission plan to ensure the desired military effect (see Figure 2).

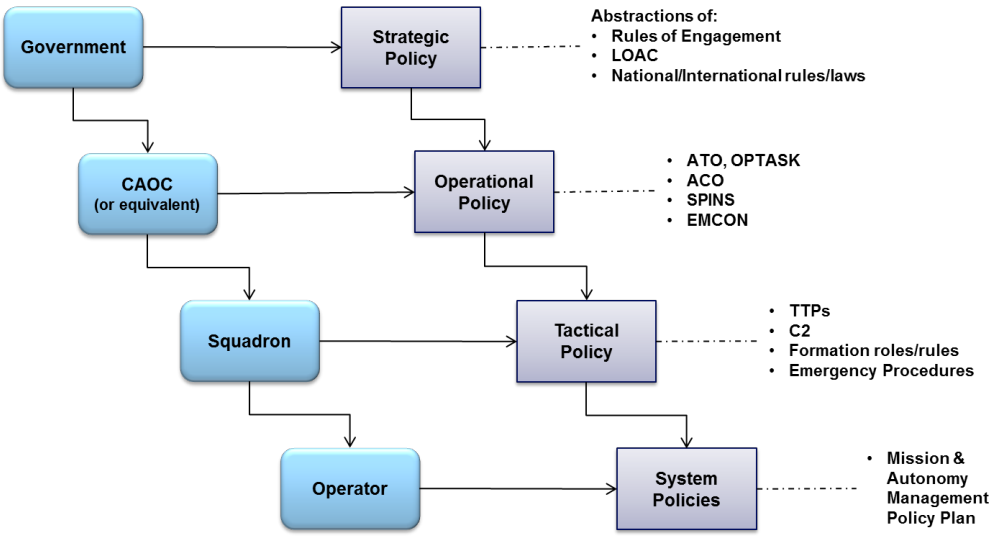


Figure 2 Policy hierarchy

As is the case in manned military aviation, there are occasions when, as a mission progresses, policies come into conflict. A typical example includes the re-planning or re-tasking of a UAS asset that requires it to enter airspace it is not permitted to enter. For manned aviation the pilot will request access to the required airspace from the relevant authority. This interaction between the pilot and the authority represents a negotiation, which is a natural consequence of policy management.

An initial model for the application of negotiations within the policy management framework is depicted in Figure 3. The detection, assessment and classification of a policy conflict is made and the determination as to whether the adjustment will require a change in Goal, Means or specific Acts will determine the process to resolve the conflict.

- A Goal adjustment is a change to the mission/task itself, and may require a negotiated resolution and approval from a higher C2 authority
- An adjustment of a Means, may require either;
  - A resolution requiring approval (from a GCS operator or Mission Commander)
  - A resolution requiring authorisation from a higher C2 authority
- An adjustment of an Act may require either;
  - An internal resolution (i.e. a minor change to the route plan)
  - A resolution requiring negotiation and approval from another agent (i.e. from a GCS operator or Mission Commander)

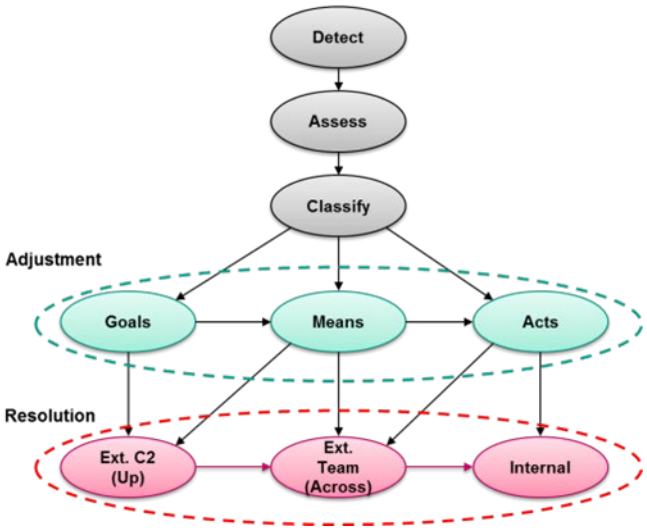


Figure 3 Policy conflict identification and resolution.

To aid the development and implementation of negotiations, as well as their validation and verification, formal methods have been applied in the form of formal agreement patterns [11]. Agreement patterns define patterns for sharing and reusing previous experience in the application of agreements (see [11] & [12]) and provide an abstract way of defining negotiations. Within COMPACT, each policy resolution can be considered to be a type of "negotiation". For instance 'platform 1' talks to 'platform 2' and exchanges some kind of information transaction forming a negotiation.

There are three parts to an agreement pattern; informal representation using classifications to direct the description, reasoning diagrams to show the elements of the agreement, and formal representations identifying the preconditions and post conditions for entering and concluding the agreement. For example, the informal representation of an Airspace Negotiation is defined below:

NAME: Auto Airspace Change
DURATION: Short Term
UAV Operator/autonomy system needs access to airspace for which there is no current authorisation (for example UAV is tracking a target that is heading outside of its currently allocated airspace and needs to relocate to maintain track)
FORCES: Pattern applicable when task requiring access to new airspace is within mission bounds (i.e. a new mission plan is not required), UAV operator/autonomy can handle any restrictions authoriser may give (such as area/height/time bounds), there is an identified authoriser (e.g. Air Battlespace Manager) and there are no other known circumstances which negate the validity of the request (e.g. emergency recovery plan invalidated due to insufficient fuel)
SOLUTION: Approver authorises the request, if there are restrictions with the approval then the UAV operator/autonomy must confirm compliance with these restrictions before entering the airspace.
EXAMPLES: COMPACT
RESULTING CONTEXT: UAV is authorised to enter the airspace to conduct task
PHASE: Agreement negotiation, Agreement enactment, Agreement renegotiation, Agreement conclusion
RELATED PATTERNS: ---

In this example,

- **Name** is a meaningful name that provides a vocabulary for identifying and discussing the negotiation agreement pattern.
- **Duration** is the time the negotiation is covering.
- **Forces** are a description of the relevant forces and constraints and how they interact with one another and with the goals and considerations to be taken into account to select a solution for a problem.
- **Solution** describes static relationships and dynamic rules describing how to realise the desired outcome.
- **Examples** are one or more sample applications of the pattern which illustrate its application and known occurrences of the pattern which help in verifying that the pattern is a proven solution to a recurring problem.
- **Resulting context** is the state or configuration of the system after the pattern has been applied.
- **Related patterns** are any other compatible patterns which can be combined with the described pattern.
- There can be any of the following **phases** in the agreement pattern:
  - **Provider selection**: Selection of the possible provider before making an agreement
  - **Agreement negotiation**: Process of establishing the conditions of the agreement
  - **Agreement enactment**: Commitments of the agreement being effective
  - **Agreement renegotiation**: Phase of changing the conditions of a previously negotiated agreement
  - **Agreement monitoring**: Process of reviewing the established agreement conditions during the agreement enactment
  - **Agreement conclusion**: Process of concluding the agreement by any party

The reasoning diagram for the Airspace negotiation uses notations defined within the CommonKADS methodology [13], specifically identifying *Knowledge Roles* and *Inferences* (Figure 4). The reasoning flow commences when the air platform needs to enter airspace which it has not been assigned (this is a knowledge role). The system then needs to assess the mission bounds, e.g. by assessing whether failing to get permission to enter this airspace will have an impact on the fuel or emergency recovery/diversion plans. Since the mission bounds need more information this is an inference role. The approver then needs to respond to the platforms request. In the example the approver is usually the airspace controller. Their response can either; allow the platform to enter the airspace, not allow the platform to enter the airspace, or allow the platform to enter the airspace with restrictions. In the example, restrictions could include duration, height, or a subset of the airspace requested. The platform then asses the approvers response and decides if it can handle the restrictions and enter the airspace. Since there is some calculation in the assessment part of the reasoning diagram this becomes and inference role. After assessing the response the platform then decides if it will enter the airspace or not.

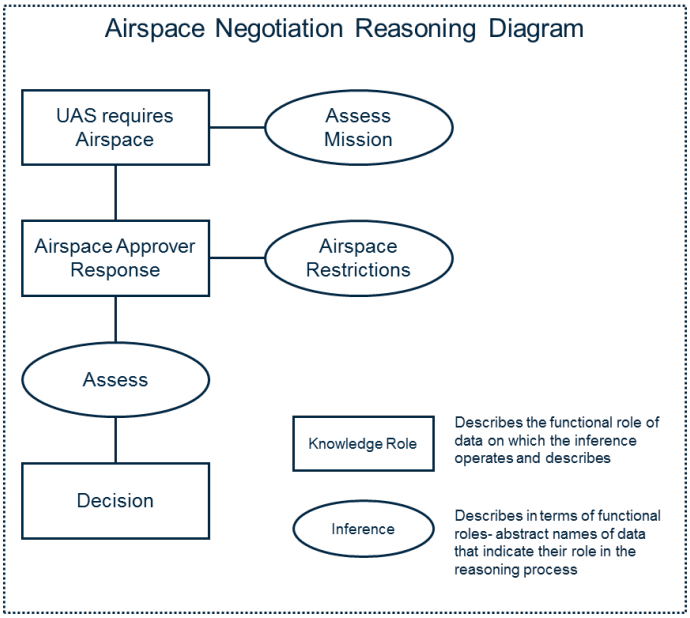


Figure 4 Airspace Negotiation reasoning diagram

The third element, formally defining the pre and post conditions, requires initial definitions of the relevant variables and their types:

variable: TYPE	**Description**
ap : ACTOR	ap is the approver
approved: (ACTOR1* ACTOR2)	Approved is a set of pairs which takes 2 actors meaning ACTOR1 is approved by ACTOR2
m: ACTOR	m is the mission bounds
flying: (ACTOR1* AIRSPACE1)	flying is a set of pairs which takes 2 actors meaning ACTOR1 is flying in AIRSPACE1
u: ACTOR	u is the unmanned air system
a_1 : AIRSPACE	a_1 is the airspace the uav is allowed to fly in
a_2 : AIRSPACE	a_2 is the airspace the uav is trying to get into
res: ACTOR -> {PARAMETERS}	res is a function which takes an actor and outputs a set of parameters (or restrictions)

The first precondition that is represented formally is that before the platform requests entry to new airspace, it must be flying in some kind of airspace already. Therefore the following applies:

$$\underline{flying(u, a_1) \wedge (\neg\ flying(u, a_2))}$$

This equation represents that the platform, in this case UAS (*u*) is *flying* in airspace1 (a_1) **and** the UAS (*u*) is **not** *flying* in airspace2 (a_2). Another precondition that should be identified before entering the negotiation is that the UAS has some restriction it will be able to manage if the approver gives any. This is defined as:

$$\underline{res\ u \neq \{\}}$$

In this equation the variable *res* takes an actor (the UAS) as an argument. Therefore the first part of the equation *res u* represents the set of restrictions that the UAS (*u*) can manage. The second part "≠ { }",

represents that the set is not empty. Therefore the UAS has to have at least some restrictions which it is able to manage. At this abstract level of representation it is not important as to what the restrictions are, just that they exist.

The post condition of the negotiation defines the outcome of the agreement. This is represented as:

$$\underline{\text{flying}(u, a_2) \wedge (\neg \text{flying}(u, a_1))}$$

This equation shows that after the agreement, the UAV (u) is flying in airspace2 (a_2) and the UAV (u) is not flying in airspace1 (a_1). This is similar to the first formal statement produced where the UAV was flying in airspace1.

The formal notation taken from the formalising of the Auto Airspace change can be used to define a formal specification. The pre- and post-conditions in the pattern agreement denote the preconditions before entering the pattern and the post conditions after the agreement. Whereas the preconditions and post conditions in the formal specification denote the states before and after the actual process has been conducted. The preconditions used in the formal specification take the preconditions from the agreement pattern and add to them any other preconditions before conducting the action. Various notations can be used within the formal specifications for negotiations. We define the notations as follows:

NameOfSchema
Declarations
Preconditions Postconditions

The **NameOfSchema** gives a name to the process. The declarations include the variables which may be used in the process. This includes options, function names etc. each variable is assigned to a type which it belongs to. The format for a declaration is "v:T", where v is the variable and T is the type. There can be none, one or many declarations in each schema.

The preconditions define what is needed before the desired outcome is to be executed. The post-conditions define what state the variables are in after the desired outcome. Additionally, the Ξ is used to define the same variables from a previous schema negotiation.

AutoAirspaceChange
ap : ACTOR approved: (ACTOR1* ACTOR2) m: ACTOR flying: (ACTOR1* AIRSPACE1) u: ACTOR a_1, a_2 : AIRSPACE res: ACTOR -> {PARAMETERS}

flying(u, a_1) ∧ (¬ flying(u, a_2))
approved(u, m)
approved(u, ap)
res u ≠ { }
res ap ⊆ res u
flying(u, a_2) ∧ (¬ flying(u, a_1))

Three new preconditions have been added to the specification which denotes the conditions which need to be satisfied in order for the aircraft to actually change its airspace.

(1) approved(u, m)

(2) approved(u, ap)

(3) res ap ⊆ res u

Precondition (1) states that the UAV (*u*) has to been approved by the mission bounds (*m*) for it to enter the new air space. In the formal notation the mission bounds (*m*) are defined as an actor. The mission bounds relate to making sure the air space change does not conflict with the mission. If there are restrictions on the mission bounds then the higher chain of command may recommend that the aircraft should change airspace as that is a priority. Therefore if the higher chain of command makes changing of airspace a priority then it has been approved. At this abstract level there is no need to identify what the mission bounds may be only that the UAS does not change the mission and therefore does not need permission, or that it does change the mission and has been approved by the appropriate party.

Precondition (2) states that the UAV (*u*) must have been approved by the approver (*ap*). The approver would usually be the airspace controller in this example but could be another authority in other examples such as a Mission Commander.

Precondition (3) states that the restriction (*res*) the approver (*ap*) is a subset of the restrictions (res) the UAV (*u*) can manage. Thus the restrictions the UAV can manage can be in a set as follows:

{Can fly between 7000-7400 feet, can fly for 4 hours, can loiter for 20 minutes before entering, … }

The approver may allow the UAV to enter the airspace but only for 4 hours. Therefore the restrictions given by the approver can defined in the following set:

{can fly for 4 hours}

Thus the restrictions given by the approver is a subset of the restrictions the UAV can handle. At this abstract view specific restrictions are unimportant but that the UAV may have some restrictions it can manage which the approver may or may not state.

Since there are 5 preconditions which must satisfied in order for the UAV to successfully change airspace then there must be 5 outcomes which may happen if one of the pre conditions has not been satisfied. A new type OUTCOME is defined to output an outcome for each pre condition if it has not been satisfied.

OUTCOME ::= success | outWithinMissionBounds | approverRefuses | uCantHandleRestricitions | uNotInCorrectAirspace | uHasNoRestrictions

For each of the preconditions a schema is identified to show what happens if the pre-conditions are compromised.

Precondition	Schema
flying(u, a1) ∧ (¬ flying(u, a2))	**NotFlyingInCorrectAirspace** Ξ AutoAirspaceChange o: OUTCOME ¬flying(u, a_1) o = uNotInCorrectAirspace
approved(u, m)	**NotWithinTheMissionBounds** Ξ AutoAirspaceChange o: OUTCOME ¬approved(u, m) o = outWithinMissionBounds
approved(u, ap)	**NotApprovedByAirspaceHead** Ξ AutoAirspaceChange o: OUTCOME ¬ approved(u, ap) o = approverRefuses
res u ≠ { }	**UAVHasNoRestrictions** Ξ AutoAirspaceChange o: OUTCOME res u = { } o = uHasNoRestrictions
res ap ⊆ res u	**UAVCan'tHandleRestrictions** Ξ AutoAirspaceChange o: OUTCOME res ap ⊄ res u o = uCantHandleRestricitions

Since there are outcomes each time a precondition is compromised, a schema can be defined to identify when no preconditions are compromised and the UAV has successfully entered a new airspace.

SuccessfullyChangesAirspace
Ξ AutoAirspaceChange o: OUTCOME
¬approved(u, m) o = success

Using the functions defined, the entire operation of the negotiation between the UAV and other actors involved can be totalised.

TotalAirspaceChange==	(AutoAirspaceChange ∧ SuccessfullyChangesAirspace) ∨ NotFlyingInCorrectAirspace ∨ NotWithinTheMissionBounds ∨ NotApprovedByAirspaceHead ∨ UAVHasNoRestrictions ∨ UAVCan'tHandleRestrictions

For a total negotiation we can have a successful negotiation to enter an airspace (**and** a success outcome) **OR** the change wouldn't be within the mission bounds **OR** on request to enter the airspace the UAV is denied access **OR** the restrictions given to enter the airspace is more than the UAV can handle. By totalising the operation all the possible outcomes can be viewed. All exceptions have been identified and there should be no unidentified outcomes when conducting this operation.

3.0 TESTING THE CONCEPT

The COMPACT concept has been subject of several years of Dstl funded research. Earlier phases of the research programme demonstrated proof of concept in laboratory environments using representative UAS control stations, C2 and rotary and fast jet mission management systems. The most recent phase of the research programme tested the concept in a Live, Virtual and Constructive (LVC) trial

The trial, called Cardigan Bay 2017 (CB17), was undertaken to address a number of key research requirements identified by UK MOD (specifically, Dstl and DE&S) some of which directly related to Human Autonomy Teaming or Manned and Unmanned Teaming. From these requirements, additional requirements were derived to form the basis for a set of research hypothesis. Use Cases were constructed against which the hypotheses could be tested and success criteria associated with each hypothesis were derived – thereby identifying whether the trial achieved its aims. Those hypotheses that directly related to Human Autonomy Teaming and Manned Unmanned Teaming were;

- More operationally effective MUMT can be obtained when remote users, including airborne users, can exercise in-mission task-level control of UAS and/or their payloads to satisfy dynamically evolving mission requirements. This was compared against one UAS per operator)
- Autonomy provides an enabler for MUMT remote users to exercise task-level control of UAS and/or their payloads
- For autonomy in support of MUMT to be militarily exploitable it will need to be applicable within multi-tier C2 infrastructures
- Automated support for Policy Management and Negotiation is an enabler to exploiting autonomy for MUMT within multi-tier C2 infrastructures

3.1 Cardigan Bay 2017

CB17 was primarily undertaken at the Llanbedr airfield in West Wales, with additional functions undertaken at Aberporth and at the QinetiQ Farnborough Site (for geography, see Figure 5). The high level scenario for CB17 involved a Forward Operating Base (FOB) at Llanbedr, which would comprise small UAS flying (the Live element) as well as host virtual manned and unmanned platforms. A Main Operating Base (MOB) at Aberporth which would include both Constructive and Live elements. Note, Aberporth is a UK MOD range which can track air platforms flying in the in Cardigan Bay range with live track data from fast-jet traffic fused during trials together with pre-recorded data from Watchkeeper UAS sorties undertaken previously. In the case of the constructive element, recorded telemetry data was replayed in the LVC environment. Live data was also be injected into the LVC environment when aircraft are operating in the area. Farnborough

acted as the main Joint Air Operations Centre (JAOC) undertaking theatre Air Battle Management. All sites were linked through wide area network that simulated various datalink and satellite communications protocols.

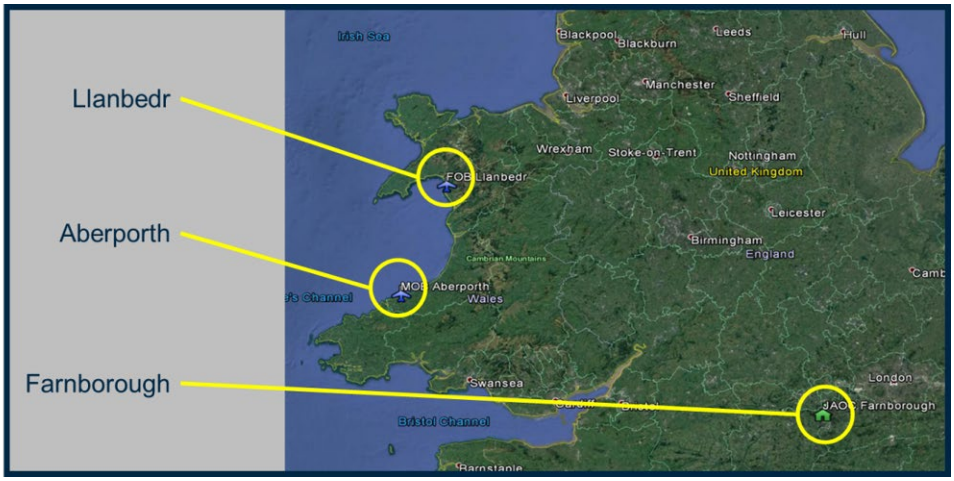


Figure 5 CB17 trial sites

The scenario used during the trial was derived from earlier activities undertaken in the West Wales and Cardigan Bay areas. Principally, it is a littoral scenario that encapsulates both maritime surveillance and land focussed Base Protection. The trials architecture, depicted in Figure 6, involved multiple platforms and stations spread between Farnborough and West Wales. Communications between stations was managed using standard datalink messaging (Link 16, STANAG 4586 and VMF) that would be applied in real operations. Many of the stations were manned by Military Participants (and some manned/operated by the technical team).

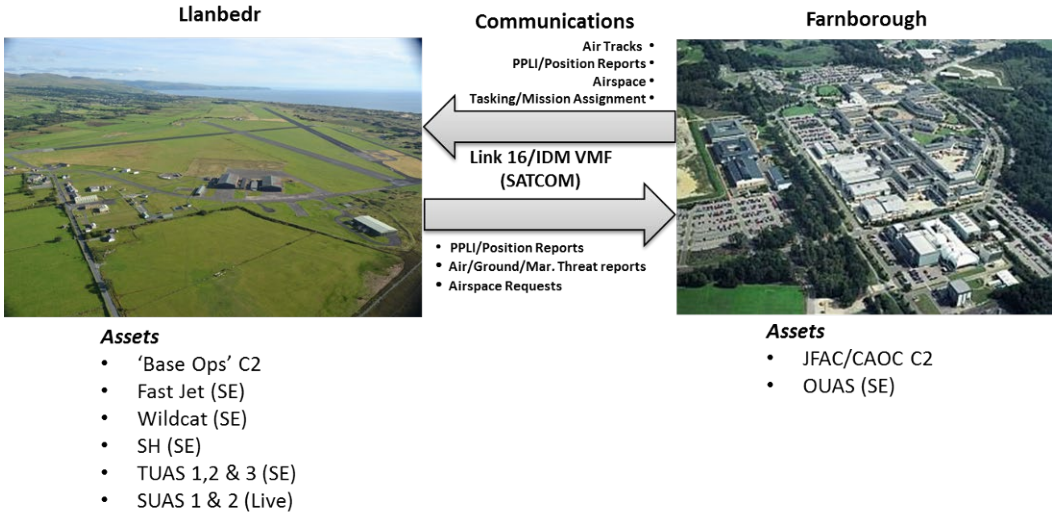


Figure 6 Trial set up

The trial design was focussed on three Use Cases that would address both littoral operations as well as Base Protection (BP). The first Use Case, in Figure 7, addressed the collection and dissemination of ISR from

different airborne platforms, some manned, but most unmanned and of different classes (size, weight and endurance). Enemy forces are operating both at sea and in-land, and so for those in tactical command of friendly forces, they must schedule and plan missions for the air assets to ensure maximum coverage of the area of interest, evolving and adapting plans and schedules as the tactical situation unfolds. In order to coordinate air assets, airspace was managed by an Air Battle Manager (ABM). As the Use Case commenced, all Air assets are allocated airspace, within which they would remain unless operators negotiate with the ABM to access to other airspace.

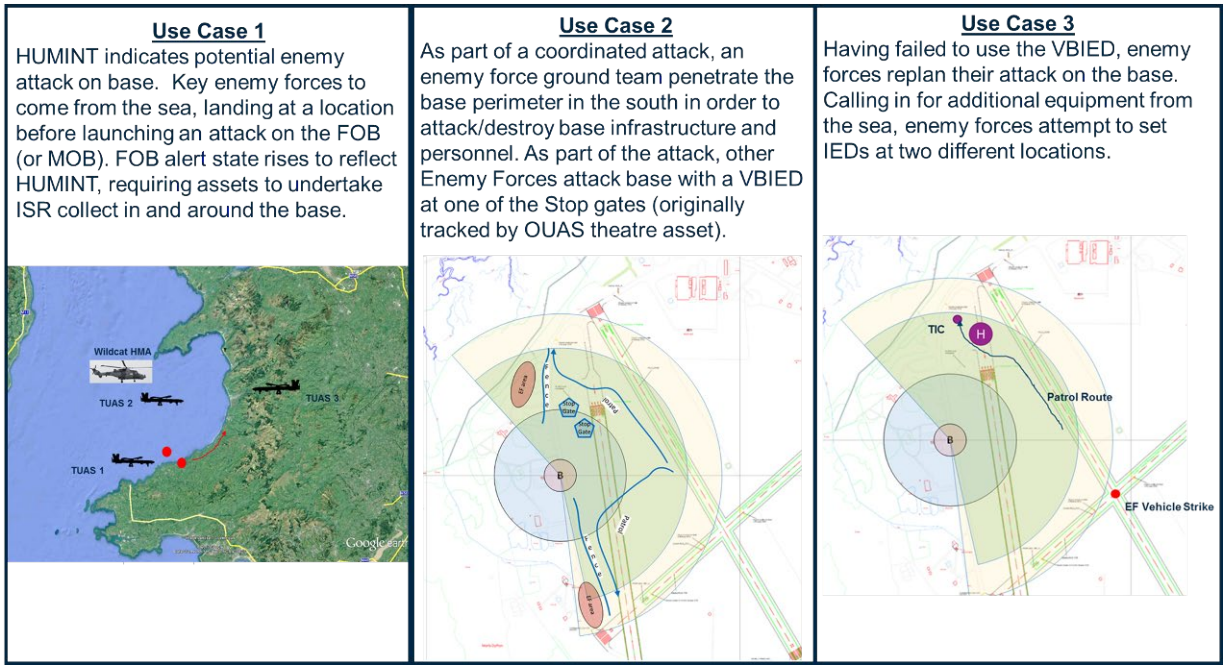


Figure 7 Use Cases for CB17

The second Use Case (Use Case 2 in Figure 7) builds on the first but includes Red Forces preparing a coordinated attack on the FOB. The focus was therefore towards the protection of the FOB through utilisation of ISR assets to locate and identify potential threats and react accordingly. This requires assessment of ISR data from multiple sources, the coordination of which spans the three sites, SUAS at Llanbedr (the Base), TUAS from Aberporth, and the OUAS operated from Farnborough (under the JAOC). The specific threats to the base included small teams attempting to infiltrate the base perimeter and a vehicle borne improvised explosive device (VBIED).

The final Use Case (Use Case 3 in Figure 7) focusses on the management of ISR assets when reacting to multiple events which occur within the base or outside or the base perimeter. This requires assessment of ISR data from multiple sources and coordination by the Base Commander to utilise assets to the best effect, ultimately requiring coordination of strike and ISR platforms to neutralise threats. In this instance, the Base perimeter is infiltrated by red forces leading to troops in contact which the Base Commander must react to.

The trials plan accounted for one assessed run for each Use Case. A training run was undertaken at the beginning of the trial to familiarise the military participants with the systems, which included instantiations of the COMPACT components.

3.2 COMPACT Assessment

The COMPACT concept is realised through provision of architectural components incorporated into

research systems that provide policy management functions and support to negotiation (via appropriate digital messaging). For the CB17 trial the systems that included COMPACT were the TUAS and SUAS control stations, manned rotary wing mission system, fixed wing mission systems, the C2 station and JTAC support system. There were 3 Military participants located at the Llanbedr trials site who were able to support the trial and operate the equipment. Their roles and systems they used in the scenario were as follows;

- Military Participant 1 was the Mission Commander at the FOB and operated the C2 and JTAC systems
- Military Participant 2 was the TUAS commander and operated the TUAS control station
- Military Participant 3 was the Rotary Wing aircraft commander (use case 1 & 2) and the SUAS operator (use case 3), and operated the Rotary Wing mission system and SUAS control Station respectively.

Each system had implemented policies and support to negotiations that were suitable for the relevant role. Details of implemented policies and negotiations are described in Table 1

Table 1. Policy Management and Negotiation support implemented on the CB17 systems.

	Policy Checks	**Negotiations**
TUAS CS	• Checking that auto-routeing (and manual routeing) conforms to airspace constraints • Checking that SAR plan conforms to SAR TTPs • Checking that Cooperative Laser Targeting with coordinating strike platform conforms to TTP and ROE • Checking that dynamic mission/route changes do not compromise Emergency Recovery Planning policy, including de-confliction and fuel planning	• Request for Airspace via Digital CHAT message (vertical cross-tier negotiation with C2) • Formatted messages for cooperative target engagement (horizontal peer-to-peer negotiation) • Digital authorisation for platform/sensor Level of Interoperability (LOI) hand-over (horizontal peer-to-peer negotiation)
SUAS CS	• Checking that auto-routeing (and manual routeing) conforms to Base Operating Procedures and airspace constraints (intended to be exercised in all use cases) • Checking that dynamic mission/route changes do not compromise Emergency Recovery Planning policy	• Request for Airspace via Digital CHAT message (cross tier negotiation with C2) • Digital authorisation for platform/sensor LOI hand-over (peer-to-peer negotiation) •

Rotary MM	• Checking that auto-routeing (and manual routeing) conforms to airspace constraints • Checking that coordinated strike behaviour with TUAS buddy-lase conforms to TTP and ROE • Warnings, Cautions and Advisories associated with TUAS platform/sensor control	• Request for Airspace via Digital CHAT message (cross-tier negotiation with C2) • Digital request for UAS platform/sensor LOI hand-over (peer-to-peer negotiation) • Digital authorisation for LOI hand-over of own EO/IR sensor to remote user (peer-to-peer negotiation) • Checking that Cooperative Laser Targeting with coordinating strike platform conforms to TTP and ROE (peer-to-peer negotiation
Fast Jet MM	• Checking that task/routes conform to airspace constraints • Checking that coordinated strike behaviour with JTAC conforms to TTP and ROE in terms of weapons effects and blue force positions	• Request for Airspace via Pre-formatted Digital message (Cross tier negotiation) • Checking that Cooperative Laser Targeting with coordinating JTAC conforms to TTP and ROE (peer-to-peer negotiation)
C2 MM	• Checking that task/routes conform to airspace constraints • Checking that task/routes are de-conflicted • Checking that air-vehicles are de-conflicted • Checking that Positive Identification (PID) is maintained during Track Handover • Checking that range, LOS, airspace and time constraints are met where a platform is employed as a Communications Relay • Checking that UAS tasking have associated emergency recovery plans that meet policy • Checking that when a Communication Relay role is to be withdrawn from a platform, it is safe/appropriate to do so • Checking that DACAS "9-Line" request can be serviced according to TTP	• Allocation of Airspace in response to digital airspace request via pre-formatted digital message (Cross tier negotiation intended to be exercised in all use cases • Track handover management • Checking that Cooperative Laser Targeting with coordinating JTAC conforms to TTP and ROE (peer-to-peer negotiation)

JTAC MM	• Checking that Airspace is allocated/de-conflicted • Checking that digital targeting messages are populated correctly • Checking that coordinated strike behaviour with TUAS buddy-lase conforms to TTP and ROE	• Allocation of airspace using J and/or K Series digital message to effector platforms (peer-to-peer negotiation) • Request for Airspace from C2 via Digital CHAT message (cross-tier negotiation with C2) • Digital request for UAS platform/sensor LOI hand-over (peer-to-peer negotiation) • Checking that Cooperative Laser Targeting with coordinating strike platform conforms to TTP and ROE (peer-to-peer negotiation)

To test the application of COMPACT concepts across these systems, suitably qualified and experienced military operators were required to operate these systems so that appropriate assessment could be undertaken. To assess the COMPACT, military operators were asked to provide a subjective assessment of the systems that they used. Principally, the assessment was concerned with measures of effectiveness and performance (MOE&P) for 4 key aspects of system capability; Autonomy, Policy Management, Negotiation and MUMT. Military operators were provided with MOE&P scoresheets to record their subjective views (Figure 8).

	Measures of Effectiveness & Performance											
	Measure of Effectiveness or Performance	Anchors (Low)	Weightings 1	2	3	4	5	6	7	N/A	Anchors (High)	Comments
Autonomy	Did the overall use of Autonomy for multiple platforms increase your SA?	*Lower SA than one person & 1 UAS Higher workload to manage UAS*									*Greater SA than one person & 1 UAS Lower workload to manage UAS*	
	Did the autonomous SUAS Search Patterns result in an increase of SA?											
	Did the autonomous SUAS Platform Scheduling and/or Planning result in an increase is SA?											
	Did the autonomous sensor coordination result in an increase in SA?											
	Did the Autonomy adaptively support the operator	*No adaption*									*Highly Adaptive*	
Policy Management	Overall, did the Policy Manager manage the Autonomy to operate within authorised parameters?	*Autonomy never operated within parameters No Confidence with Policy Manager*									*Autonomy always operated within parameters Highly confident with Policy Manager*	
	How confident were you that the UAS operated within applicable TTPS wrt match sensor to target											
	How confident were you that the UAS operated within applicable airspace											
	How confident were you that the UAS operated within applicable ROE											
Negotiation	Does negotiation enable Autonomy and PMF to dynamically adapt to meet evolving mission requirements?	*No adaption*									*Highly adaptive*	
	Was it clear when you were required to negotiate with Autonomy?	*Not obvious Unclear*									*Obvious Clear*	
	Was it clear when the Autonomy was required to negotiate with you?											
	Did the W/C/A facilitate negotiation?	*Useless Complex*									*Useful Straightforward*	
	Were required negotiations undertaken in a timely manner	*Slow*									*Fast*	
	Did/do STANAG defined digital messages support negotiation	*Undefined*									*Defined*	
	Is chat useful for undertaking Negotiation	*Not Useful Unimportant*									*Useful Vital*	
	Is Voice useful for undertaking negotiation											
MUMT	Manned Unmanned Team increases capability of the operator in conduct of a mission/task	*No appreciative increase in Capability*									*Increase in capability*	
	Did the provision of Sensor information (LOI2) increase SA	*No increase in SA*									*Improved SA*	
	Did the provision of Sensor control (LOI2 & LOI3) increase SA											
	Control of a UAS did not adversely affect operator workload and means to conduct the mission/task	*Execution of mission task impossible*									*Manageable workload*	
	Teaming with an unmanned system increased mission Performance	*Disagree*									*Agree*	
	Teaming with an unmanned system increased mission effectiveness											

Figure 8 Military Participant assessment score sheet.

Military participants undertook the assessment and completed the scoresheets after each run. The assessment process include an initial 'after action review' which acted as an opportunity to debrief and discuss the execution of the run prior to the formal assessment process. In addition to the MOE&P assessment, a formal assessment was undertaken by Dstl (results of which are reported separately).

4.0 CB17 RESULTS

The subjective results gathered from military operators are summarised in Table 2. The results generally show a steady improvement, with each aspect being rated with either low- or mid- range scores for the first

use case run, and mid- to high-range scores for the final use case run.

The scores for the TUAS GCS *Autonomy* were originally rated quite low due to the fact that the military participant using the system was unaware that the policy manager had constrained the application of the autonomy software components due to airspace policy violations. However, this is precisely what the COMPACT concept is aiming to achieve and it was clear that the means by which policy violations are communicated to the operator required additional work.

Table 2 Subjective results summary from the CB17 trial

		Use Case 1	**Use Case 2**	**Use Case 3**
TUAS GCS	*Autonomy*	Low	Low	Mid
	Policy Management	Low	Low	Mid
	Negotiation	Low	Low	Mid
	MUMT	N/A	N/A	Mid
C2/JTAC	*Autonomy*	Mid	Mid	Mid
	Policy Management	Low/Mid	Low/Mid	Mid
	Negotiation	Low	Low/Mid	Mid
	MUMT	Low	Mid	Mid
Wildcat/SUAS	*Autonomy*	Mid	High	High
	Policy Management	Mid	High	High
	Negotiation	Mid	Mid	N/A
	MUMT	Mid	High	N/A

For the first run the results from the MOE&P assessment showed that for the mission commander, airspace policy violations by UAS were not raised at the C2 station and consequently no Policy Management or automated Negotiation functionality was triggered. The ABM functionality between Mission Commander at Base Ops and TUAS operator were therefore conducted by voice, which may have contributed to the TUAS operator assessment that autonomy was not working in relation to airspace, whereas in fact any airspace that was in the TUAS control station and was attributed correctly triggered autonomy and policy management functionality including Warnings, Cautions and Alerts (WCA) – but these were overlooked/ignored by the TUAS operator during the run. MUMT was scored higher by the mission commander reflecting that he could see the platform positions and route plans for both manned and unmanned aircraft that were operating, or tasked to operate, in the airspace delegated to him by the JAOC.

The Wildcat military operator scores for autonomy, policy management, negotiation and MUMT were mid-range to high. It is likely that these reflect the relatively maturity of that particular system and the fact that

the operator had experience of using it in previous trials. Functionality exercised included instantiations of remote (TUAS) sensor feeds in integrated with the tactical map display showing remote platform stare-point and Field of View (FOV) for MUMT, together with digital LOI request and approval for remote platforms and sensors, which are a form of digital negotiation.

The TUAS operator scored autonomy low as it was not apparent to them that it was functioning. In actuality autonomy was working in the background in conjunction with policy management to produce WCA in relation to airspace, and also in relation to handling platform/sensor delegation to remote users (such as the Wildcat). The TUAS operator scored Policy Management low and this is believed again to be influenced by the operator not being clear as to what policy checking was being automatically applied behind the scenes, and only being aware of WCA when there was an issue, albeit that these were sometimes ignored. As a consequence, the operator stated that they were "not confident that that UAS operated within applicable policies" – despite this being the case (e.g. UAS staying within allocated airspace). Negotiation was scored low but this was influenced by the need to employ, for technical reasons, Chat messaging for some negotiations intended to be undertaken by preformatted datalink message. The operator found Chat messaging distracting and voice quicker. From the TUAS operator's perspective, MUMT was scored as "not applicable" because, once handed-off, they did not actively participate in MUMT with teaming exercised between RW operator and the handed-off platform. Taken in conjunction with the RW operator's scores, this should probably be taken to mean that MUMT was effective!

For the second run, the Mission Commander scored Autonomy mid to high with good SA of the Air Picture. Policy Management scores were low to mid-range because not all airspace issues were raised correctly to the operator, with the TUAS flying into unauthorised airspace. This was because of errors in the Chat message that should have triggered the automatic negotiation and airspace allocation. Negotiation scores were therefore low to mid-range. When the negotiation was repeated with the error corrected, the automated airspace allocation worked correctly and consequently there was no policy violation. As always, it was noted that voice was useful when autonomy (automatic negotiations in this case) did not work. MUMT was scored mid-range with SA assessed as having increased as a consequence of the TUAS being handed-off to the Wildcat helicopter.

The Wildcat HMA military operator scores for autonomy, policy management, negotiation and MUMT were mid-range to high, with SA assessed as having improved with hand-over of the TUAS, which correlates with the assessment of the Mission Commander. The Wildcat operator noted that control of the TUAS was achievable with simple commands.

During the third run, the TUAS operator scored autonomy higher than in the previous runs. The operator commented that the adaptive support provided by the autonomy sped up decision making, but most actions were still undertaking manually, indicating that there is potential advantage to be had from more automation/autonomy in relation the TUAS functionality exercised (which was search, track and cooperative targeting).

The TUAS operator scored Policy Management slightly higher (low to mid-range) than in previous runs, but still was not clear as to what policy checking was being automatically applied behind the scenes, and only being aware of WCA when there was an issue. Comments this time being "Operator unsure as to whether all actions were policy compliant" although the operator did qualify this with "Actions appeared to point to task (wrt airspace)". The Operator also commented "Messages conveying policies were minimal and could easily be missed" and "Alerts were clear but could also be missed" reinforcing assessments from earlier runs that some HMI development is required in relation to Policy Management.

Negotiation was also scored slightly higher than in previous runs, with the work-around of using a Chat message rather than a formatted message to request airspace now being familiar to the operator. However, the operator commented that "Chat was sometimes helpful but can slow you down depending on situation"

which is a recognised limitation of Chat and indicates that its use needs to be carefully controlled, ideally in accordance with TTP such that operators understand and are familiar with when it should be used and for what types of information (the intended mechanism for the negotiation being a computer generated, pre-formatted message).

The TUAS operator scored MUMT for the first time with scores mid-range. The operator assessed that "Target handoff was quicker" and that "Teaming sped up some processes". This is likely to reflect that the tools in the TUAS control station for supporting coordinated targeting in accordance with policy (TTP), together with the datalink integration which is a step towards automated negotiation, assisted the operator in conducting this time-critical task.

In this third run, the Mission Commander station was exercised both as a C2 station for managing airspace delegated to the Base Commander, but also as a JTAC station with support for digital targeting using the JTAC system. The Mission Commander scored Autonomy mid-range, commenting that there were "Multiple platforms but limited autonomy" which is true from the C2/JTAC perspective. C2 tools/functionality was designed to provide SA and decision support rather than autonomy. However, there were platform-to-platform de-confliction algorithms, route deconfliction algorithms in addition to the airspace de-confliction algorithms running in C2 as part of the Policy Management system.

The fact that these were not mentioned may be indicative of a similar issue to that raised in previous runs with respect to the TUAS control station – where autonomy/automation is being applied as a background task, particularly with respect to policy checking, some indication that this has happened and everything is satisfactory may be required. Notwithstanding the limited autonomy available at C2, the Mission Commander commented that "Assets used successfully" and "Platform and sensor coordination appeared to raise SA", tending to indicate that whilst more autonomy might be useful, the functionality provided was sufficient.

Policy Management scores were mid-range with the comment that operations were "Policy compliant mainly due to the operators" and that "UAS operated in allocated airspace (but none requested)". This is again indicative of the autonomy at the TUAS and SUAS control stations that attempts to ensure that all UAS operations remain within authorised (allocated) airspace was not being recognised by either UAS operators or the Mission Commander.

Negotiation scores were mid-range with the comment "Clear when negotiation required but not successful" and "Negotiations did not always work in timely manner" which indicates that there was scope for automated negotiation within clearly recognised circumstances but that the Chat work-around was not a viable solution. When undertaking the JTAC role, the operator commented that Chat was useful but that voice was a good fall-back – consistent with previous runs and comments from other operators.

In relation to MUMT, scores were mid-range with the Mission Commander/JTAC assessing that LOI2 of UAS sensors provided for greater SA and that LOI3 not used, the implication being that it was deemed to offer no additional value over LOI2 in this Use Case. A further comment was that it "Appeared control of more than 2 UAS slowed mission" and this is likely to be influenced by the amount of autonomy available, since a further comment was "If autonomy works, could increase mission performance and effectiveness".

5.0 CONCLUSIONS

The CB17 trial provided subjective results from suitably qualified and experience military participants. The complexity of the trial was a determining factor when exercising the COMPACT concept. Limited timeframes meant a short training opportunity for the military participants which led, on many occasions, to

system behaviours that were unexpected, which in turn led to lower subjective ratings – particularly for applications of Autonomy.

What became apparent was in the case of Autonomy, the policy manager was restricting execution of specific behaviours due to policy violations. This was intended, but provided a clear indication that this was not being communicated to the Military participant during the runs. Therefore, a key outcome from CB17 was the need to ensure that policy violations are more clearly alerted to the operator.

Each system's design reflects Military Advisor advice that operators would not want to be constantly distracted by information when autonomy and policy management was working with no issues/policy conflicts, but alerted when there are issues/policy conflicts. The CB17 trials results may however indicate that a compromise solution is required, perhaps similar to the "traffic lighting" of Survivability, Effectiveness, Timeliness" under previous research with "SET" greyed out when automatic behaviours and checking are not being applied (the SET concept is described in [14]).

Negotiations, particularly airspace negotiations, were deliberately accounted for in the design of the use cases, such that there would always be a need for platforms to require entry into new airspace. A limiting factor in the design of the systems was the need to use a specifically formatted freetext message to convey the negotiation (this is because there is no support for negotiations in standard digital messaging). Military Participants initially found this cumbersome however, as the trial continued, began using the negotiations as intended. What was apparent was that Negotiations were continuously being undertaken, albeit via voice communications, validating the intent to investigate automation of these in future trials of COMPACT.

The CB17 trial was in part designed to support the evaluation and assessment of the current instantiation of the COMPACT concept architecture. The trial demonstrated that a key element of COMPACT, the Policy management function, has the potential to provide a viable means of controlling and managing the application of Autonomy on complex systems and that applying negotiations provides a means to manage and mitigate and potential policy conflicts.

6.0 REFERENCES

[1] Cottrell, R. J. Thomas, M. J, Hybrid Architectures For Adaptable Autonomy, QinetiQ, November 2013.

[2] Stevens, B. Investigation of Advanced Novel Computing Techniques Interfacing with Integrated Modular Avionics, DERA UK MOD, 1998 (Unpublished)

[3] Geddes, N. D. Shalin, V. L. Intelligent Decision Aiding for Aviation, Linkoping, Sweden: Centre for Human Factors in Aviation, 1997

[4] Banks, S. B. Lizza, C. S. Pilot's Associate – A Cooperative, Knowledge-Based System Application, IEEE Expert, June 1991

[5] Miller, C.A Hannen, M. D. The Rotorcraft Pilot's Associate: design and evaluation of an intelligent user interface for cockpit information management, Knowledge Based Systems, 12 (p443-456), 1999

[6] Catford, J. R. The Mission Management Aid – A New Level of Systems Integration in Future Fighter Aircraft, IEEE Colloquium Integrating Issues in Aerospace Control, April 1991

[7] Gibson, C. P. Garrett, A. J. Towards a Future Cockpit – The Prototyping and Pilot Integration of the Mission Management Aid (MMA), AGARD Conference Proceedings No. 478 – Situation Awareness

in Aerospace Operations

[8] Howells, H. Davies, A. Macauley, B. Zanconato, R. Large Scale Knowledge Based Systems for Airborne Decision Support, Knowledge Based Systems, 12 p215-222, 1999

[9] Taylor, R. M. et al. Cognitive Cockpit Engineering: Coupling Functional State Assessment, Task Knowledge Management, and Decision Support of Context-Sensitive Aiding, in Cognitive Systems Engineering in Military Aviation Environments: Avoiding Cogminutia Fragmentosa! (eds McNeese, M.D Vidulicj, M. A) 2002

[10] ECOA – European Component orientated Architecture, www.ecoa.technology, accessed 23/08/2018

[11] Burski, L. A. *Agreement Patterns for Agent Negotiations, 2017, QinetiQ (unpublished)*

[12] A. Iglesias and Mercedes Garijo and Jose Ignacio Fernandez-Villamor and Jose Javier Duran, Agreement Patterns, 2009

[13] Luck M, McBurney P, Computing as interaction: agent and agreement technologies. In: IEEE SMC conference on distributed human-machine systems, pp 1–6, 2008

[14] Schreiber, A, Th. et al (2000), Knowledge Engineering and Management - The CommonKADS Methodology, MIT Press

[15] Cottrell R.J., Dixon D.G., Hope T. and Taylor RM. *Operator in the Loop? Adaptive decision support for military air operations*, In HCI International 2005, Proceedings of the 1st International Conference on Augmented Cognition, Las Vegas, AZ, 22-27 July 2005. Mawhah, NJ: Lawrence Erlbaum, 2005.

Policy Management and Negotiation: Enabling Effective Human Autonomy Teaming

Michael Thomas
QinetiQ
Farnborough, GU14 0LX
UNITED KINGDOM
mthomas1@qinetiq.com

Rob Cottrell
QinetiQ
Farnborough, GU14 0LX
UNITED KINGDOM
rjcottrell@qinetiq.com

Antony Grabham
Dstl
Portsmouth, PO17 6AD
UNITED KINGDOM
apgrabham@mail.dstl.gov.uk

ABSTRACT

Autonomy technology is being developed and deployed in support of military unmanned systems. However issues persist, particularly in relation to future aspirations for unmanned aircraft operations, on how to balance requirements for flight certification and attendant needs for transparency and the trust of machine autonomy by human operators and team members, with the need for increasingly complex autonomy to achieve mission objectives. With autonomy being essential to unmanned aircraft operations during certain mission phases or under specific circumstances, human autonomy teaming cannot be effectively achieved by merely managing levels of autonomy. Effective human autonomy teaming needs to include paradigms for ensuring that autonomy technologies are employed that are appropriate to the prevailing mission and operational context and which satisfy the applicable policies. Furthermore, human autonomy teaming needs to include provision for autonomy technology to electronically negotiate with controlling authorities where such negotiations are expected such that routine adaptations can be accommodated. In such cases human intervention is only required (by policy) when human authority or decision complexity demands it. This paper discusses research into a Configurable Operating Model for Policy Automation and Control of Tasks (COMPACT) which seeks to employ higher-integrity policy management technology to enable lower integrity with more complex autonomy technology invoked according to the prevailing context and with provision for electronic negotiation between operating tiers of command. The paper also summarises Live Virtual Constructive trials conducted in 2017 to test elements of the research with military participants in militarily representative scenarios.

1.0 INTRODUCTION

This paper provides an overview of a conceptual adaptable autonomy architecture that is currently the subject of a UK MOD research programme undertaken by QinetiQ on behalf of Dstl. The paper also describes a recent trial during which the architecture was tested using military personnel against the backdrop of a militarily representative scenario.

The Configurable Operating Model for Policy Automation and Control of Tasks (COMPACT) was originally conceived [1] as a means of addressing the problems associated with the employment of advanced and novel computing techniques that can deliver the complex, and often ‘intelligent’, software that underpins automated and autonomous systems.

The role of automation and autonomy software in aircraft, particularly unmanned aircraft, has significantly increased and has led to more complex sets of requirements to provide the necessary functionality to enable human operators and pilots’ greater scope and capacity for undertaking missions.

In many cases Tthese complex requirements can only be delivered through provision of novel and non-traditional software development concepts, such as Artificial Intelligence (AI) and Machine Learning (ML). This can lead to software components that are difficult to certify and therefore exploit on operational

platforms [2]. In addition, the increased software complexity can sometimes be difficult to interpret by the operators and pilots who use them, leading to issues of trust and comprehension.

For a number of decades, this potential division of roles and responsibilities between the human and the system has been subject to much research. Early efforts focussed on the role of a system providing pilots of manned systems some level of assisted, associate or coach support [3]. Examples include the US PA [4]and RPA [5], and the UK MMA [6] & [7], FOAEW TDSS [8] and Cognitive Cockpit [9].

In many of these cases, the final responsibility for decision making was retained by the human operator as they were ultimately in control of their platform. But today, with operators of unmanned systems potentially expected to control multiple platforms, and multiple platform types, the human component has much less cognitive capacity to effectively manage their responsibilities without the application of novel software able to provide the required levels of decision support, automation and autonomy. Furthermore, it is expected that there will be mission phases or specific circumstances when autonomy is routinely required as an essential component of future unmanned aircraft operations, rather than being required as a crew aid as was often the case for the research focussed on manned platforms. The COMPACT concept provides a potential means to bridge many of these issues and provide a robust mechanism for ensuring complex software exploitation and trusted decision making.

2.0 THE COMPACT CONCEPT

The COMPACT concept is intended to provide a means for monitoring and controlling Unmanned Air Systems (UAS) that necessarily employ various autonomy and automation technologies, based on a set of pre-determined (and configurable) rules.

When characterising the COMPACT concept, it is worth looking back to the human machine teaming concepts reported in [3]; the assistant, associate and coach. Within these definitions, the COMPACT concept has a role similar to that of a coach, in that its purpose is to monitor both the performance of the system and the human (in terms of how the human interacts with the system), and determines whether this performance is consistent and compliant with applicable rules and constraints.

In many respects, the COMPACT concept architecture (see Figure 1) shares many of the capabilities of the Situational Assessment Support System (SASS) and certain elements of the Task Interface Manager (TIM) developed under the Cognitive Cockpit Programme [9]. The SASS comprised a rule based system which, by understanding the pilot's cognitive state, the platform status in relation to the mission plan (and associated task model), and the environment, could determine the appropriate course of action in response to dynamic events and most appropriate means of conveying information to the pilot for the current level of autonomy (using the Pilot Authority and Control of Tasks taxonomy – PACT). In relation to COMPACT, the UAS operator's state could be an analogue of pilot's cognitive state, however the initial research focus has been on the assessment of the UAS in relation to its mission goals, mission plan and the prevailing context.

In order to ensure that COMPACT can achieve the desired aims, it requires information and knowledge that relates to the platform/system state, mission and operational context, mission plans and goals, and the task models that describe tactics, techniques and procedures (TTPs). Using this information COMPACT is will be able to monitor the platform state against the platform model, task model, mission and task plans and identify discrepancies. COMPACT will have knowledge of courses of action when discrepancies in the platform state against the platform or task model, or the sortie, mission and task plan occur. COMPACT will expedite defined actions to ensure consistency is maintained in the platform state against the platform and task models and the sortie, mission and task plans unless this conflicts with the control state.

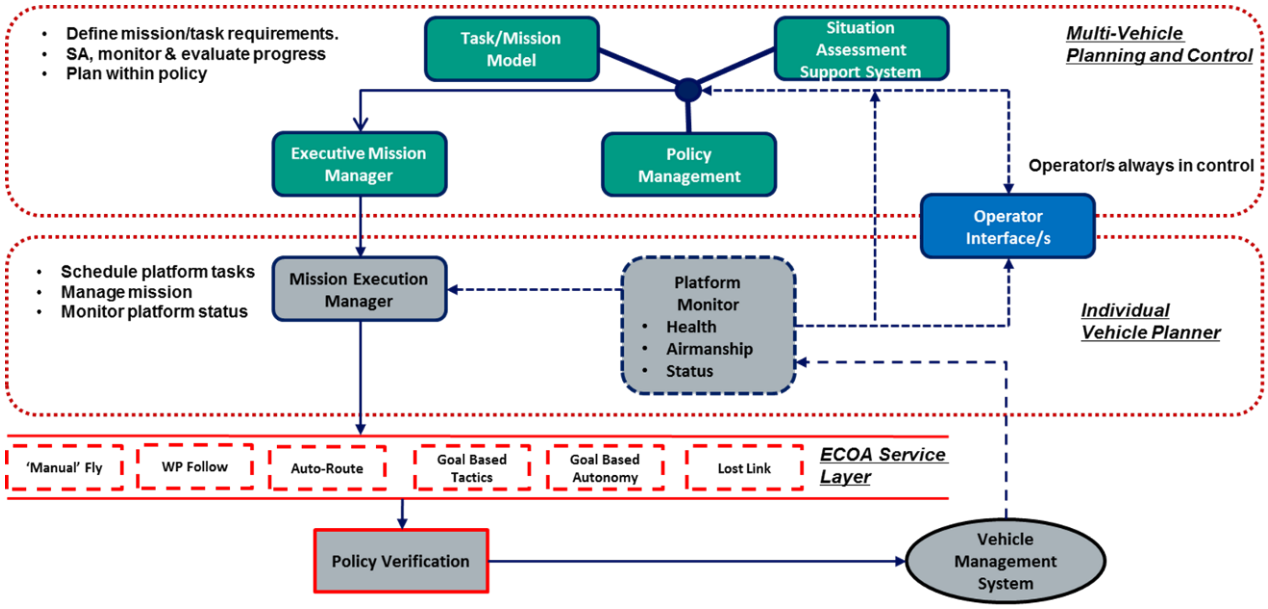


Figure 1 COMPACT conceptual architecture

COMPACT is predicated on the idea that there are a collection of software agents that can invoke various types of software components (including novel and AI techniques) able to provide platform or system decision support, automation and autonomy functions (which may be implemented using the European Component Orientated Architecture – ECOA, see [10]). The invocation of these agents is controlled by a Task or Mission Management function in response to evolving mission requirement and situation assessment.

The Task Management component continuously monitors which agents can be used through Situation Assessment (SASS) and policy management. The SASS Module will continuously monitor current status. To include:

- Environmental: terrain, landscape, weather
- Platform: position, altitude, speed, direction, fuel,
- Mission Phase– current and future (as advised by Task Manager)
- Policy Manager - Airspace / Mission Compliance; rules, constraints and Rules Of Engagement (ROE) as defined by Policy data

The Task Manager Module has knowledge of how the mission should be carried out (based on a priori knowledge) and will 'decide' which behaviours are to be conducted in order to achieve the task. These behaviours may include those delivered by the agents, the operator, or combinations thereof.

2.1 Policy Management and Negotiation

Policy Management is the means by which the performance and configuration of a system is monitored by COMPACT to ensure its operation is compliant with current policies. The term policy refers to the constraints and guidelines within which military operations are to be conducted. At the highest level, different nations apply laws and regulations that must be adhered to, such as (in the UK) the Law of Armed Conflict. At the operational level, the ROE provide a set of constraints that determine how military force should be applied. Tactical operations are managed and directed using a multitude of policies and in the case of aircraft; airspace management and the Airspace Control Order (ACO) are of particular relevance as is the

Air Tasking Order (ATO). At the platform level, Tactics, Techniques and Procedures (TTP) shape behaviours that maximise platform performance and the mission plan to ensure the desired military effect (see Figure 2).

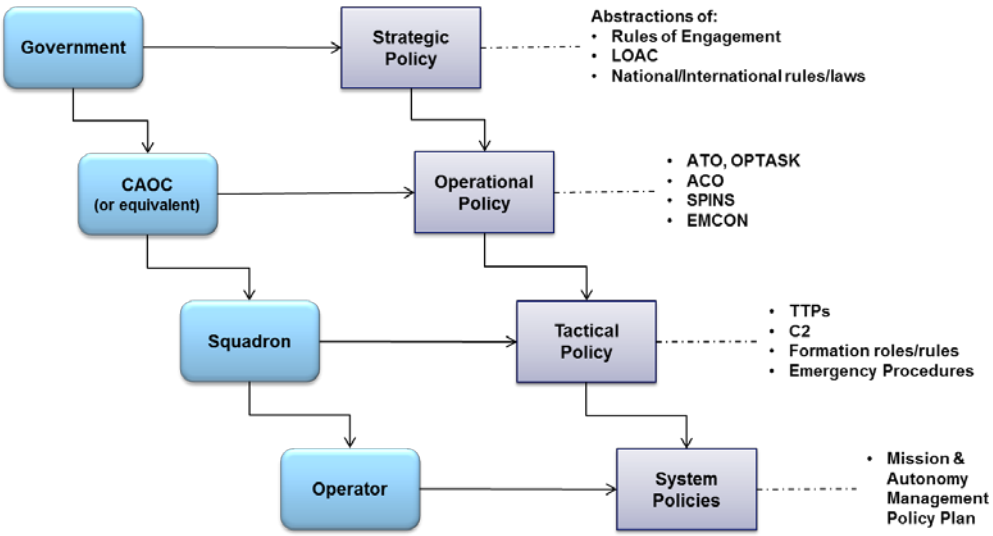


Figure 2 Policy hierarchy

As is the case in manned military aviation, there are occasions when, as a mission progresses, policies come into conflict. A typical example includes the re-planning or re-tasking of a UAS asset that requires it to enter airspace it is not permitted to enter. For manned aviation the pilot will request access to the required airspace from the relevant authority. This interaction between the pilot and the authority represents a negotiation, which is a natural consequence of policy management.

An initial model for the application of negotiations within the policy management framework is depicted in Figure 3. The detection, assessment and classification of a policy conflict is made and the determination as to whether the adjustment will require a change in Goal, Means or specific Acts will determine the process to resolve the conflict.

- A Goal adjustment is a change to the mission/task itself, and may require a negotiated resolution and approval from a higher C2 authority
- An adjustment of a Means, may require either;
  - A resolution requiring approval (from a GCS operator or Mission Commander)
  - A resolution requiring authorisation from a higher C2 authority
- An adjustment of an Act may require either;
  - An internal resolution (i.e. a minor change to the route plan)
  - A resolution requiring negotiation and approval from another agent (i.e. from a GCS operator or Mission Commander)

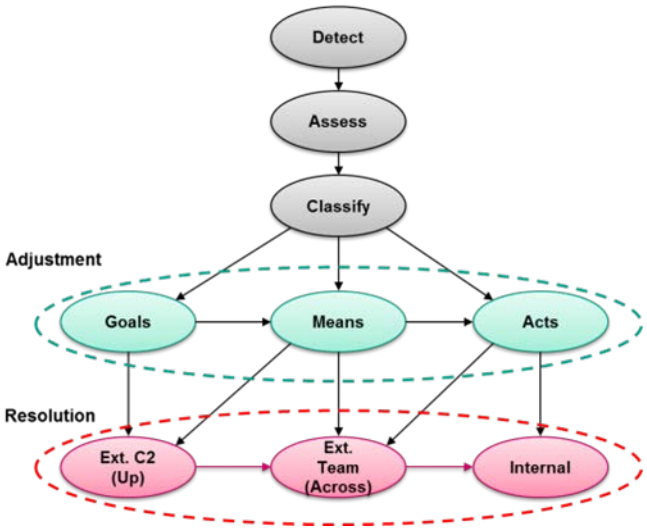


Figure 3 Policy conflict identification and resolution.

To aid the development and implementation of negotiations, as well as their validation and verification, formal methods have been applied in the form of formal agreement patterns [11]. Agreement patterns define patterns for sharing and reusing previous experience in the application of agreements (see [11] & [12]) and provide an abstract way of defining negotiations. Within COMPACT, each policy resolution can be considered to be a type of "negotiation". For instance 'platform 1' talks to 'platform 2' and exchanges some kind of information transaction forming a negotiation.

There are three parts to an agreement pattern; informal representation using classifications to direct the description, reasoning diagrams to show the elements of the agreement, and formal representations identifying the preconditions and post conditions for entering and concluding the agreement. For example, the informal representation of an Airspace Negotiation is defined below:

NAME: Auto Airspace Change
DURATION: Short Term
UAV Operator/autonomy system needs access to airspace for which there is no current authorisation (for example UAV is tracking a target that is heading outside of its currently allocated airspace and needs to relocate to maintain track)
FORCES: Pattern applicable when task requiring access to new airspace is within mission bounds (i.e. a new mission plan is not required), UAV operator/autonomy can handle any restrictions authoriser may give (such as area/height/time bounds), there is an identified authoriser (e.g. Air Battlespace Manager) and there are no other known circumstances which negate the validity of the request (e.g. emergency recovery plan invalidated due to insufficient fuel)
SOLUTION: Approver authorises the request, if there are restrictions with the approval then the UAV operator/autonomy must confirm compliance with these restrictions before entering the airspace.
EXAMPLES: COMPACT
RESULTING CONTEXT: UAV is authorised to enter the airspace to conduct task
PHASE: Agreement negotiation, Agreement enactment, Agreement renegotiation, Agreement conclusion
RELATED PATTERNS: ---

In this example,

- **Name** is a meaningful name that provides a vocabulary for identifying and discussing the negotiation agreement pattern.
- **Duration** is the time the negotiation is covering.
- **Forces** are a description of the relevant forces and constraints and how they interact with one another and with the goals and considerations to be taken into account to select a solution for a problem.
- **Solution** describes static relationships and dynamic rules describing how to realise the desired outcome.
- **Examples** are one or more sample applications of the pattern which illustrate its application and known occurrences of the pattern which help in verifying that the pattern is a proven solution to a recurring problem.
- **Resulting context** is the state or configuration of the system after the pattern has been applied.
- **Related patterns** are any other compatible patterns which can be combined with the described pattern.
- There can be any of the following **phases** in the agreement pattern:
  - **Provider selection**: Selection of the possible provider before making an agreement
  - **Agreement negotiation**: Process of establishing the conditions of the agreement
  - **Agreement enactment**: Commitments of the agreement being effective
  - **Agreement renegotiation**: Phase of changing the conditions of a previously negotiated agreement
  - **Agreement monitoring**: Process of reviewing the established agreement conditions during the agreement enactment
  - **Agreement conclusion**: Process of concluding the agreement by any party

The reasoning diagram for the Airspace negotiation uses notations defined within the CommonKADS methodology [13], specifically identifying *Knowledge Roles* and *Inferences* (Figure 4). The reasoning flow commences when the air platform needs to enter airspace which it has not been assigned (this is a knowledge role). The system then needs to assess the mission bounds, e.g. by assessing whether failing to get permission to enter this airspace will have an impact on the fuel or emergency recovery/diversion plans. Since the mission bounds need more information this is an inference role. The approver then needs to respond to the platforms request. In the example the approver is usually the airspace controller. Their response can either; allow the platform to enter the airspace, not allow the platform to enter the airspace, or allow the platform to enter the airspace with restrictions. In the example, restrictions could include duration, height, or a subset of the airspace requested. The platform then asses the approvers response and decides if it can handle the restrictions and enter the airspace. Since there is some calculation in the assessment part of the reasoning diagram this becomes and inference role. After assessing the response the platform then decides if it will enter the airspace or not.

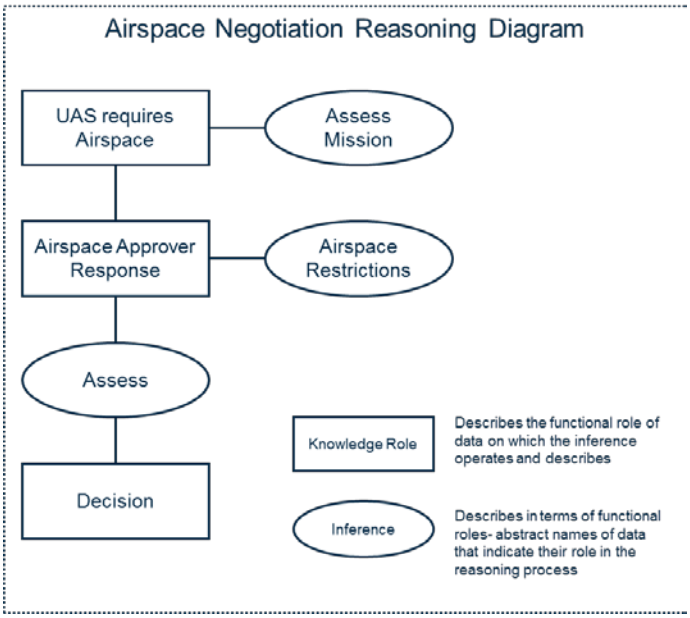


Figure 4 Airspace Negotiation reasoning diagram

The third element, formally defining the pre and post conditions, requires initial definitions of the relevant variables and their types:

variable: TYPE	**Description**
ap : ACTOR	ap is the approver
approved: (ACTOR1* ACTOR2)	Approved is a set of pairs which takes 2 actors meaning ACTOR1 is approved by ACTOR2
m: ACTOR	m is the mission bounds
flying: (ACTOR1* AIRSPACE1)	flying is a set of pairs which takes 2 actors meaning ACTOR1 is flying in AIRSPACE1
u: ACTOR	u is the unmanned air system
a_1 : AIRSPACE	a_1 is the airspace the uav is allowed to fly in
a_2 : AIRSPACE	a_2 is the airspace the uav is trying to get into
res: ACTOR -> {PARAMETERS}	res is a function which takes an actor and outputs a set of parameters (or restrictions)

The first precondition that is represented formally is that before the platform requests entry to new airspace, it must be flying in some kind of airspace already. Therefore the following applies:

$$\underline{\text{flying}(u, a_1) \wedge (\neg \text{ flying}(u, a_2))}$$

This equation represents that the platform, in this case UAS (*u*) is *flying* in airspace1 (a_1) **and** the UAS (*u*) is **not** *flying* in airspace2 (a_2). Another precondition that should be identified before entering the negotiation is that the UAS has some restriction it will be able to manage if the approver gives any. This is defined as:

$$\underline{\text{res } u \neq \{\}}$$

In this equation the variable *res* takes an actor (the UAS) as an argument. Therefore the first part of the equation *res u* represents the set of restrictions that the UAS (*u*) can manage. The second part "≠ { }",

represents that the set is not empty. Therefore the UAS has to have at least some restrictions which it is able to manage. At this abstract level of representation it is not important as to what the restrictions are, just that they exist.

The post condition of the negotiation defines the outcome of the agreement. This is represented as:

$$\underline{\text{flying}(u, a_2) \wedge (\neg \text{flying}(u, a_1))}$$

This equation shows that after the agreement, the UAV (u) is flying in airspace2 (a_2) and the UAV (u) is not flying in airspace1 (a_1). This is similar to the first formal statement produced where the UAV was flying in airspace1.

The formal notation taken from the formalising of the Auto Airspace change can be used to define a formal specification. The pre- and post-conditions in the pattern agreement denote the preconditions before entering the pattern and the post conditions after the agreement. Whereas the preconditions and post conditions in the formal specification denote the states before and after the actual process has been conducted. The preconditions used in the formal specification take the preconditions from the agreement pattern and add to them any other preconditions before conducting the action. Various notations can be used within the formal specifications for negotiations. We define the notations as follows:

NameOfSchema
Declarations
Preconditions Postconditions

The **NameOfSchema** gives a name to the process. The declarations include the variables which may be used in the process. This includes options, function names etc. each variable is assigned to a type which it belongs to. The format for a declaration is "*v*:*T*", where *v* is the variable and *T* is the type. There can be none, one or many declarations in each schema.

The preconditions define what is needed before the desired outcome is to be executed. The post-conditions define what state the variables are in after the desired outcome. Additionally, the Ξ is used to define the same variables from a previous schema negotiation.

AutoAirspaceChange
ap : ACTOR approved: (ACTOR1* ACTOR2) m: ACTOR flying: (ACTOR1* AIRSPACE1) u: ACTOR a_1, a_2 : AIRSPACE res: ACTOR -> {PARAMETERS}

flying(u, a_1) ∧ (¬ flying(u, a_2))
approved(u, m)
approved(u, ap)
res u ≠ { }
res ap ⊆ res u
flying(u, a_2) ∧ (¬ flying(u, a_1))

Three new preconditions have been added to the specification which denotes the conditions which need to be satisfied in order for the aircraft to actually change its airspace.

(1) approved(u, m)

(2) approved(u, ap)

(3) res ap ⊆ res u

Precondition (1) states that the UAV (*u*) has to been approved by the mission bounds (*m*) for it to enter the new air space. In the formal notation the mission bounds (*m*) are defined as an actor. The mission bounds relate to making sure the air space change does not conflict with the mission. If there are restrictions on the mission bounds then the higher chain of command may recommend that the aircraft should change airspace as that is a priority. Therefore if the higher chain of command makes changing of airspace a priority then it has been approved. At this abstract level there is no need to identify what the mission bounds may be only that the UAS does not change the mission and therefore does not need permission, or that it does change the mission and has been approved by the appropriate party.

Precondition (2) states that the UAV (*u*) must have been approved by the approver (*ap*). The approver would usually be the airspace controller in this example but could be another authority in other examples such as a Mission Commander.

Precondition (3) states that the restriction (*res*) the approver (*ap*) is a subset of the restrictions (res) the UAV (*u*) can manage. Thus the restrictions the UAV can manage can be in a set as follows:

{Can fly between 7000-7400 feet, can fly for 4 hours, can loiter for 20 minutes before entering, … }

The approver may allow the UAV to enter the airspace but only for 4 hours. Therefore the restrictions given by the approver can defined in the following set:

{can fly for 4 hours}

Thus the restrictions given by the approver is a subset of the restrictions the UAV can handle. At this abstract view specific restrictions are unimportant but that the UAV may have some restrictions it can manage which the approver may or may not state.

Since there are 5 preconditions which must satisfied in order for the UAV to successfully change airspace then there must be 5 outcomes which may happen if one of the pre conditions has not been satisfied. A new type OUTCOME is defined to output an outcome for each pre condition if it has not been satisfied.

OUTCOME ::= success | outWithinMissionBounds | approverRefuses | uCantHandleRestricitions | uNotInCorrectAirspace | uHasNoRestrictions

For each of the preconditions a schema is identified to show what happens if the pre-conditions are compromised.

Precondition	Schema
flying(u, a1) ∧ (¬ flying(u, a2))	**NotFlyingInCorrectAirspace** Ξ AutoAirspaceChange o: OUTCOME ¬flying(u, a_1) o = uNotInCorrectAirspace
approved(u, m)	**NotWithinTheMissionBounds** Ξ AutoAirspaceChange o: OUTCOME ¬approved(u, m) o = outWithinMissionBounds
approved(u, ap)	**NotApprovedByAirspaceHead** Ξ AutoAirspaceChange o: OUTCOME ¬ approved(u, ap) o = approverRefuses
res u ≠ { }	**UAVHasNoRestrictions** Ξ AutoAirspaceChange o: OUTCOME res u = { } o = uHasNoRestrictions
res ap ⊆ res u	**UAVCan'tHandleRestrictions** Ξ AutoAirspaceChange o: OUTCOME res ap ⊄ res u o = uCantHandleRestricitions

Since there are outcomes each time a precondition is compromised, a schema can be defined to identify when no preconditions are compromised and the UAV has successfully entered a new airspace.

SuccessfullyChangesAirspace
Ξ AutoAirspaceChange o: OUTCOME
¬approved(u, m) o = success

Using the functions defined, the entire operation of the negotiation between the UAV and other actors involved can be totalised.

TotalAirspaceChange==	(AutoAirspaceChange ∧ SuccessfullyChangesAirspace) ∨ NotFlyingInCorrectAirspace ∨ NotWithinTheMissionBounds ∨ NotApprovedByAirspaceHead ∨ UAVHasNoRestrictions ∨ UAVCan'tHandleRestrictions

For a total negotiation we can have a successful negotiation to enter an airspace (**and** a success outcome) **OR** the change wouldn't be within the mission bounds **OR** on request to enter the airspace the UAV is denied access **OR** the restrictions given to enter the airspace is more than the UAV can handle. By totalising the operation all the possible outcomes can be viewed. All exceptions have been identified and there should be no unidentified outcomes when conducting this operation.

3.0 TESTING THE CONCEPT

The COMPACT concept has been subject of several years of Dstl funded research. Earlier phases of the research programme demonstrated proof of concept in laboratory environments using representative UAS control stations, C2 and rotary and fast jet mission management systems. The most recent phase of the research programme tested the concept in a Live, Virtual and Constructive (LVC) trial

The trial, called Cardigan Bay 2017 (CB17), was undertaken to address a number of key research requirements identified by UK MOD (specifically, Dstl and DE&S) some of which directly related to Human Autonomy Teaming or Manned and Unmanned Teaming. From these requirements, additional requirements were derived to form the basis for a set of research hypothesis. Use Cases were constructed against which the hypotheses could be tested and success criteria associated with each hypothesis were derived – thereby identifying whether the trial achieved its aims. Those hypotheses that directly related to Human Autonomy Teaming and Manned Unmanned Teaming were;

- More operationally effective MUMT can be obtained when remote users, including airborne users, can exercise in-mission task-level control of UAS and/or their payloads to satisfy dynamically evolving mission requirements. This was compared against one UAS per operator)
- Autonomy provides an enabler for MUMT remote users to exercise task-level control of UAS and/or their payloads
- For autonomy in support of MUMT to be militarily exploitable it will need to be applicable within multi-tier C2 infrastructures
- Automated support for Policy Management and Negotiation is an enabler to exploiting autonomy for MUMT within multi-tier C2 infrastructures

3.1 Cardigan Bay 2017

CB17 was primarily undertaken at the Llanbedr airfield in West Wales, with additional functions undertaken at Aberporth and at the QinetiQ Farnborough Site (for geography, see Figure 5). The high level scenario for CB17 involved a Forward Operating Base (FOB) at Llanbedr, which would comprise small UAS flying (the Live element) as well as host virtual manned and unmanned platforms. A Main Operating Base (MOB) at Aberporth which would include both Constructive and Live elements. Note, Aberporth is a UK MOD range which can track air platforms flying in the in Cardigan Bay range with live track data from fast-jet traffic fused during trials together with pre-recorded data from Watchkeeper UAS sorties undertaken previously. In the case of the constructive element, recorded telemetry data was replayed in the LVC environment. Live data was also be injected into the LVC environment when aircraft are operating in the area. Farnborough

acted as the main Joint Air Operations Centre (JAOC) undertaking theatre Air Battle Management. All sites were linked through wide area network that simulated various datalink and satellite communications protocols.

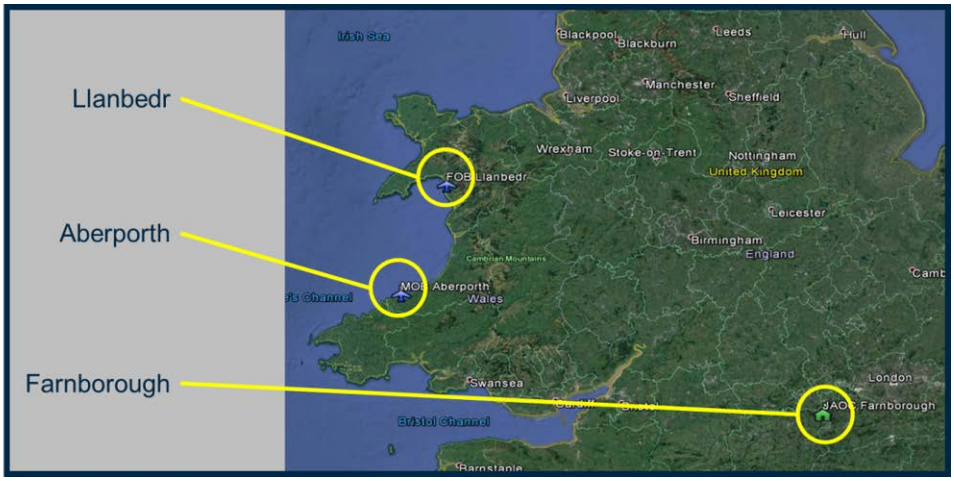


Figure 5 CB17 trial sites

The scenario used during the trial was derived from earlier activities undertaken in the West Wales and Cardigan Bay areas. Principally, it is a littoral scenario that encapsulates both maritime surveillance and land focussed Base Protection. The trials architecture, depicted in Figure 6, involved multiple platforms and stations spread between Farnborough and West Wales. Communications between stations was managed using standard datalink messaging (Link 16, STANAG 4586 and VMF) that would be applied in real operations. Many of the stations were manned by Military Participants (and some manned/operated by the technical team).

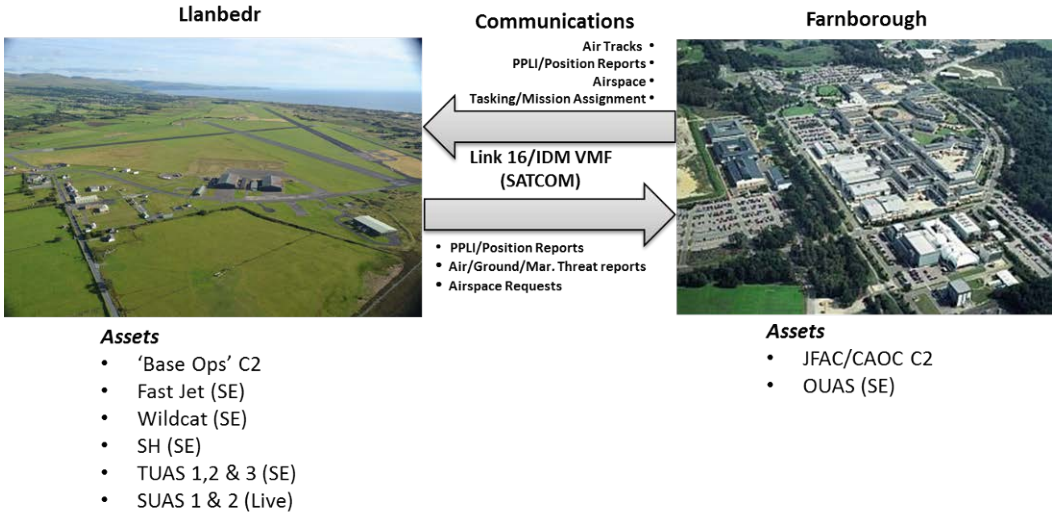


Figure 6 Trial set up

The trial design was focussed on three Use Cases that would address both littoral operations as well as Base Protection (BP). The first Use Case, in Figure 7, addressed the collection and dissemination of ISR from

different airborne platforms, some manned, but most unmanned and of different classes (size, weight and endurance). Enemy forces are operating both at sea and in-land, and so for those in tactical command of friendly forces, they must schedule and plan missions for the air assets to ensure maximum coverage of the area of interest, evolving and adapting plans and schedules as the tactical situation unfolds. In order to coordinate air assets, airspace was managed by an Air Battle Manager (ABM). As the Use Case commenced, all Air assets are allocated airspace, within which they would remain unless operators negotiate with the ABM to access to other airspace.

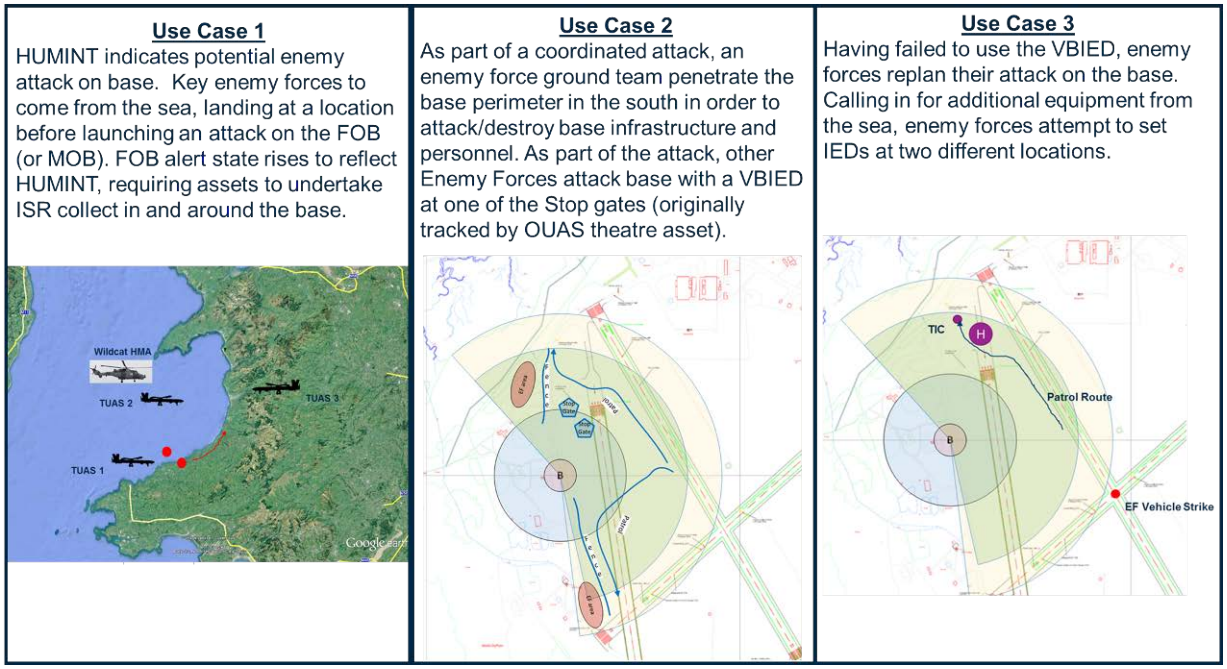


Figure 7 Use Cases for CB17

The second Use Case (Use Case 2 in Figure 7) builds on the first but includes Red Forces preparing a coordinated attack on the FOB. The focus was therefore towards the protection of the FOB through utilisation of ISR assets to locate and identify potential threats and react accordingly. This requires assessment of ISR data from multiple sources, the coordination of which spans the three sites, SUAS at Llanbedr (the Base), TUAS from Aberporth, and the OUAS operated from Farnborough (under the JAOC). The specific threats to the base included small teams attempting to infiltrate the base perimeter and a vehicle borne improvised explosive device (VBIED).

The final Use Case (Use Case 3 in Figure 7) focusses on the management of ISR assets when reacting to multiple events which occur within the base or outside or the base perimeter. This requires assessment of ISR data from multiple sources and coordination by the Base Commander to utilise assets to the best effect, ultimately requiring coordination of strike and ISR platforms to neutralise threats. In this instance, the Base perimeter is infiltrated by red forces leading to troops in contact which the Base Commander must react to.

The trials plan accounted for one assessed run for each Use Case. A training run was undertaken at the beginning of the trial to familiarise the military participants with the systems, which included instantiations of the COMPACT components.

3.2 COMPACT Assessment

The COMPACT concept is realised through provision of architectural components incorporated into

research systems that provide policy management functions and support to negotiation (via appropriate digital messaging). For the CB17 trial the systems that included COMPACT were the TUAS and SUAS control stations, manned rotary wing mission system, fixed wing mission systems, the C2 station and JTAC support system. There were 3 Military participants located at the Llanbedr trials site who were able to support the trial and operate the equipment. Their roles and systems they used in the scenario were as follows;

- Military Participant 1 was the Mission Commander at the FOB and operated the C2 and JTAC systems
- Military Participant 2 was the TUAS commander and operated the TUAS control station
- Military Participant 3 was the Rotary Wing aircraft commander (use case 1 & 2) and the SUAS operator (use case 3), and operated the Rotary Wing mission system and SUAS control Station respectively.

Each system had implemented policies and support to negotiations that were suitable for the relevant role. Details of implemented policies and negotiations are described in Table 1

Table 1. Policy Management and Negotiation support implemented on the CB17 systems.

	Policy Checks	**Negotiations**
TUAS CS	• Checking that auto-routeing (and manual routeing) conforms to airspace constraints • Checking that SAR plan conforms to SAR TTPs • Checking that Cooperative Laser Targeting with coordinating strike platform conforms to TTP and ROE • Checking that dynamic mission/route changes do not compromise Emergency Recovery Planning policy, including de-confliction and fuel planning	• Request for Airspace via Digital CHAT message (vertical cross-tier negotiation with C2) • Formatted messages for cooperative target engagement (horizontal peer-to-peer negotiation) • Digital authorisation for platform/sensor Level of Interoperability (LOI) hand-over (horizontal peer-to-peer negotiation)
SUAS CS	• Checking that auto-routeing (and manual routeing) conforms to Base Operating Procedures and airspace constraints (intended to be exercised in all use cases) • Checking that dynamic mission/route changes do not compromise Emergency Recovery Planning policy	• Request for Airspace via Digital CHAT message (cross tier negotiation with C2) • Digital authorisation for platform/sensor LOI hand-over (peer-to-peer negotiation) •

Rotary MM	• Checking that auto-routeing (and manual routeing) conforms to airspace constraints • Checking that coordinated strike behaviour with TUAS buddy-lase conforms to TTP and ROE • Warnings, Cautions and Advisories associated with TUAS platform/sensor control	• Request for Airspace via Digital CHAT message (cross-tier negotiation with C2) • Digital request for UAS platform/sensor LOI hand-over (peer-to-peer negotiation) • Digital authorisation for LOI hand-over of own EO/IR sensor to remote user (peer-to-peer negotiation) • Checking that Cooperative Laser Targeting with coordinating strike platform conforms to TTP and ROE (peer-to-peer negotiation
Fast Jet MM	• Checking that task/routes conform to airspace constraints • Checking that coordinated strike behaviour with JTAC conforms to TTP and ROE in terms of weapons effects and blue force positions	• Request for Airspace via Pre-formatted Digital message (Cross tier negotiation) • Checking that Cooperative Laser Targeting with coordinating JTAC conforms to TTP and ROE (peer-to-peer negotiation)
C2 MM	• Checking that task/routes conform to airspace constraints • Checking that task/routes are de-conflicted • Checking that air-vehicles are de-conflicted • Checking that Positive Identification (PID) is maintained during Track Handover • Checking that range, LOS, airspace and time constraints are met where a platform is employed as a Communications Relay • Checking that UAS tasking have associated emergency recovery plans that meet policy • Checking that when a Communication Relay role is to be withdrawn from a platform, it is safe/appropriate to do so • Checking that DACAS “9-Line” request can be serviced according to TTP	• Allocation of Airspace in response to digital airspace request via pre-formatted digital message (Cross tier negotiation intended to be exercised in all use cases • Track handover management • Checking that Cooperative Laser Targeting with coordinating JTAC conforms to TTP and ROE (peer-to-peer negotiation)

JTAC MM	• Checking that Airspace is allocated/de-conflicted • Checking that digital targeting messages are populated correctly • Checking that coordinated strike behaviour with TUAS buddy-lase conforms to TTP and ROE	• Allocation of airspace using J and/or K Series digital message to effector platforms (peer-to-peer negotiation) • Request for Airspace from C2 via Digital CHAT message (cross-tier negotiation with C2) • Digital request for UAS platform/sensor LOI hand-over (peer-to-peer negotiation) • Checking that Cooperative Laser Targeting with coordinating strike platform conforms to TTP and ROE (peer-to-peer negotiation)

To test the application of COMPACT concepts across these systems, suitably qualified and experienced military operators were required to operate these systems so that appropriate assessment could be undertaken. To assess the COMPACT, military operators were asked to provide a subjective assessment of the systems that they used. Principally, the assessment was concerned with measures of effectiveness and performance (MOE&P) for 4 key aspects of system capability; Autonomy, Policy Management, Negotiation and MUMT. Military operators were provided with MOE&P scoresheets to record their subjective views (Figure 8).

	Measures of Effectiveness & Performance											
	Measure of Effectiveness or Performance	Anchors (Low)	Weightings 1	2	3	4	5	6	7	N/A	Anchors (High)	Comments
Autonomy	Did the overall use of Autonomy for multiple platforms increase your SA?	*Lower SA than one person & 1 UAS Higher workload to manage UAS*									*Greater SA than one person & 1 UAS Lower workload to manage UAS*	
	Did the autonomous SUAS Search Patterns result in an increase of SA?											
	Did the autonomous SUAS Platform Scheduling and/or Planning result in an increase is SA?											
	Did the autonomous sensor coordination result in an increase in SA?											
	Did the Autonomy adaptively support the operator	*No adaption*									*Highly Adaptive*	
Policy Management	Overall, did the Policy Manager manage the Autonomy to operate within authorised parameters?	*Autonomy never operated within parameters No Confidence with Policy Manager*									*Autonomy always operated within parameters Highly confident with Policy Manager*	
	How confident were you that the UAS operated within applicable TTPS wrt match sensor to target											
	How confident were you that the UAS operated within applicable airspace											
	How confident were you that the UAS operated within applicable ROE											
Negotiation	Does negotiation enable Autonomy and PMF to dynamically adapt to meet evolving mission requirements?	*No adaption*									*Highly adaptive*	
	Was it clear when you were required to negotiate with Autonomy?	*Not obvious Unclear*									*Obvious Clear*	
	Was it clear when the Autonomy was required to negotiate with you?											
	Did the W/C/A facilitate negotiation?	*Useless Complex*									*Useful Straightforward*	
	Were required negotiations undertaken in a timely manner	*Slow*									*Fast*	
	Did/do STANAG defined digital messages support negotiation	*Undefined*									*Defined*	
	Is chat useful for undertaking Negotiation	*Not Useful Unimportant*									*Useful Vital*	
	Is Voice useful for undertaking negotiation											
MUMT	Manned Unmanned Team increases capability of the operator in conduct of a mission/task	*No appreciative increase in Capability*									*Increase in capability*	
	Did the provision of Sensor information (LOI2) increase SA	*No increase in SA*									*Improved SA*	
	Did the provision of Sensor control (LOI2 & LOI3) increase SA											
	Control of a UAS did not adversely affect operator workload and means to conduct the mission/task	*Execution of mission task impossible*									*Manageable workload*	
	Teaming with an unmanned system increased mission Performance	*Disagree*									*Agree*	
	Teaming with an unmanned system increased mission effectiveness											

Figure 8 Military Participant assessment score sheet.

Military participants undertook the assessment and completed the scoresheets after each run. The assessment process include an initial 'after action review' which acted as an opportunity to debrief and discuss the execution of the run prior to the formal assessment process. In addition to the MOE&P assessment, a formal assessment was undertaken by Dstl (results of which are reported separately).

4.0 CB17 RESULTS

The subjective results gathered from military operators are summarised in Table 2. The results generally show a steady improvement, with each aspect being rated with either low- or mid- range scores for the first

use case run, and mid- to high-range scores for the final use case run.

The scores for the TUAS GCS *Autonomy* were originally rated quite low due to the fact that the military participant using the system was unaware that the policy manager had constrained the application of the autonomy software components due to airspace policy violations. However, this is precisely what the COMPACT concept is aiming to achieve and it was clear that the means by which policy violations are communicated to the operator required additional work.

Table 2 Subjective results summary from the CB17 trial

		Use Case 1	**Use Case 2**	**Use Case 3**
TUAS GCS	*Autonomy*	Low	Low	Mid
	Policy Management	Low	Low	Mid
	Negotiation	Low	Low	Mid
	MUMT	N/A	N/A	Mid
C2/JTAC	*Autonomy*	Mid	Mid	Mid
	Policy Management	Low/Mid	Low/Mid	Mid
	Negotiation	Low	Low/Mid	Mid
	MUMT	Low	Mid	Mid
Wildcat/SUAS	*Autonomy*	Mid	High	High
	Policy Management	Mid	High	High
	Negotiation	Mid	Mid	N/A
	MUMT	Mid	High	N/A

For the first run the results from the MOE&P assessment showed that for the mission commander, airspace policy violations by UAS were not raised at the C2 station and consequently no Policy Management or automated Negotiation functionality was triggered. The ABM functionality between Mission Commander at Base Ops and TUAS operator were therefore conducted by voice, which may have contributed to the TUAS operator assessment that autonomy was not working in relation to airspace, whereas in fact any airspace that was in the TUAS control station and was attributed correctly triggered autonomy and policy management functionality including Warnings, Cautions and Alerts (WCA) – but these were overlooked/ignored by the TUAS operator during the run. MUMT was scored higher by the mission commander reflecting that he could see the platform positions and route plans for both manned and unmanned aircraft that were operating, or tasked to operate, in the airspace delegated to him by the JAOC.

The Wildcat military operator scores for autonomy, policy management, negotiation and MUMT were mid-range to high. It is likely that these reflect the relatively maturity of that particular system and the fact that

the operator had experience of using it in previous trials. Functionality exercised included instantiations of remote (TUAS) sensor feeds in integrated with the tactical map display showing remote platform stare-point and Field of View (FOV) for MUMT, together with digital LOI request and approval for remote platforms and sensors, which are a form of digital negotiation.

The TUAS operator scored autonomy low as it was not apparent to them that it was functioning. In actuality autonomy was working in the background in conjunction with policy management to produce WCA in relation to airspace, and also in relation to handling platform/sensor delegation to remote users (such as the Wildcat). The TUAS operator scored Policy Management low and this is believed again to be influenced by the operator not being clear as to what policy checking was being automatically applied behind the scenes, and only being aware of WCA when there was an issue, albeit that these were sometimes ignored. As a consequence, the operator stated that they were "not confident that that UAS operated within applicable policies" – despite this being the case (e.g. UAS staying within allocated airspace). Negotiation was scored low but this was influenced by the need to employ, for technical reasons, Chat messaging for some negotiations intended to be undertaken by preformatted datalink message. The operator found Chat messaging distracting and voice quicker. From the TUAS operator's perspective, MUMT was scored as "not applicable" because, once handed-off, they did not actively participate in MUMT with teaming exercised between RW operator and the handed-off platform. Taken in conjunction with the RW operator's scores, this should probably be taken to mean that MUMT was effective!

For the second run, the Mission Commander scored Autonomy mid to high with good SA of the Air Picture. Policy Management scores were low to mid-range because not all airspace issues were raised correctly to the operator, with the TUAS flying into unauthorised airspace. This was because of errors in the Chat message that should have triggered the automatic negotiation and airspace allocation. Negotiation scores were therefore low to mid-range. When the negotiation was repeated with the error corrected, the automated airspace allocation worked correctly and consequently there was no policy violation. As always, it was noted that voice was useful when autonomy (automatic negotiations in this case) did not work. MUMT was scored mid-range with SA assessed as having increased as a consequence of the TUAS being handed-off to the Wildcat helicopter.

The Wildcat HMA military operator scores for autonomy, policy management, negotiation and MUMT were mid-range to high, with SA assessed as having improved with hand-over of the TUAS, which correlates with the assessment of the Mission Commander. The Wildcat operator noted that control of the TUAS was achievable with simple commands.

During the third run, the TUAS operator scored autonomy higher than in the previous runs. The operator commented that the adaptive support provided by the autonomy sped up decision making, but most actions were still undertaking manually, indicating that there is potential advantage to be had from more automation/autonomy in relation the TUAS functionality exercised (which was search, track and cooperative targeting).

The TUAS operator scored Policy Management slightly higher (low to mid-range) than in previous runs, but still was not clear as to what policy checking was being automatically applied behind the scenes, and only being aware of WCA when there was an issue. Comments this time being "Operator unsure as to whether all actions were policy compliant" although the operator did qualify this with "Actions appeared to point to task (wrt airspace)". The Operator also commented "Messages conveying policies were minimal and could easily be missed" and "Alerts were clear but could also be missed" reinforcing assessments from earlier runs that some HMI development is required in relation to Policy Management.

Negotiation was also scored slightly higher than in previous runs, with the work-around of using a Chat message rather than a formatted message to request airspace now being familiar to the operator. However, the operator commented that "Chat was sometimes helpful but can slow you down depending on situation"

which is a recognised limitation of Chat and indicates that its use needs to be carefully controlled, ideally in accordance with TTP such that operators understand and are familiar with when it should be used and for what types of information (the intended mechanism for the negotiation being a computer generated, pre-formatted message).

The TUAS operator scored MUMT for the first time with scores mid-range. The operator assessed that "Target handoff was quicker" and that "Teaming sped up some processes". This is likely to reflect that the tools in the TUAS control station for supporting coordinated targeting in accordance with policy (TTP), together with the datalink integration which is a step towards automated negotiation, assisted the operator in conducting this time-critical task.

In this third run, the Mission Commander station was exercised both as a C2 station for managing airspace delegated to the Base Commander, but also as a JTAC station with support for digital targeting using the JTAC system. The Mission Commander scored Autonomy mid-range, commenting that there were "Multiple platforms but limited autonomy" which is true from the C2/JTAC perspective. C2 tools/functionality was designed to provide SA and decision support rather than autonomy. However, there were platform-to-platform de-confliction algorithms, route deconfliction algorithms in addition to the airspace de-confliction algorithms running in C2 as part of the Policy Management system.

The fact that these were not mentioned may be indicative of a similar issue to that raised in previous runs with respect to the TUAS control station – where autonomy/automation is being applied as a background task, particularly with respect to policy checking, some indication that this has happened and everything is satisfactory may be required. Notwithstanding the limited autonomy available at C2, the Mission Commander commented that "Assets used successfully" and "Platform and sensor coordination appeared to raise SA", tending to indicate that whilst more autonomy might be useful, the functionality provided was sufficient.

Policy Management scores were mid-range with the comment that operations were "Policy compliant mainly due to the operators" and that "UAS operated in allocated airspace (but none requested)". This is again indicative of the autonomy at the TUAS and SUAS control stations that attempts to ensure that all UAS operations remain within authorised (allocated) airspace was not being recognised by either UAS operators or the Mission Commander.

Negotiation scores were mid-range with the comment "Clear when negotiation required but not successful" and "Negotiations did not always work in timely manner" which indicates that there was scope for automated negotiation within clearly recognised circumstances but that the Chat work-around was not a viable solution. When undertaking the JTAC role, the operator commented that Chat was useful but that voice was a good fall-back – consistent with previous runs and comments from other operators.

In relation to MUMT, scores were mid-range with the Mission Commander/JTAC assessing that LOI2 of UAS sensors provided for greater SA and that LOI3 not used, the implication being that it was deemed to offer no additional value over LOI2 in this Use Case. A further comment was that it "Appeared control of more than 2 UAS slowed mission" and this is likely to be influenced by the amount of autonomy available, since a further comment was "If autonomy works, could increase mission performance and effectiveness".

5.0 CONCLUSIONS

The CB17 trial provided subjective results from suitably qualified and experience military participants. The complexity of the trial was a determining factor when exercising the COMPACT concept. Limited timeframes meant a short training opportunity for the military participants which led, on many occasions, to

system behaviours that were unexpected, which in turn led to lower subjective ratings – particularly for applications of Autonomy.

What became apparent was in the case of Autonomy, the policy manager was restricting execution of specific behaviours due to policy violations. This was intended, but provided a clear indication that this was not being communicated to the Military participant during the runs. Therefore, a key outcome from CB17 was the need to ensure that policy violations are more clearly alerted to the operator.

Each system's design reflects Military Advisor advice that operators would not want to be constantly distracted by information when autonomy and policy management was working with no issues/policy conflicts, but alerted when there are issues/policy conflicts. The CB17 trials results may however indicate that a compromise solution is required, perhaps similar to the "traffic lighting" of Survivability, Effectiveness, Timeliness" under previous research with "SET" greyed out when automatic behaviours and checking are not being applied (the SET concept is described in [14]).

Negotiations, particularly airspace negotiations, were deliberately accounted for in the design of the use cases, such that there would always be a need for platforms to require entry into new airspace. A limiting factor in the design of the systems was the need to use a specifically formatted freetext message to convey the negotiation (this is because there is no support for negotiations in standard digital messaging). Military Participants initially found this cumbersome however, as the trial continued, began using the negotiations as intended. What was apparent was that Negotiations were continuously being undertaken, albeit via voice communications, validating the intent to investigate automation of these in future trials of COMPACT.

The CB17 trial was in part designed to support the evaluation and assessment of the current instantiation of the COMPACT concept architecture. The trial demonstrated that a key element of COMPACT, the Policy management function, has the potential to provide a viable means of controlling and managing the application of Autonomy on complex systems and that applying negotiations provides a means to manage and mitigate and potential policy conflicts.

6.0 REFERENCES

[1] Cottrell, R. J. Thomas, M. J, Hybrid Architectures For Adaptable Autonomy, QinetiQ, November 2013.

[2] Stevens, B. Investigation of Advanced Novel Computing Techniques Interfacing with Integrated Modular Avionics, DERA UK MOD, 1998 (Unpublished)

[3] Geddes, N. D. Shalin, V. L. Intelligent Decision Aiding for Aviation, Linkoping, Sweden: Centre for Human Factors in Aviation, 1997

[4] Banks, S. B. Lizza, C. S. Pilot's Associate – A Cooperative, Knowledge-Based System Application, IEEE Expert, June 1991

[5] Miller, C.A Hannen, M. D. The Rotorcraft Pilot's Associate: design and evaluation of an intelligent user interface for cockpit information management, Knowledge Based Systems, 12 (p443-456), 1999

[6] Catford, J. R. The Mission Management Aid – A New Level of Systems Integration in Future Fighter Aircraft, IEEE Colloquium Integrating Issues in Aerospace Control, April 1991

[7] Gibson, C. P. Garrett, A. J. Towards a Future Cockpit – The Prototyping and Pilot Integration of the Mission Management Aid (MMA), AGARD Conference Proceedings No. 478 – Situation Awareness

in Aerospace Operations

[8] Howells, H. Davies, A. Macauley, B. Zanconato, R. Large Scale Knowledge Based Systems for Airborne Decision Support, Knowledge Based Systems, 12 p215-222, 1999

[9] Taylor, R. M. et al. Cognitive Cockpit Engineering: Coupling Functional State Assessment, Task Knowledge Management, and Decision Support of Context-Sensitive Aiding, in Cognitive Systems Engineering in Military Aviation Environments: Avoiding Cogminutia Fragmentosa! (eds McNeese, M.D Vidulicj, M. A) 2002

[10] ECOA – European Component orientated Architecture, www.ecoa.technology, accessed 23/08/2018

[11] Burski, L. A. *Agreement Patterns for Agent Negotiations, 2017, QinetiQ (unpublished)*

[12] A. Iglesias and Mercedes Garijo and Jose Ignacio Fernandez-Villamor and Jose Javier Duran, Agreement Patterns, 2009

[13] Luck M, McBurney P, Computing as interaction: agent and agreement technologies. In: IEEE SMC conference on distributed human-machine systems, pp 1–6, 2008

[14] Schreiber, A, Th. et al (2000), Knowledge Engineering and Management - The CommonKADS Methodology, MIT Press

[15] Cottrell R.J., Dixon D.G., Hope T. and Taylor RM. *Operator in the Loop? Adaptive decision support for military air operations*, In HCI International 2005, Proceedings of the 1st International Conference on Augmented Cognition, Las Vegas, AZ, 22-27 July 2005. Mawhah, NJ: Lawrence Erlbaum, 2005.

Human-System Interaction Standards for Robotic, Intelligent, Autonomous Systems

Dr Jonathan Vaughan Earthy
Lloyd's Register EMEA, Global Technology Centre
Southampton Boldrewood Innovation Campus
Burgess Road SO16 7QF
UNITED KINGDOM

Jonathan.earthy@lr.org

Dr Joshua Lee Downs
The Boeing Company
325 JS McDonnell Blvd, MC S306-4130
Hazelwood, MO 63042-2513
UNITED STATES OF AMERICA

Joshua.L.Downs@boeing.com

ABSTRACT

ISO TC159/SC4 Ergonomics of human-system interaction is developing Technical Report (TR) ISO 9241-810 Human-system issues of robotic, intelligent and autonomous systems to address product development involving Robotic, Intelligent, Autonomous Systems (RIAS). This TR will review the human-system issues for a range of RIAS, describe the ergonomics risks that have to be considered in the application of these technologies, and identify priorities for future standardisation work. The focus of the TR is breadth not depth and issues not answers. This paper presents the approach ISO TC159/SC4 Working Group 6 is using to identify and develop the list of issues and priorities for future standardisation work regarding RIAS. Categories of issues that highlight the significant ergonomics challenges RIAS present and the ramifications for the practice of Ergonomics/Human Factors Engineering will be discussed.

1.0 INTRODUCTION

Product development of systems with robot, intelligent and autonomous characteristics is rapidly progressing. Given the human-system issues unique to systems with these characteristics, timely guidance covering these issues is necessary to help all sectors of industry to design, field and operate first-time quality RIAS, and build appropriate trust in such systems.

The paradigm behind human-systems interaction standards so far has been that of tool use. The current 9241 series of standards are for interactive tools and the physical environment within which they are used. RIAS necessitate a new paradigm. Agents developed using these technologies will be more connected, complex, probabilistic, social, and augment human capabilities. Interaction with these agents could become a relationship, their interface a personality, and users and agents may form complex human machine teams, working together towards a shared goal.

There is an urgent need for a Technical Report from ISO explaining the existing, emerging and potential human-system issues and consequences for use and users associated with systems that have robotic, intelligent and autonomous characteristics. The scope of RIAS addressed includes:

- Physically embodied RIAS, such as robots and autonomous vehicles with which users will physically interact
- Systems embedded within the physical environment with which users do not typically interact, but which collect data and/or modify the environment within which people live or work (e.g. smart buildings)
- Intelligent software tools and agents with which users actively interact through some form of user interface

- Intelligent software agents which act without active user input to modify or tailor the system to the users behaviour, task or some other purpose (e.g. providing context specific content/information, tailoring adverts to a user based on information about them, "ambient intelligence", etc.)
- The effect of interaction of several RIAS on users
- The impact on the wider system of systems / socio-technical systems on which the use of RIAS will impact.

It is not intended to explore the philosophical, ethical or political issues surrounding robotics, artificial intelligence, machine learning, and intelligent machines or environments. These concerns are addressed in BS 8611:2016 Guide to the ethical design and application of robots and robotic systems and IEEE project P7000 Model process for addressing ethical concerns during system design. However, the study identifies where and why ethical issues need to be taken into account for a wide range of systems and contexts, and as such it provides information relevant to the broader debate regarding RIAS.

The target audience for the report is the standards development community and ergonomists involved in developing, acquiring and/or commissioning RIAS. Designers and engineers who have to consider the ergonomics point of view, futurists, researchers, technology developers, regulators and legislators may also find the document useful.

Although the report addresses a wide range of technology applications, and sector and stakeholder views on the issues, the treatment of each may not be complete. The report will list the areas of concern that remain to be investigated.

2.0 METHOD

Boundaries were not set for this study beyond analysis by example. These will come from other work (e.g. robotics, Artificial Intelligence (AI), safety) and lack of representation for some areas of application. Figure 1 presents the space explored, including the scale of system, the ramifications of the complete product lifecycle, and the maturity of application in relation to the scope of ergonomics. RIAS can be located in these dimensions. The location prompts consideration of particular use/user issues/problems. Because of the possible interpretation of "issue" as meaning a known topic for a specific project that is to be addressed by a particular Ergonomics method or technique, the combined phrase issue/problem was introduced to indicate a more general area of concern requiring further consideration.

The analysis projected forwards from current applications of technology to more connected, complex, probabilistic and non-deterministic social systems/entities/agents and human augmentation. Social in this context also included physical interaction. Applications the team considered included robots, intelligent systems and environments, and autonomous agents/systems. The range of users for each was explored. The futurology literature, as well as reports of current and planned products was considered. The range of published sources and other expertise considered was broad, and included extensive discussions and analysis by the project team. A bibliography of key/interesting sources is therefore provided in lieu of references.

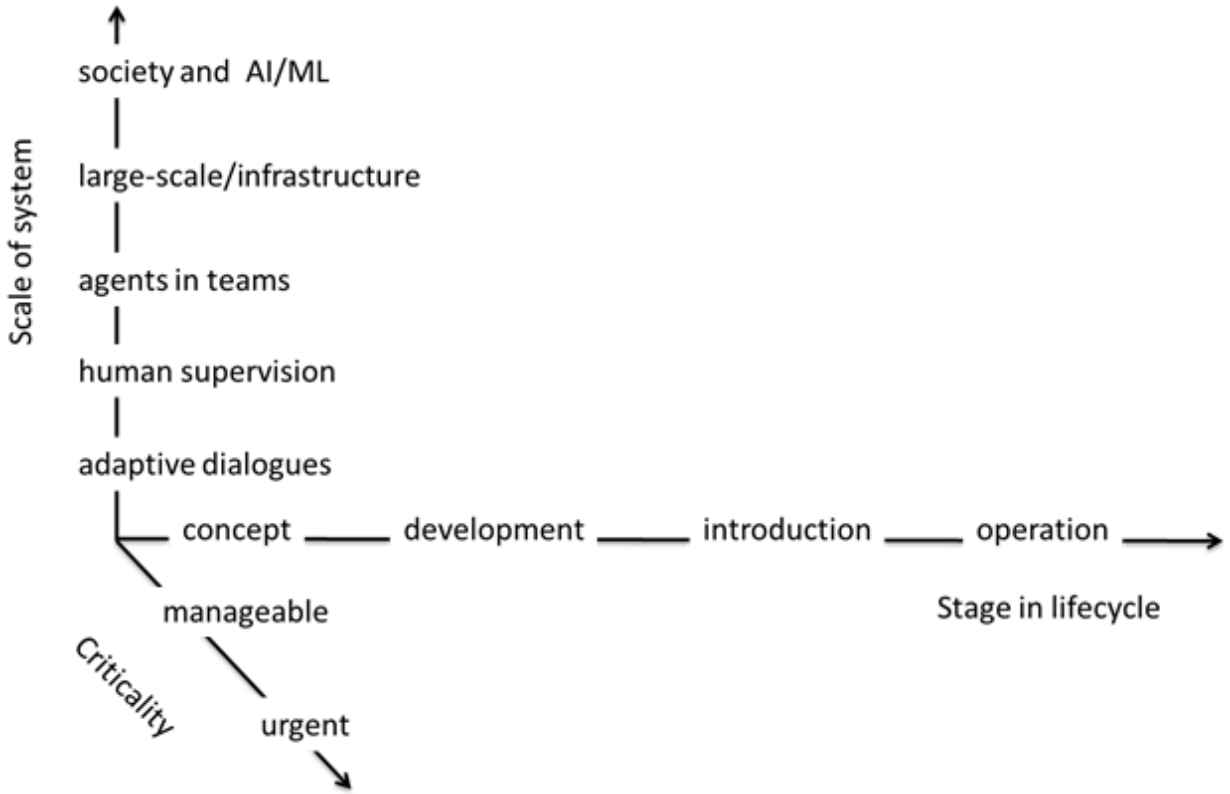


Figure 1: Dimensions considered in the review of RIAS issues/problems.

Human-system issues or problems were assessed at each level/context from society's relationship to large-scale AI (e.g. being kept as pets or not; employee, colleague, independent agent), through large-scale systems such as autonomous transport, through autonomous entities within teams, through human supervision of autonomous systems, to the details of adaptive dialogues. These analyses were based on existing research where available. The stage in the lifecycle where the issues/problems emerged and could be addressed was considered. The current level of maturity of addressing issues/problems and the urgency of addressing them were also considered.

The team considered views and concerns of users of RIAS and other stakeholders including industry sector, roles, visibility, trust, rights and culture. The limits for ergonomics were considered, with some attempt to identify areas of change. Regulatory work and astute observers were also considered.

An example of highly social interaction with autonomous entities was explored. Particular emphasis was placed on identifying what was of interest to ergonomics. The team was encouraged to go beyond the current assumed scope of ergonomics, considering what ergonomics absorbed from psychology, sociology, anthropology, and engineering. Cockpit resource management and how ergonomics contributed is an example of a successful application of domain knowledge and techniques that had not previously been considered by ergonomics. Verbal and chat interfaces, and physical interactions between people were explored, including how these interfaces change between a person and an artificial intelligence.

The issues identified were grouped under the minimum set of categories described in Section 3. The hazards and possible harm that could result were identified and are described in Section 4. The ergonomics aspects addressed by standards are described in Section 5. The necessary extensions to the scope of ergonomic analysis are explored in Section 6.

3.0 CATEGORIES

The identified human-RIAS issues/problems fall into six categories. These are summarised in Figure 2 and described in the following sub-sections.

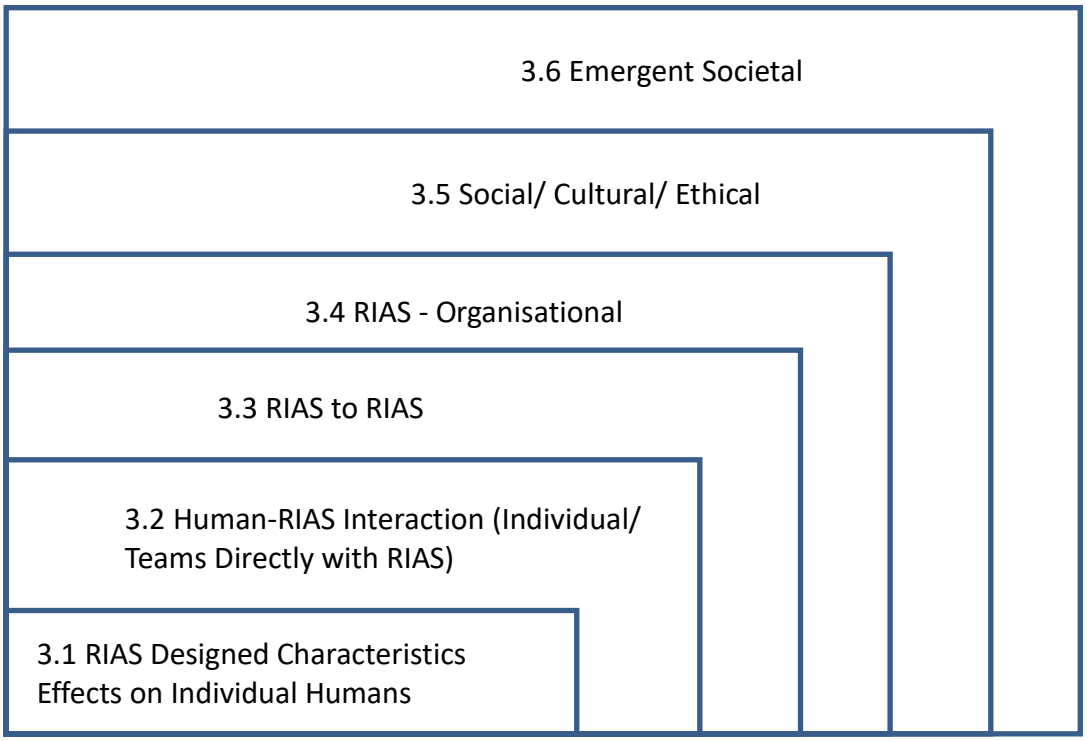


Figure 2: The six categories of RIAS issues identified.

3.1 RIAS designed characteristics effects on individual humans

This first category represents the impact that the characteristics of RIAS have on individual humans who are in the environment where it is operating. This includes physical, cognitive, affective, behavioural, and motivational responses.

3.2 Human-RIAS interaction issues (individual/teams interacting directly with RIAS)

Unless a RIAS is completely autonomous, humans interact with it directly at some level, or to some degree. The humans in this category constitute users of the RIAS. The issues in the category specifically cover the consequences and impact of the design of the user interface on user interactions with the RIAS.

1.3 RIAS to RIAS issues

One can anticipate that the time will come when multiple RIAS operate within the same environment, if it has not happened already. It is difficult to predict exactly what issues may arise, but compatibility and communication issues between one RIAS and another are, perhaps, foreseeable, and can be expected to affect users, as well as with the humans present in environments where multiple RIAS are operating.

3.4 RIAS – organisational issues

The RIAS is apt to affect the activities (e.g., work) that occurs in the organization, organizational processes, and the roles that people in the organization play. These issues raise the question of how we might optimise the existing organisational structures/working practices etc. to make best use of a new RIAS into an existing organisation. Although the issues in this category may not arise for every RIAS, they are likely to arise for any RIAS that is implemented at the level of an organization.

3.5 Social/cultural/ethical issues

A RIAS does not operate in isolation, but within a social and cultural context. To be effective, it must be seamlessly integrated within that context, as a member of the “team” or a “cooperating entity". The issues in this category describe how the design of the RIAS can impact the social interactions, group attitudes and motivation, collective group behaviours positively or negatively, depending on how seamlessly the RIAS “fits” within the social and cultural context.

3.6 Emergent societal issues

Issues will emerge during or soon after a new scientific or technological development is applied. These issues generally involve society as a whole, and are dealt with via policies, standards, regulations, and procedures. Thus, although human factors/ergonomics specialists may have a role to play in addressing some of these issues, many other types of professionals, as well as the general public, are apt to work together in addressing these issues.

4.0 HAZARD FROM RIAS TECHOLOGY IF ERGONOMICS ARE NOT ADDRESSED

4.1 The benefits of a human-centred approach to RIAS

Human interaction with RIAS agents can be predicted to take a variety of formats and, perhaps, even more that cannot be predicted. Regardless, it must be expected that many of these, if not all, will require a relationship with humans as individuals or with society as a whole. Some might even form complex human machine teams, working together towards a shared goal. Highly usable systems and products tend to be more successful both technically and commercially. RIAS developed following human-centred principles will offer several more benefits. These include:

- Improved ability for user/human to take control when required
- Increased accessibility for people with a limited range of capabilities
- Improved user experience
- Improved acceptance and integration of RIAS in human teams
- Improved integration and support of human tasks and activities
- Improved system performance across all system states (primary/normal, degraded, emergency, and reversionary (when the RIAS is unavailable)) and during the transitions between these states where particular issues can arise.

That can lead to:

- Increased product/system/service acceptance
- Increased product/system/service confidence and trust

- Reduced supporting costs
- A competitive advantage
- Reduced risk of human/human machine teaming errors and systems performing sub-optimally or entering unsafe or undesirable modes of operation.

4.2 The risks of not addressing RIAS human-system issues/problems

Not considering the human element in the design of a RIAS carries significant risks. The potential is high for negative impacts on the performance of individuals using the system and the behaviour of those otherwise affected by the RIAS. These include such things as:

- Inappropriate levels of trust/confidence in the system (having the correct level of trust or confidence in a RIAS is important. This is task, context and situation dependent, i.e. a RIAS may be more reliable in some situations than others and the ability of a user to understand this is critical, through transparency/predictability of system behaviour, etc.):
  - Under-trust can lead to: increases in workload due to additional checking and supervisory activities, users not adopting the system (if they have a choice) and anticipated benefits of the RIAS not being fully realized
  - Over-trust can lead to errors being made or missed, task disengagement and dissatisfaction with the RIAS when it doesn't perform as expected.
- Loss of situational awareness and disengagement with a task
- Over-reliance on the RIAS to complete tasks
- Increases in mental workload
- Degradation of knowledge, skills and abilities which may be required in reversionary, emergency and degraded system states or in the absence of the system
- Increases in training needs
- Perceived loss of control over one's work or one's environment, which may impact on job satisfaction and staff retention
- Ambiguity or loss of accountability and responsibility
- Poor usability impacting on sales and customer satisfaction.

These factors, in turn, can compromise safety by:

- Increasing human error, leading to accidents and injuries
- Increasing the time taken for users to reengage with a task (re-entering the loop) when they are required to do so, potentially delaying corrective action in time to prevent an unsafe state or action from occurring
- Causing people to deliberately (or unknowingly) develop ways of working around the RIAS and violating processes in order to complete tasks satisfactorily.

Similarly, when the human element is not sufficiently considered, there are negative social impacts for groups of people using the RIAS or functioning within the context of a RIAS. These include:

- General dissatisfaction with the social aspects of the environment
- Disruptions in the functioning of teams using a RIAS, or aberrations in social interactions due to the presence or functioning of a RIAS

- General diffusion of responsibility in the case of negative events related to the RIAS
- Perceived invasion of privacy by social groups and/or user teams
- Perceived decrease in security by social groups and/or user teams
- Generalized stress due to uncertainty created by the presence of the RIAS.

These things, in turn, can result in rejection of the RIAS, negatively impacting both the RIAS company's reputation and its bottom line. They can also serve to create a dysfunctional workplace, with all the difficulties this entails.

Therefore it is essential that a human-centred approach (human enhancement) be considered at all levels (of a system, organisation and scale of system), where RIAS and overall system design is focused around the decisions and tasks that must be retained by the human user (across all potential use cases: optimal, degraded, reversionary and emergency) with the RIAS supporting these. The human must be kept in the loop to the extent that is needed to achieve required levels of performance and safety across all system states. This is particularly important when systems transition between states, e.g. normal operation into an emergency state. Determining the needed extent is part of system risk assessment.

Similarly, it is critical to seamlessly integrate the RIAS with the social context in which it operates, to ensure that its operation does not unintentionally disrupt the social and cultural fabric of the environment. Thus, it is important that effect change management procedures are implemented whenever a RIAS is introduced into an environment that includes teams or other groups of people whose social interactions could be affected. Although change management is important during the introduction of any new or modified system, it is especially important for these systems to help ensure their acceptance, and minimize disruptions of group activities in the environment.

5.0 AREAS OF RIAS ERGONOMICS ADDRESSED BY STANDARDS

Standards for ergonomics already address many of the issues described in Section 3 and hazards described in Section 4, particularly those related to user interface, organisational issues and lifecycle processes. This section summarises how these standards can be used to address RIAS.

5.1 Principles of Ergonomics

ISO 26800:2011 *Ergonomics – General approach, principles and concepts.*

ISO 26800 describes the general ergonomics approach and specifies basic ergonomics principles and concepts applicable to the design and evaluation of tasks, jobs, products, tools, equipment, systems, organizations, services, facilities, and environments. The underlying principles of ergonomics remain the same, although the relative emphasis placed on them will vary. The principles and concepts described in this standard are fundamental to the design process wherever human involvement is expected, in order to ensure the optimum integration of human requirements and characteristics into a design.

There are a number of standards on ergonomics and human factors based on these principles and concepts which can be used by managers, engineers, and designers in selecting, designing, and managing systems and equipment to ensure that they are effective, efficient, and satisfying to use. These are outlined below.

ISO 27500:2016 *The human-centred organization – Rationale and general principles.*

This International Standard draws on that extensive body of ergonomics and human factors knowledge and presents the rationale and general principles of human-centredness in a concise form for *inter alia* executive board members. It explains the principles which characterize a human-centred organization. These

principles are as follows:

- Capitalize on individual differences as an organizational strength
- Make usability and accessibility strategic business objectives
- Adopt a total system approach
- Ensure health, safety, and well-being are business priorities
- Value employees and create a meaningful work environment
- Be open and trustworthy
- Act in socially responsible ways.

These provide a framework for organisational behaviour when RIAS are considered or implemented.

5.2 Human-Centred Design Process

ISO 9241-210:2010 *Human-centred design for interactive systems*.

ISO/FDIS 9241-220:2018 *Processes for enabling, executing and assessing human-centred design within organisations*.

ISO/TS 18152:2010 *Specification for the process assessment of human-system issues*.

Human-centred design is an approach to interactive systems development that aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, and usability knowledge and techniques. This approach enhances effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability; and counteracts possible adverse effects of use on human health, safety and performance.

ISO 9241-210 provides requirements and recommendations for human-centred design principles and activities throughout the life cycle of computer-based interactive systems. It is intended to be used by those managing design processes, and is concerned with ways in which both hardware and software components of interactive systems can enhance human–system interaction. The approach described complements existing systems design approaches. It can be incorporated in approaches as diverse as object-oriented, waterfall and rapid application development.

The principles and activities of a human-centred approach to design are elaborated in two human-centred design process models. 9241-220 describes the processes that ensure human-centred quality of interactive systems and 18152 describes the processes that address human-system issues in the engineering of systems. The processes in these standards go a lot further than the current norm in HCD practice (e.g. saying that HCD people must get involved early, defining the user requirements and then drive the system/technical/platform level requirements) and will apply to RIAS as much as to other types of system and any set of human-systems issues.

5.3 Interaction and Interface

ISO 9241-110:2006 *Dialogue principles*.

ISO 9241-112:2017 *Principles for the presentation of information*.

ISO 9241-13:1998 *User guidance*.

ISO 9241-129:2010 *Guidance on software individualisation*.

These parts of ISO 9241 set general design principles the ergonomic design of interactive systems and information presented without reference to situations of use, application, environment or technology. They

provide a framework for applying those principles to the analysis, design and evaluation of interactive systems. While they are applicable to all types of interactive systems, they do not cover the specifics of every context of use (e.g. safety critical systems, collaborative work). The principles are generally independent of any specific design style or application. Part 129 specifically addresses adaptive dialogues. Other parts address novel and multiple modalities.

A dialogue is the "*interaction between a user and an interactive system as a sequence of user actions (inputs) and system responses (outputs) in order to achieve a goal*", where user actions include not only entry of data but also navigational and other (control) actions of the user. Principles relating to dialogue and individualisation help prevent usability problems such as:

- Additional unnecessary steps not required as part of the task
- Misleading information
- Insufficient and poor information on the user interface
- Unexpected response of the interactive system
- Navigational limitations during use, and
- Inefficient error recovery.

Principles related to information and guidance provide a variety of benefits to users including improvements in speed, accuracy, mental effort, and user experience. They also help prevent users from experiencing usability problems with presented information. Examples of such problems include the following:

- Not detecting information despite it being present
- Other information distracting from information they are focusing on
- Not discriminating between pieces of information, since they appear to be identical
- Misinterpreting information, since the meaning of the information is ambiguous
- Expending unnecessary time in understanding information, since the information is presented is unnecessarily lengthy
- Unknown conventions leading to lack of understanding information.

5.4 Workspace and Workload

ISO 6385:2016 *Ergonomics principles in the design of work systems*.

ISO 9241-2:1992 *Guidance on task requirements*.

ISO 10075 *Ergonomic principles related to mental workload (Parts 1-3)*.

These standards provide ergonomic principles for the design of tasks, work and work systems intended to encourage balanced attention on human, social and technical requirements in a balanced manner during the design process.

"Work system" covers a large variety of working and leisure situations, including permanent and flexible work places. Work systems involve combinations of people and equipment, within a given space and environment, and the interactions between these components within an organization. Work systems vary in complexity and characteristics.

The principles specified in these standards support design of optimal working conditions with regard to task performance, workload, human well-being, safety and health; including the development of existing skills and the acquisition of new ones - while taking into account technological and economic effectiveness and

efficiency.

Technological, economic, organizational and human factors affect task performance, behaviour and well-being of people as part of a work system. Applying ergonomic knowledge in the light of practical experience in the design of a work system is intended to satisfy human requirements including changes to type and measurement of workload.

The systems approach in these standards assists in both existing and new situations, such as introduction of RIAS. Ergonomic evaluations of existing or new work systems will show the need for, and encourage attention to, workload and the role of the worker with/within those systems. Following the principles and requirements described in these standards will support management in making better decisions, for instance related to the sustainability of investments in RIAS and work system innovation.

5.5 Context and Environment

ISO 9241-11:2018 *Usability: Definitions and concepts*.

ISO 11064 *Ergonomic design of control centres (Parts 1-7).*

Standards from TC159/SC5 *Ergonomics of the thermal/physical environment*. For example, ISO TR 15742 *Determination of the combined effects of environmental components on people*; ISO/DTS 16418 *Mathematical model for predicting and evaluating the dynamic human physiological responses to the thermal environments*; ISO 28803 *Application of international standards to people with special requirements*.

ISO 9241-11 provides a framework for understanding the concept of usability and applying it to situations where people use interactive systems (including RIAS), and other types of systems (including built environments), and products (including industrial and consumer products) and services (including technical and personal services). Usability is a scalable, task-based measure of the degree to which users are enabled to achieve goals effectively, efficiently and with satisfaction, taking account of the context of use. 9241-11 explains how usability can be interpreted in terms of user performance and satisfaction, and emphasizes that usability is dependent on the context of use, the specific circumstances in which a system, product or service is used.

Control and control centres present an early and widespread application of RIAS. The overall strategy for dealing with user requirements in control centres is presented in ISO 11064-1. ISO 11064-2 provides guidance on the design and planning of the control room in relation to its supporting areas. Requirements for the layout of the control room are covered by ISO 11064-3. Ergonomic requirements, recommendations and guidelines for the design of workplaces in control centres are established in ISO 11064-4. Displays and controls, human computer interaction and the physical working environment are presented in ISO 11064-5 and ISO 11064-6. Evaluation principles are dealt with in ISO 11064-7.

Standards for modelling, measuring and assessing the impact of properties of the physical and thermal environment are defined by ISO TC159/SC5. As RIAS increase their presence in environments, in particular controlling dynamic integrated environments, personal capability, comfort and safety will increasingly depend on machine application of these standards and correct interpretation of the integrated effect of environmental factors.

6.0 CHANGES REQUIRED FOR ERGONOMICS TO BETTER ADDRESS RIAS TECHNOLOGY

Ergonomics involves the design and evaluation of the user interface and its context. The ergonomist develops and studies the concept of operations for a system, defines the tasks to be performed, and allocates

those tasks to the user, the machine, or some combination of user and machine. The physical presentation and layout of controls, the content and format of information displayed to the user, and the physical workspace are of interest to the ergonomist.

The toolbox of principles and techniques applied by the ergonomist to develop user interfaces in a system has evolved to support systems that function according to inputs provided by the user in context with engineered models of the environment in which the system is to be operating. Automation runs on such models, often following scripted logic for performing tasks the user would have otherwise had to perform.

The toolbox of principles and techniques applied by the ergonomist will need to be modified to address the unique interaction challenges autonomy presents. As automation extends to become more autonomous, dynamic, probabilistic behaviours that are more adaptive to the operational needs of a system, merely concerning ourselves with the design and layout of control panels will be insufficient to define the user interaction with the system. That is we need to use human-centred design approaches that are consistent and linked to the wider systems engineering and design processes for RIAS (HF requirements identification, architecting, modelling, design, human in the loop test and acceptance processes etc.). Integration with engineering, management and design are already addressed in ISO TS 18152 and ISO 9241-220.

With advanced autonomy that functions like another person or team of people might when performing the same task, collaborative awareness and crew resource management are examples of constructs and techniques from teaming and supervisory command and control that may apply. Likewise, abnormal situation management may also provide relevant guidance.

Like the use of stereotypes and affordances, biases in social expectations and social contracts of a specific user population will be essential for the developers of autonomy to consider and control for in their designs. This may be as simple as ensuring the proper use of turn signals in an auto-driving car, or as complex as deciding the level of medical care to be given to an elder.

These are just a few examples of areas of interest that may be important to the ergonomist dealing with RIAS analysis and design. Our consideration of issues, principles, and techniques that may be relevant to ergonomists dealing with the development of autonomy needs to include that middle ground between what we know about how to design effective interfaces between people and automatic systems, and the industrial and social psychology of teaming and supervisory command and control. Other domains like sociology and anthropology might have principles and techniques that could prove useful to ergonomists dealing with autonomy, as could computer science.

The areas in which Ergonomics will need to develop are listed below:

- Social aspects of RIAS and effects on society:
  - Highly social interaction
  - Cultural ergonomics
  - ISO personality dimensions for RIAS
  - Organisational psychology/ergonomics
  - Management of emergent behaviour.
- Human State Measurement and models:
  - Human State Measurement
  - Human Models

    - General models of humans
    - Models of specific users/user groups.
- Human machine teaming
    - Who is the user?
    - Dynamic function allocation between humans, machines
    - Dynamic Goal Setting
    - Dynamic task allocation between human team members
    - Dynamic task allocation between humans and machines
    - Dynamic task allocation between machines
    - Transparent interaction and transparent users
    - Knowledge acquisition/management
    - Small data *vs* big data – learning / evolving systems vs models.
- Guidance and RIAS:
    - Transparent systems
    - Automated education
    - Training users of RIAS systems
    - Design of systems for zero training.
- Principles for RIAS ergonomics.

7.0 DISCUSSION AND CONCLUSIONS

7.1 Human Autonomy Teaming in this context

Table 1 assigns Symposium presentations to the RIAS human-system issue categories. The purpose is to illustrate the categories using the terminology and concepts of human autonomy teaming.

This distribution is not surprising, being close to that seen in the source literature studied in our analysis. The focus of ergonomics seems to be on the things that we can address in a physical and engineering context, and to dabble a bit with regard to organisation and teaming, but the vast body of social science is not applied by ergonomists. The analysis of human-system issues and consequential risks reported above indicates that this approach will not suffice for RIAS. Examples include: requiring workers to collaborate with robots in the workplace, or development of collaborative robots; the changes to pedestrian behaviour and transport infrastructure required by self-driving cars; changes in the construction of personal privacy and security brought about by the internet; changes in human behaviour with increasing ambient intelligence.

Table 1: Symposium papers by RIAS issue category.

Issue category	Symposium paper number
1. RIAS designed characteristics effects on individual humans	2, 4, 10, 11, 12
2. Human-RIAS interaction issues (individual/teams interacting directly with RIAS)	1, 3, 5, 6, 7, 8, 9, 13, 16, 23, W2

Issue category	Symposium paper number
3. RIAS to RIAS issues	21
4. RIAS – organisational issues	K1, W1, W3
5. Social/cultural/ethical issues	14, 15, 20, 22
6. Emergent societal issues	

7.2 Conclusions

This paper has described the scope and methodology employed by ISO TC159/SC4 Working Group 6 to identify established, emerging, and potential human-system issues with the introduction and use of RIAS technologies. Summaries of the categories of these issues, and benefits and hazards for use and users associated with systems that have RIAS characteristics were also described. An alignment of these categories to this symposium's papers was presented to demonstrate the broad and deep relevance of the issues to the wider community of practitioners and researchers. Likewise, existing ISO standards that may address the RIAS ergonomics issues were identified.

Working Group 6 has identified a number of issues and categories of issues that will likely require that HFE and the greater HSI community extend the principles and techniques they have become comfortable with if the study and practice of ergonomics is to remain relevant for future RIAS technologies. To name a few, methodologies in social and organizational psychology, sociology, anthropology, and computer science may prove useful in defining the interaction requirements for RIAS' connected, complex, probabilistic, social, and human augmentation characteristics across the international stage.

Next steps for Working Group 6 include exploring the degree to which existing ISO standards address the issues, and how the standards may need to be updated or new standards defined. The characteristics and ramifications of an ISO standard personality for RIAS will continue to be explored. Publication of a technical report detailing the group's findings will then follow. Wide dissemination of the findings is required to notify the existence of human-systems issues in relation to RIAS, and to demonstrate the existence of available solutions. At the same time the Ergonomics community have to be persuaded of the need to extend the scope of practice to address a new scope of interaction.

8.0 ACKNOWLEDGEMENTS

Standards are team efforts, but traditionally anonymous. The contributions of the 9241-810 project team within ISO TC159/SC4/WG6, and the new members and liaisons that joined for this work are gratefully appreciated.

The authors thank their respective organisations for their support in ISO work. This paper presents the personal opinions of the authors and not those of their respective organisations or of ISO as an organisation. This document is not an ISO International Standard. It is made available for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

The authors wish to assert copyright on this work. Portions of this paper may be used in future ISO documents.

9.0 BIBLIOGRAPHY

Section 2 outlines the information meta-analysis applied during the study and why references are not cited. A bibliography of key documents and other information items is provided instead for interested parties.

[1] "The Role of Human Causal Factors in U.S. Army Unmanned Aerial Vehicle Accidents" AARL Fort Rucker AL, USAARL Report No. 2004-11

[2] Beer, J.M., Fisk, A.D., Rogers, W.A., "Toward a Framework for Levels of Robot Autonomy" in Human-Robot Interaction, Journal of Human-Robot Interaction, Vol. 3, No. 2, 2014, Pages 74-99, DOI 10.5898/JHRI.3.2.Beer.

[3] Billings, Charles E. "Human-Centered Aircraft Automation: A Concept and Guidelines." NASA Technical Memorandum 103885, Aug. 1991

[4] Bjorn-Andersen, N. 'Are "Human Factors" human?', Contribution to Man Machine Integration, State of the Art Report, Pergamon Infotec, Jan 1985

[5] Carroll, J.M., Neale, D.C., Isenhour, P.L., Rosson, M.B., McCrickard, D.S. (2003) "Notification and awareness: synchronizing task-oriented collaborative activity." Int. J. Human-Computer Studies 58, p.605-632.

[6] Carroll, J.M., Rosson, M.B., Convertino, G., Ganoe, C.H. (2006) "Awareness and Teamwork in Computer-Supported Collaborations." Interacting with Computers, Volume: 18, Issue: 1.

[7] Charalambous, G., Fletcher, S., & Webb, P. (2015). "Identifying the key organisational human factors for introducing human-robot collaboration in industry: an exploratory study." The International Journal of Advanced Manufacturing Technology, 81(9-12), 2143-2155.

[8] Chetouani, M. Advancing intuitive human-machine interaction with human-like social capabilities for education in schools (ANIMATAS) EU HORIZON 2020 project, list of PhD positions. 2018 http://www.animatas.eu/index.php?perma=Positions, accessed 2018-09-02

[9] Christoffersen, K., Woods, D. 1. "How to make automated systems team players", Advances in Human Performance and Cognitive Engineering Research, Volume 2. Emerald Group Publishing Limited, pp.1 – 12, 2002

[10] Cummings, M.L., Mitchell, P.M., Sheridan, T.B. (2005 - DRAFT) "CH XX: Human Supervisory Control Challenges in Network Centric Operations." https://pdfs.semanticscholar.org/69f6/13a50a6483ea3a0ef5ac2fde7e853a74b7e8.pdf, accessed on 2018-09-05

[11] Cummings, M.L., Platts, J.T., Sulmistras, A. (2006) "Human Performance Considerations in the Development of Interoperability Standards for UAV Interfaces." Moving Autonomy Forward Conference 2006, De Vere Hotel Belton Woods, Lincoln, UK.

[12] "Mission Command: Command and Control of Army Forces" FM 6-0, Department of the Army, August 2003.

[13] De Backer, K., DeStefano, T., Menon, C., & Suh, J. R. (2018). "Industrial robotics and the global organisation of production." oecd-ilibrary.org.

[14] De Stafano, V., "Negotiating the algorithm", Automation, artificial intelligence and labour protection,

EMPLOYMENT working paper no. 246, Employment Policy Department, International Labour Organisation. 2018

[15] Fukuzumi, S. et al, 2018, "Extraction of new guideline items from the view point of ELSI (Ethics, Legal, Social Issues) for service utilized AI" – focus on healthcare area, 32nd domestic conference on Japanese society for Artificial Intelligence, 3H1-OS-25a-04 (in Japanese).

[16] Future of Privacy Forum, "Unfairness by algorithm: distilling the harms of automated decision making" Dec. 2017. Future of Privacy Forum, fpf.org, Washington, DC.

[17] Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, Å. (2003). "Key principles for user-centred systems design." Behaviour and Information Technology, 22(6), 397-409.

[18] Hollnagel, E., & Woods, D.D. (2005). Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. CRC Press.

[19] Innovate. UK technology strategy board report: The UK Landscape for Robotics and Autonomous Systems, 2015

[20] ISO/IEC JTC1/SC7 2017 Study group on Autonomous Systems and Ambient Intelligence Environment Engineering (ASAIEE).

[21] Kaber, D.B., & Endsley, M.R. (1998) "Team Situation Awareness for Process Control Safety and Performance." Process Safety Progress, 17(1), Spring 1998, p.43-48.

[22] Knight, W. "Intelligent Machines: Forget Killer Robots - Bias is the Real AI Danger." MIT Technology Review, 3 Oct. 2017

[23] Knott, B., MINERVA research initiative topics of interest, B. Special Interest Area 1: Understanding the social impact of autonomy. 2017 https://minerva.defense.gov/Portals/47/documents/Research_Topics/2017%20Topics%20for%20website%20document.pdf?ver=2017-01-30-163533-870, accessed 2018-09-02.

[24] Moray, N. "Human factors in process control". In G. Salvendy (Ed.), Handbook of human factors and ergonomics (2nd ed.). New York: Wiley. 1997

[25] Ministry of internal affairs and communications in Japan: draft AI R&D guidelines, The Conference towards AI Network Society, 2017. http://www.soumu.go.jp/main_content/000507517.pdf, accessed 2018-09-02.

[26] Nunes, A. Bryan Reimer, B. Coughlin, J.F. (2018) "People must retain control of autonomous vehicles." Nature Vol 556, p169-171

[27] "Autonomy levels for unmanned systems (ALFUS) framework Volume II: Framework Models Version 1.0." NIST Special Publication 1011-II-1.0

[28] "Autonomy levels for unmanned systems (ALFUS) framework Volume I: Terminology Version 2.0." NIST Special Publication 1011-I-2.0.

[29] "Defense Science Board Study on Unmanned Aerial Vehicles and Uninhabited Combat Aerial Vehicles." OUSD (ATL), February 2004.

[30] Parasuraman, R. (2000). "Designing automation for human use: Empirical studies and quantitative models. Ergonomics, 43, 931-951.

[31] Parasuraman, R., and Mouloua, M. (Eds.) (1996). Automation and human performance: Theory and application". Mahwah, NJ: Erlbaum.

[32] Parasuraman, R., and Riley, V. (1997). "Humans and automation Use, misuse, disuse, abuse". Human Factors, Vol 39(2), 230-253.

[33] Parnas, D.L, 2017, "The real risks of artificial intelligence", Comms ACM, Vol 60, no 10, p 27-31.

[34] "Enhanced Situational Awareness for Robotic and EVA Operations." SAE/TP 2007-01-3231, 2007

[35] "Collaborating Human – Robot Swarms to Achieve Robust Space Exploration Capabilities." SAE/TP 2006-01-2022

[36] "Toward Human-Robot Interface Standards: Use of Standardization and Intelligent Subsystems for Advancing Human-Robotic Competency in Space Exploration." SAE/TP 2006-01-2019

[37] Scholtz, J.C. "Human-Robot Interactions: Creating Synergistic Cyber Forces." NIST. 2002

[38] Sheridan, T. B. Humans and automation: System design and research issues. New York: Wiley. 2002

[39] Sheridan, T. B., and Parasuraman, R. (2006). "Human-automation interaction". In R. Nickerson (Ed.), Review of human factors and ergonomics. Santa Monica, CA: Human Factors and Ergonomics Society.

[40] Sherwood Jones, B. Earthy, J. "Using Human-Centred Design to mitigate the risk of human error in software intensive systems", Safety Critical Systems Club Bulletin, Jan 2016, P 11-15.

[41] Totterdell P.A., Norman M.A, and Brown D.P, 1987, "Levels of adaptivity in interface design", Human-Computer Interaction - INTERACT'97, Bullinger and Shackel eds. IFIP.

[42] UK MOD, "Joint Concept Note 1/18: Human-Machine Teaming." Development, Concepts and Doctrine Centre, Ministry of Defence. May 2018

[43] "Command and Control" MCDP 6, U.S. Marine Corps, PCN 142 000001 00.

[44] "UAS Ground Control Station Human-Machine Interface Development and Standardization Guide" Version 1.0: USDA Technology and Logistics, 2012

[45] Woods, D. (2016) The Risks of Autonomy: Doyle's Catch. Journal of Cognitive Engineering and Decision Making. 10. 10.1177/1555343416653562.

Automation as an Intelligent Teammate: Social Psychological Implications

José Kerstholt, Jonathan Barnhoorn, Tom Hueting & Lotte Schuilenborg
TNO
PO Box 23
3769 ZG Soesterberg
THE NETHERLANDS

Jose.Kerstholt@tno.nl

ABSTRACT

In order to increase the effectiveness of NATO operations with flexible and robust human-autonomy teams, insight is needed into the psychological mechanisms underlying teaming with artificial agents. We conducted a literature scan to provide an overview of current knowledge on factors affecting both objective and subjective outcomes of human-autonomy teaming. The main results from our study are that current research mainly addressed factors affecting overall task performance, like situation awareness, and neglected relational aspects and human appraisal. However, social outcomes ensuing from two-way interactions will become more and more important with increased agency of AI. We therefore argue that in order to fully utilize the potential of autonomous systems, more insight is needed into (unconsciously operating) relational aspects, and how they interact with task-related factors.

1.0 INTRODUCTION

Lieutenant Bob is performing an overt reconnaissance mission in a hilly rural area together with his soldiers. The squad uses multiple motorised vehicles to transport itself. At some point in time a swarm of unmanned aerial vehicles (UAVs) that accompanies them locates enemy personnel. The swarm, communicating through an artificial agent called Smart Bud, transmits the enemy positions to all friendly personnel and systems in the area. This includes a 60mm mortar crew which immediately starts preparations to potentially engage the enemy. During this procedure, Smart Bud informs Lieutenant Bob of the 60mm mortar crew's selected targets and advises Lieutenant Bob to cancel this plan on the basis of a disproportional amount of expected collateral damage. Instead, re-planning is advised. Based on his prior combat experience and the challenging operational situation at hand, Lieutenant Bob disagrees with Smart Bud's advice. He informs Smart Bud about his concerns, resulting in a short dialogue about pros and cons of both courses of action.

Artificial Intelligence (AI) has brought about major changes to almost every aspect of our lives and although expectations differ wildly, it may reach a point where it equals or even surpasses human intelligence (Grace, Salvatier, Dafoe, Zhang, & Evans, 2017). Approaching this point requires rethinking our relationship with technology, as AI is going beyond intelligent (decision) support by augmenting human knowledge and skills and facilitating unprecedented capabilities. The scenario above illustrates a situation in which AI (the Smart Bud) is on equal footing with the human when making consequential choices. In other words, Smart Bud has become a team member with a high level of agency (Nyholm, 2017). What would the dialogue between the Smart Bud and Bob look like regarding arguments being exchanged, or even more advanced, regarding social-skills being applied to convince the other team member of the supremacy of one's position? Would it affect the actual decision being taken?

Even though military decision-making is preferably considered as a rational process (e.g., Boyd, 1995), in reality social-psychological mechanisms like team cohesion are central to team functioning (Salas, Sims, & Shawn Burke, 2005). Thus, it may well be that an interaction between a human and AI as described above

triggers such psychological mechanisms, which therefore need to be taken into account when designing AI. How will humans react to team members being replaced by AI? Will it affect their performance? Their motivation? Will we see AI as an actual team member or will we always see it as a 'support' system? The present paper tries to explore some issues related to this new area of AI.

2.0 STATE OF THE ART

The present paper reports the results of a literature review which was meant to identify knowledge gaps with regard to (social-) psychological factors that are relevant to teaming with AI that has high levels of agency and assuming the role of a team member. This is often referred to as a human-autonomy team (HAT). We have organised our findings in terms of an IMOI framework (input-mediator-outcome-input) postulated by Ilgen, Hollenbeck, Johnson, & Jundt (2005), which is commonly used in the literature on team effectiveness (see Figure 1). The model looks at the processes of a team as an integrated system with the organisation, the team itself and the individual members serving as input. The processes that happen during the teamwork serve as mediators, and multiple criteria count as outcomes of the team work (i.e. team satisfaction or taskwork outcomes). In addition, there are feedback loops throughout the process (new input) so that outcomes, for instance, can have an effect on the mediators and input. For instance, team satisfaction as an output can, in a subsequent feedback loop, have influence on inputs or mediators (e.g. team cohesion). Despite being used extensively in the human-human teaming literature (Ilgen, Hollenbeck, Johnson, & Jundt, 2005), this model has also been used for a similar but somewhat less exhaustive analysis by You and Robert (2017) as well as by Stowers et al. (2017) in the HAT literature. This paper adds to this discussion by a stronger focus on the social-psychological mechanisms in HATs which have been largely ignored in previous work. In addition, knowledge gaps related to these mechanisms are identified for the purpose of future research.

This paper is structured along the lines of the IMOI framework. First, we describe the outcomes relevant to HAT. Second, we describe the mediators and how they affect the outcomes, especially as compared to human-human teaming, and lastly, we describe the inputs. The model is described in reverse order to clearly show the feedback loops and the gaps in the literature.

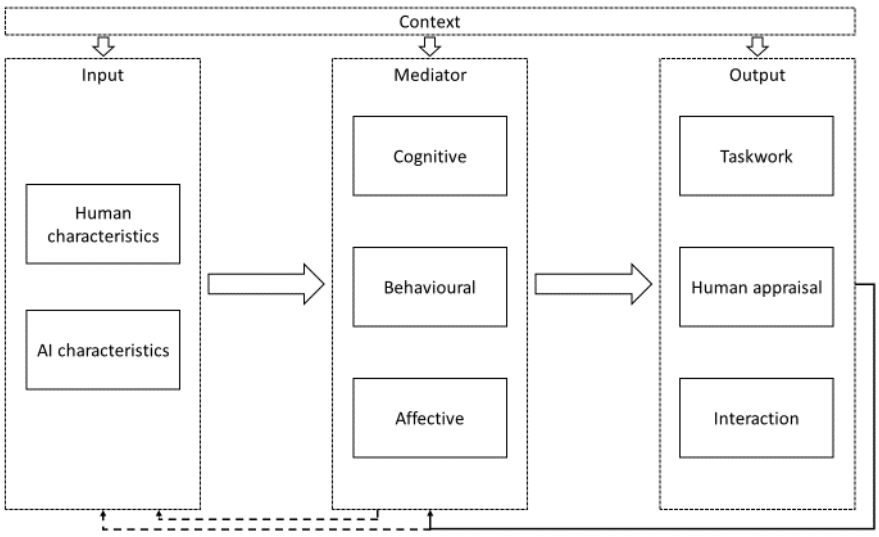


Figure 1: Input-Mediator-Output (IMO) team Effectiveness Framework

2.1 Outcomes

The model distinguishes three classes of outcomes: taskwork outcomes (relatively objective variables indicating how well the task is completed, e.g., errors made), human appraisal, (e.g., a person's satisfaction), and interaction outcomes (that inherently emanate from working in a team, e.g., calibrated trust).

2.1.1 Taskwork outcomes

A general finding from our literature review is that most studies have focused on task performance, or the completion of goals and sub goals. Task performance is often measured in terms of time needed to perform a task, error rate, quality, or efficiency (e.g., number of mouse clicks needed to reach a certain level of performance (Clare, Cummings, & Repenning, 2015)). One study into adaptive aiding in military HATs, for example, measured mission completion time, detection performance and classification performance (Visser & Parasuraman, 2011). Another study using a multi-UxV planning system 'IMPACT' measured the rate at which operators correctly accepted or rejected plans proposed by an intelligent assistant (Mercado et al., 2016). Other measures used include target detection performance in a gunnery task with an intelligent aided target recognition system (Chen & Terrence, 2009), and area coverage in a surveillance task collaborating with a swarm of UAVs (Clare et al., 2015).

Another observation, related to task work outcomes, is that studies on AI in a military context are primarily anchored in the current operating paradigm of fighting and defeating enemies (Spiegeleire, Maas, & Sweijs, 2017). As such, AI is seen as an extension of technological development to increase the effectiveness of operations, with lower risks and higher speed. Think, for example, of algorithms to interpret imagery from drone surveillance feeds, remotely-piloted systems, and AI-coordinated swarms. This focus may explain the relative abundance of studies on optimising task performance.

2.1.2 Human appraisal

Human appraisal is deeply engrained into the human-human teaming literature (Ilgen et al., 2005). Examples of measures of human appraisal include various forms of satisfaction (e.g. with the team, with the outcome or with the collaboration), attribution and agency, coping/mental resilience, acceptance and self-efficacy. In general such research is relatively sparse in the context of HAT, even though some of these measures have been studied.

Some studies have, for example, focused on topics like technology acceptance, the attribution of blame and self-efficacy. It can, for instance, be measured how (much) blame is attributed to humans and AI for an action with a negative (unintended) effect (Malle, Scheutz, Arnold, Voiklis, & Cusimano, 2015a). Closely related are studies on the attribution of moral permissibility which can be measured by the degree to which people find the actions of the AI or the human justifiable and/or permissible (Malle et al., 2015a). Self-efficacy has been studied in relation to trust, showing that a person's confidence in his or her own skills predicts the extent to which he relies on the automation (e.g., Prinzel, 2002).

2.1.3 Interaction outcomes

A number of outcomes are related to the interaction within the team. For instance, the effectiveness of communication and, related to this, situation awareness (SA) on the individual, team, or even organisational level (Stanton, Salmon, Walker, Salas, & Hancock, 2017). Another outcome of the team interaction processes is the shared mental model of, for instance, the task, goals and member capabilities the team has developed during task execution. An optimally shared mental model results, according to these studies, in a shared and correct view of how to reach objectives (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). A correct (shared) mental model also results in appropriately calibrated trust in the AI, which is another interaction outcome that has received extensive attention (Schaefer, Chen, Szalma, & Hancock,

2016). Overall, interaction outcomes seem to have received a fair share of attention in HAT research.

2.1.4 Gaps with regard to outcomes

With regard to outcomes that have been addressed in current research, we concluded that least attention has been paid to human appraisal, that is, subjective evaluations of team performance and functioning. There has been a focus on overall task performance in current HAT studies, as the main aim seems to be to improve overall team outcome by optimally combining the qualities of human and non-human agents (e.g., Onnasch, Wickens, Li, & Manzey, 2014). In such a context important questions are, for example, how to divide tasks across actors by taking workload into account (adaptive automation), or which knowledge AI needs in order to optimally support a human agent. A broader view, however, would be to consider factors like motivation of human team members and human mental resilience. This broader view would then assume the AI to be more than a decision-support system, but a real team member.

Going back to our initial scenario for example, it would be necessary to know how lieutenant Bob evaluates the arguments of Smart Bud, whether he trusts and follows them, and how the team would continue their interaction on both task and social level after experiencing negative consequences due to wrong decisions (resilience).

2.2 Mediators

In order to describe the mediators, we made a distinction between cognitive and behavioural factors on the one hand, and affective factors on the other.

2.2.1 Cognitive and behavioural

Many cognitive processes and behaviours have been investigated in terms of their effect on human-autonomy teaming. These range from the mental models that individuals develop of systems, to cognitive workload, to the extent to which operators tend to monitor system performance.

Mental models are frameworks that individuals construct in order to generate predictions and support their understanding of a system's or a person's drivers, abilities and behaviour. Mental models are constructed and continuously updated by translating observations to internal representations. They are cognitive short-hands that provide guidance for interactions, without necessarily being accurate reflections of the outside-world (Cooper, Reimann, & Cronin, 2011). In HATs, the mental model that a human operator has of an AI determines the capabilities and reliability he expects from it. This way, mental models can affect many other mediators and almost all possible outcomes of a HAT. For instance, according to a recent meta-analysis, mental models are an important antecedent of trust (Schaefer et al., 2016).

Specifically, the mental model that the human has of the AI (and vice versa), regarding the expected level of knowledge, is of great influence on the style and effectiveness of communication. Previous research has identified that in order for communication to be effective there has to be a certain level of mutual understanding, or common ground, between sender and receiver (Lee et al., 2005). It has been found that people formulate their messages according to their estimate of the knowledge shared between them and the receiver. This is done intuitively between humans, albeit sometimes incorrect and/or based on preconceptions. For effective communication to occur between humans and AI, the AI has to support this as well. This means that the human must be able to gauge the level of knowledge the AI has, but the AI must also be aware of the knowledge the human has, as to avoid talking past each other.

2.2.2 Affective

Team members' motivation expands beyond attaining overall task goals, as the main drivers of military

motivation are affective in nature and concern the relation with other team members (Wong, Kolditz, Millen & Potter, 2003). Team identification is important for a team to function efficiently, meaning that all members of the team see themselves as part of the team. Team identification can be created through interdependence, where the accomplishment of one individual relies on that of a teammate. Creating team interdependence can have an effect on performance and on human appraisal outcomes. For instance, Walliser, Mead, and Shaw (2017) found that human agent teams that were structured as a team through interdependence had higher performance, and higher affect then when they were not structured as a team, that is, team identity was not created. In the field of autonomous synthetic characters (i.e. autonomous players in for instance Role Playing Games) some research has also been done into team identification. Prada and Paiva (2009) showed that when social skills where programmed such that the autonomous characters exhibited behaviours consistent with the group's composition, context and structure, team identification was higher.

Trust is one of the affective mediating mechanism that has received most attention in research on working with automated or autonomous systems. The importance of this mechanism originates from the finding that trust is a strong predictor of how systems are used. For instance, too much trust results in misuse through overestimation of system capabilities, potentially leading to dangerous situations (Baker & Keebler, 2017). Undertrust, on the other hand, results in disuse of systems, potentially leading to increased workload for the human, and reduced efficiency. An extensive literature has investigated the antecedents of trust, summarized in a meta-analysis by Schaefer, Chen, Szalma and Hancock (2016). Predictors of trust include a person's age, his or her understanding of the automation, the appearance and anthropomorphism of the system, and many others. Models are under development to gain insight into how trust changes over time (e.g., Gao, Clare, Macbeth, & Cummings, 2013). Furthermore, researchers have experimented with priming participants who tend to overtrust systems (in this case, gamers), which resulted in successful calibration of trust and improved performance of the human-machine team (Clare, Cummings, & Repenning, 2015). Yet other researcher propose general models for when and how to repair trust in human machine interaction (de Visser, Pak, & Shaw, 2018).

Another important affective mediator is our perception of how our team members, in this case the AI, sees us. Even though only few studies investigated this factor, it may have profound effect on team functioning. For instance it was found that rejection by the robot after playing a game ("That was boring! I don't want to play with you again") significantly lowered self-esteem of the participant. Robot appraisal ("That was fun, I would like to play with you again some time.") on the other hand, had no influence on self-esteem (Nash et al., 2018).

2.2.4 Gaps related to mediators

A number of topics have received relatively little attention. First, when considering the functioning of HATs in a military context, we identified a gap regarding research on the human's mental model of the AI's ethical awareness, which would contribute to a correct perception of meaningful human control (Ekelhof, 2015). Meaningful human control entails the notion that humans rather than computers should remain in control of, and be morally responsible for, relevant decisions about (lethal) military operations. Second, the effectiveness of HATs is not only induced by analytic processes such as optimizing (shared) mental models but also by less rational processes such as when robot rejection lowers self-esteem. With the exception of trust, we see that such affective and irrational mediators have often been overlooked in the HAT literature. Research focusses on how we as humans view AI, but it does not yet envision AI as a team member triggering affective processes within a team. As team identification is an important prerequisite for accurate team functioning more research is needed on how to design socially skilled AI without jeopardizing task performance.

In our scenario above in which lieutenant Bob is reasoning with his Smart Bud about the best course of action, it would, for example, be interesting to know how Bob would actually decide as a function of the type of dialogue: what if Smart Bud uses some kind of persuasion, or affective non-verbal signs like a frown, or

natural speech rather than text? All these factors will have large effects on the information processing of Bob and consequently on the accuracy of the decisions being made.

2.3 Input

2.3.1 Human characteristics

Literature on working with automated and autonomous systems provides many insights into individual human differences that are known to affect the way in which HATs function. Without being exhaustive, we describe the most relevant variables.

Attentional control is one's ability to focus and shift attention in a flexible manner. The literature on the effect of individual differences in attentional control on the functioning of HATs is summarised by Chen & Barnes (2014b). Attentional control was identified as one of the most important abilities that affect performance of UAV operators, and in a broader sense this ability affects performance on tasks ranging from driving to flight training. Regarding collaboration with automated systems, it has been suggested that operators with lower attentional control rely more heavily on the automation.

Spatial ability is another relevant factor for human control of automation. Amongst others, spatial ability is essential for navigation, UAV control, visual scanning and target detection (Chen & Barnes, 2014).

Gaming experience has been investigated as a predictor of human trust in HAT systems. This makes sense because many computer games resemble realistic HAT scenarios. In strategy games, for example, gamers often control groups of elements (like tanks) that are implemented via AI. In general, gaming experience has been found to increase trust in autonomous systems. In some cases, this leads to over-trust, which can be alleviated using simple interventions like presenting negative quotes about system performance from previous operators (Clare et al., 2015).

Besides self-efficacy being an outcome from a team interaction, it is also a trait that humans bring into the interaction. In working with automation, research shows that self-efficacy can have an influence on complacency and thereby on performance. Specifically, having low self-efficacy can induce strategies that increase the likelihood of automation-induced complacency. In other words, people who think poorly of themselves are more likely to follow the system. People with high self-efficacy, on the other hand, are more likely to offload too little work to the system, resulting in high workload (Prinzel, 2002).

Prior experience with robots leads to more positive attitudes towards robots and increased trust (Takayama, Takayama, & Pantofaru, 2016). Prior experience is mentioned here as a human characteristic but might also be fostered through interaction and can therefore also be seen as a mediator.

2.3.2 Robot characteristics

According to a meta-analysis by Hancock et al. (2011), characteristics of autonomous systems that contribute to trust in the system fall into either of two categories. The first category consists of performance-based factors, related to the reliability and effectiveness of the robot. The second category is related to robot attributes such as 'personality', and visual appearance. Both categories contribute to the development and maintenance of trust in the robotic system. Performance-based factors include failure rate (Hancock et al., 2011; Merritt & Ilgen, 2008), reliability, and false alarm rate (Hancock et al., 2011).

Anthropomorphism, the degree to which an agent exhibits human characteristics, has a major impact on how the agent is perceived and, consequently, the schema's that are triggered in interaction. Research indicates, for example, that emotions are easily attributed even to robots that show simple dog behaviour (Gacsi et al., 2016). It has been found that anthropomorphic agents are trusted more than machines (Visser et al. 2016).

Appearance also has an effect on feeling responsible for the task. One study showed, for example, that participants relied more on a human-like than a machine-like collaborator (Hinds et al., 2004). In a study by Biswas and Murray (2017) it was found that the preference for humanlike machines even extended to biased performance as humans preferred to interact with a robot containing biased algorithms as compared to unbiased robots.

A central problem of many advanced autonomous systems is that they (partially) rely on algorithms that remain opaque to the user, which has stimulated research into the topic of explainable AI (XAI, Lent, Rey, Fisher & Mancuso, 2004). The goal of this area of research is to develop intuitive measures and explanations that systems can use to calibrate operator's mental models and trust (e.g., Waa, Diggelen, & Neerincx, n.d.). Furthermore, we know from social psychology that a lack of justification from a teammate can negatively affect interaction. In all, even though the technical challenges here are immense, progress on XAI is likely essential to further develop HAT effectiveness.

The design of AI can greatly influence the way information is communicated and thus the way it interacts with its teammates. Some research has been done in the field of aviation involving the effects of speech vs text as a modality of communications. It was found that using speech instead of textual communication had an advantage, e.g. noticing more deviations in the monitoring task (Stedmon et al., 2007). The same research describes another experiment where the effects of real versus synthesised speech were compared. It was found that by using real speech there was an increased level of trust in the system and reduced reaction times (Stedmon et al., 2007). In general, the use of speech is of great influence on the HAT and trust development, particularly as speech invokes characteristics resembling a personality.

2.3.3 Gaps related to input

The number of human and robot characteristics is quite substantial and they all have their own effects on the outcomes and mediators previously mentioned. With regard to human characteristics it is particularly noteworthy that effects may change over time, for example due to positive or negative experiences in interacting with AI. To date, however, experiments mostly addressed short term effects and longitudinal data are scarce. With regard to robot design more attention should be paid to input variables affecting human-robot dialogue. In the context of explainable AI more insight is needed on how robot design and type of reasoning affect human trust and motivation. Human-like AI triggers mental processes that may be positive for social interaction but less advantage for task performance. In designing AI both factors, and their mutual effects, need to be taken into account.

All the input variables mentioned above could also matter in the interaction between Lieutenant Bob and Smart Bud: does Bob for example feel in control in debating with AI and does he understand the arguments provided? How does Bob assess the level of accuracy of Smart Bud, does he have any positive or negative experiences? And would he decide differently if Smart Bud was a robot rather than an algorithm informing him through his personal device?

3.0 DISCUSSION

3.1 Human appraisal

Most studies on HATs focused on the question of how to optimise task processes as to maximise task-related outcomes. As noted above, one reason for this emphasis might be that AI is often considered as technological advancement further increasing effectiveness and speed of current operations. However, even in kinetic operations, team performance is not solely driven by cognitive mechanisms such as having adequate SA or relevant task knowledge. Going back to the scenario we started with, lieutenant Bob might well feel compromised by Smart Buds superior situation awareness or restrained in his autonomy by

constantly being monitored by Smart Bud. To date, knowledge is lacking concerning the influence of such motivational factors, i.e. human appraisal, in cooperating with AI, and how these factors interact with task-related factors. We expect human appraisal to become even more important in the future however, as AI becomes more and more engrained in team functioning. A view incorporating a wider view on motivational factors would significantly elaborate research questions with regard to HAT. For example, if Smart Bud is able to detect my level of workload or fatigue, how would that affect my level of experienced autonomy, and would I need emotional support from Smart Bud when in a dangerous situation? Furthermore, we know from research on perceived autonomy that it predicts persistence and adherence, and improves effective performance especially in complex tasks (Deci & Ryan, 2008).

Research on human appraisal in HATs should take results from the social-robotics field into account, as this field particularly focused on such factors. Note, however, that these studies did not consider human-automation interaction in a team or in a military context, but instead investigated AI as human support or assistant (for instance being a buddy to elderly, Looije, Neerincx & Cnossen,, 2010; Wada, Shibata, Musha, & Kimura, 2008). A main gap is therefore research on human appraisal in team interaction, which dominates the human team literature, but is rather absent in the HAT literature.

Another important outcome, or property, of sociotechnical systems is resilience. Resilience, or "the capability of recovering safely and efficiently from abnormal events" (Patriarca, Bergström, Di Gravio, & Costantino, 2018) has recently been reviewed in the context of HATs (Matthews, Barber, Teo, Wohleber, & Lin, 2016), and is related to many of the mediators and inputs we discuss here. For the military context however, we believe that 'mental resilience', or the individual ability to cope with events that can have an impact on mental well-being, should also be investigated in the light of HATs. This is because we know that team properties can affect mental resilience of individual members (e.g., Kamphuis, van Hemert, van Wouwe, van den Berg, & van Boxmeer, 2012), and replacing humans by systems strongly affects those properties.

3.2 Human information processing

As far as mediators are concerned many studies deal with adequate (shared) mental models as a prerequisite for optimal team performance. Team members have expectations with regard to each other's knowledge level, reliability, needs and behaviour with regard to joint task performance. Human aware AI, for example, seeks to design accurate user models allowing the AI to best anticipate on human needs and support, as to optimally complement human task performance.

However, as also noted by Patterson (2017), military thinking as well as research on human-autonomy teaming usually assumes, implicitly, a rational and deliberate decision-making process (Williams, 2010). For instance, military theory often uses the OODA loop (observe, orient, decide, act) to model decision-making. In reality however, many of these steps are taken unconsciously for most of our decisions. A useful model that describes human decision-making is Kahneman's dual process theory. It proposes that humans make decisions using one of two systems. System 1 is fast, intuitive, automatic and heuristic-based, while system 2 is slow, effortful and analytic (Kahneman, 2011). Williams (2010) discusses the relevance of decision-making theory for military operations and presents several examples of biases occurring in practice. Even though such biases will also occur in HATs, empirical research on both presence and mitigation is still sparse (see also Bosch and Bronkhorst, 2018).

Some research suggests that humans prefer biased robots over unbiased ones as they are more humanlike (Biswas 2017). Mirnig et al. (2017) showed, for example, that robots who gave erroneous instructions during a LEGO task, were perceived as more likeable, but equally intelligent, than robot who gave correct instructions. Salem et al. (2013) showed the same effect in a moving task where the robot gave instructions on where to put the items. The robot in the incongruent speech-gesture condition (e.g. say it goes left and point right) was perceived more likeable than the robot in the congruent speech-gesture condition. Moreover,

they found that participants were more positive about future interactions in the incongruent condition. Participants did perform significantly worse, however, when they got incongruent instructions.

Presumably, intuitive processes will become more important when level of agency of AI increases. Not only because of increased intentionality ascribed to AI, but also because of all kinds of social processes. These unconscious and automatic processes have substantial influences on human thinking, task performance and social interactions. Social robotics use these natural tendencies through anthropomorphised AI. By making AI more human-like or animal-like, specific cognitive schema's are automatically triggered, supporting goals to be attained. These goals are obviously different than goals in military operations. However, also in a military context unconscious processes play significant roles, which should be taken into consideration when designing AI.

3.3 Conclusion

AI will become more and more important in future military operations. We concluded from our literature review that current research in the military domain mostly focuses on how to optimize task outcomes, and that AI is still mainly viewed as an 'add-on' to human decision makers (Spiegeleire, et al., 2017). However, as also discussed by Bosch & Bronkhorst (2018) there is a trend towards collaborative decision making, where AI is regarded as a true team-player. Such a situation would require profound mutual awareness, in order to jointly adapt to complex and dynamic situations. In addition to this, we argued that increased agency of AI will also trigger social behaviours, as these are deeply anchored into our (largely) social brain. So in addition to cognitive challenges, more insight is needed into how humans react socially to artificial teammates to further optimize HATs. As these aspects are well recognised in other domains such as health care (see for example Looije, Neerincx, & Cnossen, 2010), a main challenge would be to combine these different research strands, and to investigate how trade-offs are made with regard to task and relation in mixed human-AI teams.

REFERENCES

[1] Baker, A. L., & Keebler, J. R. (2017). Advances in Human Factors in Robots and Unmanned Systems, *499*, 331–340. https://doi.org/10.1007/978-3-319-41959-6

[2] Bosch, K., van den & Bronkhorst, A. (2018). Human-AI Cooperation to Benefit Military Decision Making. *Human-AI Cooperation to Benefit Military Decision Making, STO-MP-IST*, 1–12.

[3] Boyd, J. (1995). The Essence of Winning and Losing, by John R. Boyd, (August).

[4] Chen, J. Y. C., & Barnes, M. J. (2014). Human–Agent Teaming for Multirobot Control: A Review of Human Factors Issues. *IEEE Transactions on Human-Machine Systems*, *44*(1), 13–29. https://doi.org/10.1109/THMS.2013.2293535

[5] Clare, A. S., Cummings, M. L., & Repenning, N. P. (2015). Influencing trust for human-automation collaborative scheduling of multiple unmanned vehicles. *Human Factors*, *57*(7), 1208–1218. https://doi.org/10.1177/0018720815587803

[6] Deci, E. L., & Ryan, R. M. (2008). Facilitating optimal motivation and psychological well-being across life's domains. *Canadian Psychology/Psychologie Canadienne*, *49*(1), 14–23. https://doi.org/10.1037/0708-5591.49.1.14

[7] Ekelhof, M. A. C. (2015). Autonome Wapens. *Militaire Spectator*, *184*(5), 232–245.

[8] Gao, F., Clare, A. S., Macbeth, J. C., & Cummings, M. L. (2013). Modeling the Impact of Operator

Trust on Performance in Multiple Robot Control,. *AAAI Spring Symposium: Trust and Autonomous Systems*, 16–22.

[9] Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2017). When Will AI Exceed Human Performance? Evidence from AI Experts, 1–21.

[10] Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in Organizations: From Input-Process-Output Models to IMOI Models. *Annual Review of Psychology*, *56*(1), 517–543. https://doi.org/10.1146/annurev.psych.56.091103.070250

[11] Kamphuis, W., van Hemert, D. A., van Wouwe, N., van den Berg, C. E., & van Boxmeer, L. E. L. M. (2012). Een model van mentale veerkracht Hoe kan Defensie herstel na uitzendingen bevorderen ? *Militaire Spectator*, *181*(11), 495–507.

[12] Lent, M. Van, Rey, M., Fisher, W., & Mancuso, M. (2004). An Explainable Artificial Intelligence System for Small-unit Tactical Behavior. *In Proceedings of the Sixteenth Conference on Innovative Applications of Artificial Intelligence*, 900–907.

[13] Looije, R., Neerincx, M. A., & Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human Computer Studies*, *68*(6), 386–397. https://doi.org/10.1016/j.ijhcs.2009.08.007

[14] Mathieu, J. E., Heffner, T. S., Goodwin, G., Salas, E., & Cannon-Bowers, J. (2000). The Influence of Shared Mental Models on Team Process and Performance. *Journal of Applied Psychology*, *85*(2), 273–283.

[15] Nyholm, S. (2017). Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci. Science and engineering ethics, 1-19.

[16] Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, *56*(3), 476–488. https://doi.org/10.1177/0018720813501549

[17] Prinzel, L. J. I. (2002). The Relationship of Self-Efficacy and Complacency in Pilot-Automation Interaction. *Technical Memorandum No. TM-2002-211925, NASA.*

[18] Salas, E., Sims, D. E., & Shawn Burke, C. (2005). Is there A "big five" in teamwork? *Small Group Research*, *36*(5), 555–599. https://doi.org/10.1177/1046496405277134

[19] Sanders, T. L., Macarthur, K., Volante, W., Hancock, G., Macgillivray, T., Shugars, W., & Hancock, P. A. (2017). Trust and Prior Experience in Human-Robot Interaction, (2011), 1809–1813. https://doi.org/10.1177/1541931213601934

[20] Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A Meta-Analysis of Factors Influencing the Development of Trust in Automation. *Human Factors*, *58*(3), 377–400. https://doi.org/10.1177/0018720816634228

[21] Spiegeleire, S., de Maas, M., & Sweijs, T. (2017). *Artificial Intelligence and the Future of Defense – Strategic Implications for Small- and Medium-Sized Force Providers. The Hague Centre for Strategic Studies (HCSS).* Retrieved from http://www.hcss.nl/report/artificial-intelligence-and-future-defense

[22] Stanton, N. A., Salmon, P. M., Walker, G. H., Salas, E., & Hancock, P. A. (2017). State-of-science:

situation awareness in individuals, teams and systems. *Ergonomics*, *60*(4), 449–466. https://doi.org/10.1080/00140139.2017.1278796

[23] Stowers, K., Oglesby, J., Sonesh, S., Leyva, K., Iwig, C., & Salas, E. (2017). A framework to guide the assessment of human-machine systems. *Human Factors*, *59*(2), 172–188. https://doi.org/10.1177/0018720817695077

[24] Takayama, L., Takayama, L., & Pantofaru, C. (2016). Influences on Proxemic Behaviors in Human-Robot Interaction Influences on Proxemic Behaviors in Human-Robot Interaction, (NOVEMBER 2009), 5495–5502. https://doi.org/10.1109/IROS.2009.5354145

[25] Visser, E. J., de Pak, R., & Shaw, T. H. (2018). From "automation" to "autonomy": The importance of trust repair in human-machine interaction. *Ergonomics*, *139*, 1–33. https://doi.org/10.1080/00140139.2018.1457725

[26] Waa, J., van der Diggelen, J., & Neerincx, M. (n.d.). The design and validation of an intuitive confindence measure. https://doi.org/10.1145/1235

[27] Wada, K., Shibata, T., Musha, T., & Kimura, S. (2008). Robot therapy for elders affected by dementia. *IEEE Engineering in Medicine and Biology Magazine*, *27*(4), 53–60. https://doi.org/10.1109/MEMB.2008.919496

[28] Wong, L., Kolditz, T. A., Millen, R. A., & Potter, T. M. (2003). Why they fight: combat motivation in the Iraq war. ARMY WAR COLL STRATEGIC STUDIES INST CARLISLE BARRACKS PA.

[29] You, S., & Robert, L. (2017). Teaming up with Robots : An IMOI (Inputs-Mediators-Outputs- Inputs) Framework of Human-Robot Teamwork, (October).

Agent Transparency and Human-Autonomy Teaming Effectiveness in Multi-Robot Management Contexts

Jessie Y. C. Chen, Kimberly Stowers, Ryan Wohleber, Michael Barnes
US Army Research Laboratory, Aberdeen Proving Ground, MD
UNITED STATES OF AMERICA

yun-sheng.c.chen.civ@mail.mil; kim.stwrs@gmail.com; ryan.wohleber@gmail.com;
michael.j.barnes.civ@mail.mil

ABSTRACT

As intelligent agents (IAs) become more sophisticated, it is crucial to develop user interfaces that make the IA's intent, reasoning and predicted future states transparent to the human operator. Chen et al. leveraged Endsley's model of situation awareness (SA) and developed the Situation awareness-based Agent Transparency (SAT) model to organize the issues involved with supporting the human's awareness of an agent's: current actions and plans (L1), reasoning process (L2), and outcome predictions (L3). Projection uncertainty (U) may also be communicated. The current paper presents two human factors studies that investigated the effects of agent transparency on the effectiveness of human-agent teaming in military multi-robot management contexts. The first study investigated effects of levels of agent transparency; the second study focused on effects of agent communication framing (agent's style of communication). The user interfaces employed in the experiments were based on the SAT model. Overall, the results showed that perceptions of the IA (its aptitude and usefulness) increased with agent transparency. Trust survey results showed that the complimentary IA was rated as less trustworthy compared with the critical IA, especially in low transparency conditions. Participants tended to agree more with the critical IA when it was more opaque than when it was more transparent; on the other hand, participants' agreements with the complimentary IA did not differ regardless of the IA's transparency.

1.0 INTRODUCTION

As autonomous systems become more intelligent and more capable of complex decision making [1][2], it becomes increasingly difficult for humans to understand the reasoning process behind the system's output. However, such understanding is crucial for both parties to collaborate effectively, particularly in dynamic environments [3][4]. Indeed, *Explainable AI* [5], a DARPA program announced in August 2016, underscores the importance of making intelligent systems more understandable to the human. The U.S. Defense Science Board recently published a report, *Summer Study on Autonomy* [6], in which the Board identified six barriers to human trust in autonomous systems, with 'low observability, predictability, directability, and auditability' as well as 'low mutual understanding of common goals' being among the key issues (p. 15).

1.1 Situation awareness-based Agent Transparency (SAT) Framework

Chen et al. [3][7] leveraged Endsley's model of situation awareness (SA) [8][9] and developed the Situation awareness-based Agent Transparency (SAT) model, which mirrors the three levels (perception, comprehension, and projection) of Endsley's model (Figure 1). At the first SAT level (L1), the operator is provided with the basic information about the agent's current state and goals, intentions, and proposed actions (*perception* of the environment from the agent's perspective). At the second level (L2), the operator is provided with information about the agent's reasoning process behind those actions and the constraints/affordances that the agent considers when planning those actions (*comprehension*). At the third level (L3), the operator is provided with information regarding the agent's projection of future states, predicted consequences, likelihood of success/failure, and any uncertainty associated with the

aforementioned projections (*projection*). Understanding the effects of uncertainty on the third level of SAT was considered of particular importance, because projection is predicated on many factors whose outcomes are never known precisely [7][10]-[13]. Prior research has shown that agents conveying uncertainty can improve the joint human-agent team performance and are also perceived as more trustworthy [10][14][15].

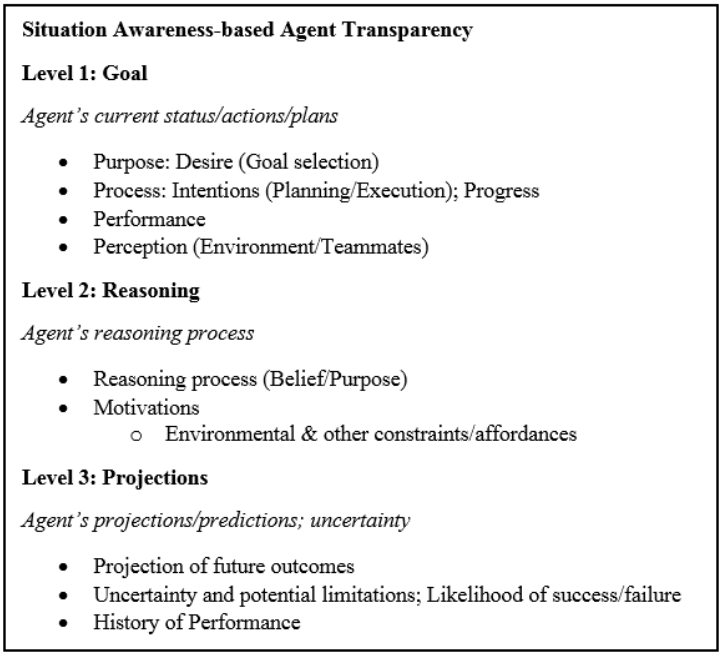


Figure 1: Situation awareness-based Agent Transparency (SAT) Framework

1.2 Impact

Successful operational employment of autonomous systems requires command and control agility. Indeed, agility in tactical decision-making, mission management, and control is one of the key attributes for enabling heterogeneous unmanned vehicle (UxV) teams to successfully manage the 'fog of war' with its inherent complex, ambiguous, and time-challenged conditions. Mission effectiveness relies on rapid identification and management of uncertainties that can disrupt an autonomous team's ability to complete complex operations. Increasingly, research and development efforts are focusing on developing intelligent agents (IAs) that can work with human operators on managing the UxV teams [3][16]. One of those efforts is the Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT) project, which is currently funded by the ARPI. IMPACT is investigating issues associated with flexible play-calling [17][18], global cooperative control and local adaptive/reactive capability, and bi-directional human-autonomy interaction in military mission contexts [19]. However, in order for the IA to collaborate with human operators, the IA must be able to communicate its current and future states and receive commands to alter them when necessary [3]. Furthermore, knowledge of the IA's state of uncertainty has proven to be a crucial factor in the ability of humans to interact with automated systems [4][7][10]-[13][20]. In particular, it is important for operators to understand not only the locus of the uncertainty, but also to understand its source [21].

1.3 Current Studies

Two studies conducted under the IMPACT project are summarized in this paper. The first study investigated effects of levels of agent transparency; the second study focused on effects of agent communication framing (agent's style of communication). In both studies, participants completed blocks of mission vignettes, in

which they were instructed to choose one of two plans provided by an IA, which always recommended an optimal plan (A) and a back-up plan (B). Three out of eight times, plan B was in fact better than plan A, based on updated intelligence or commander's instructions provided to the participants. The participant's task was to select the best plan based on current mission requirements. The user interfaces employed in the experiments were based on the SAT model.

2.0 STUDY 1

2.1 Methodology

This experiment utilized a within-subjects design with agent transparency as the independent variable. A previous study using the same UI from IMPACT tested the utility of Level 1 transparency and found that it was most useful when combined with higher levels of transparency [10]. Thus, transparency was only tested at three higher levels: (a) Level 1+2: containing reasoning information, (b) Level 1+2+3: containing reasoning and projection information *without* uncertainty, and (c) Level 1+2+3+U containing reasoning and projection *with* uncertainty. The UI varied per condition by showing corresponding pieces of SAT-level information on the map, in text, and on a sliding bar scale (Figure 2).

Figure 2: IMPACT Study 1

Fifty-three students from an American university were recruited for this experiment. The average age of participants was 21.7 years (*sd* = 3.6). Prior to beginning the experiment, participants completed several individual difference surveys, including an Implicit Association Test (IAT) to track any bias they may have against technology. They then received about 1 hour of training. The subsequent experiment was divided into 3 blocks of 8 missions, with conditions counter-balanced. Using the interface presented in Figure 2, participants were presented with two plans to complete each mission, and were instructed to choose just one of them to move forward. For each set of plans, the agent always recommended one plan over the other, but the human always made the final decision. Reliability of the agent was held constant such that it was right in 5 out of every 8 scenarios. Information regarding mission objectives and commander's instructions were given before each set of decision options were presented, while information regarding specific mission parameters and vehicle capabilities were given on the decision screen (Figure 2). Each experimental block (including corresponding surveys) took participants about 30 minutes to complete.

After each block, participants completed surveys, including a measure of workload on an interval scale from 1 (low) to 10 (high) (NASA-TLX [22]), and a trust survey using a 7-point Likert scale (modified Trust in Automated Systems survey with two subscales [23][24]). Performance was measured throughout by tracking total correct responses, "proper use" (instances when the human agreed with the IA and answered correctly) and "correct rejections" (instances when the human disagreed with the IA and answered correctly). This

performance measure was adapted from signal detection theory [25], but signal detection analyses were not completed.

2.2 Results

2.2.1 Operator Performance - Total Correct Response

A repeated-measures ANOVA showed significance, indicating an increase in operator performance (Total Correct Response, which combines Proper Use and Correct Rejections; Figure 3) with transparency level, $F(2, 104) = 10.31$, $p < .001$, $\eta_p^2 = .17$. Pairwise comparisons showed a significant difference between Level 1+2 and Level 1+2+3 ($p < .05$), as well as between Level 1+2 and Level 1+2+3+U ($p < .001$). Total correct responses were highest in the Level 1+2+3+U SAT condition and lowest in the Level 1+2 SAT condition.

2.2.2 Operator Performance - Proper Use and Correct Rejections

To examine exactly what areas of performance were most impacted, a repeated-measures MANOVA of proper use and correct rejections was performed. A significant multivariate effect indicated an increase in performance with transparency across both sub-measures: $F(4, 206) = 6.05$, $p < .001$, $\eta_p^2 = .11$. A univariate effect was found for proper use, $F(2, 104) = 10.66$, $p < .001$, $\eta_p^2 = .17$. Pairwise comparisons with Bonferroni correction indicated a significant difference between Level 1+2 and Level 1+2+3 ($p < .01$), as well as between Level 1+2 and Level 1+2+3+U ($p < .001$). As can be seen in Figure 3, both proper use and correction rejections were highest in the Level 1+2+3+U condition.

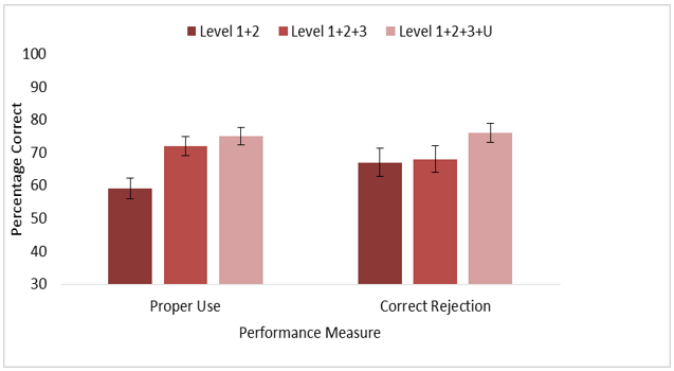


Figure 3. Percentage correct for both IA proper use and IA correct rejection rates.

2.2.3 Operator's Perceived Workload

In order to examine operator workload, we conducted a repeated-measures ANOVA on the total NASA-TLX scores, as well as a repeated measures MANOVA on the 6 unweighted subscales of the NASA-TLX. The effect of total workload was not significant, $F(2, 104) = .86$, $p = .43$, $\eta_p^2 = .02$. The effect of the subscales was also not significant, $F(12, 198) = .87$, $p = .58$, $\eta_p^2 = .05$.

2.2.4 Operator's Perceived Trust in the IA

Individual differences in IAT scores were covaried with participant trust scores in order to account for pre-existing biases that may affect trust. Trust was thus examined using a repeated-measures MANCOVA across the two subscales of our modified trust survey. A significant multivariate effect was found, $F(4, 202) = 2.61$, $p < .05$, $\eta_p^2 = .05$. Additionally, significant main effects of both subscales were found: (a) for integrating and diplaying information – $F(2, 102) = 3.48$, $p < .05$, $\eta_p^2 = .06$, and (b) for suggested decisions – $F(2, 102) =$

4.08, $p < .05$, $\eta_p^2 = .07$. Pairwise comparisons with Bonferroni correction indicated a significant difference between Level 1+2 and Level 1+2+3 in *suggesting decisions* ($p < .05$; Figure 4).

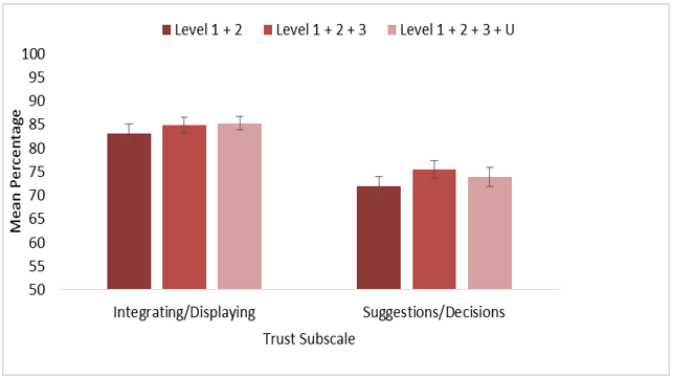


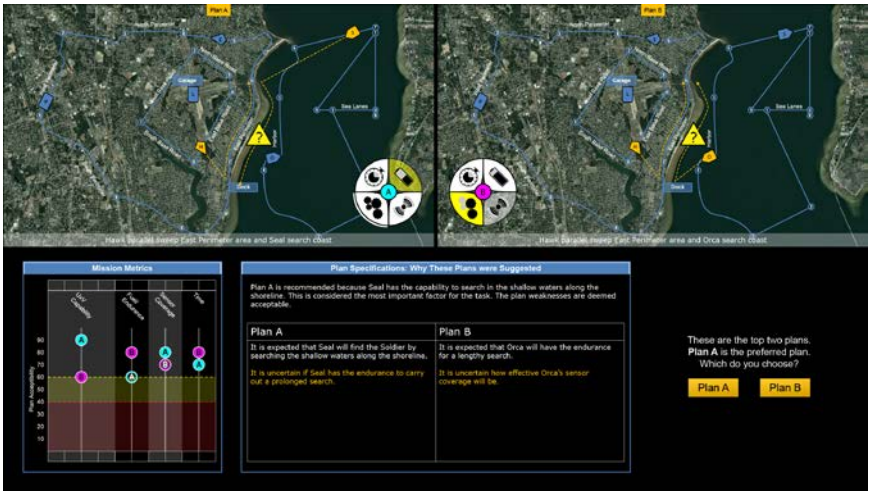
Figure 4. Trust in integrating and displaying information, as well as suggesting decisions.

3.0 STUDY 2

The second study focuses on the human factors aspects of the agent's transparency and communication framing in the context of human-autonomy teaming via IMPACT technologies and capabilities. The overall goal of the this study was to understand the interaction between level of agent transparency communication, according to the SAT model [3], and the agent's framing of communication. In line with Dzindolet et al.'s findings [21], we expect trust and evaluation of the agent to be higher with a high transparency interface than with a low transparency interface. When the agent is more transparent, and critical of the participant's plan decisions (critical framing), it will be perceived better and trusted more than a complimentary agent as it highlights reasons for error. On the other hand, we expect that a more opaque yet complimentary agent will increase trust in the agent by helping to counteract the overweighing of faults.

3.1 Methodology

Twenty-nine students from an American university were recruited for cash payment. Data were analyzed for 26 (17 men, 12 women, M_{age} = 20.03, SD_{age} = 2.09). Three were omitted from analysis due to technical issues. This experiment involved a 2x2 mixed design with agent transparency as the within-subjects independent variable and communication framing as the between-subjects independent variable. Agent transparency was tested at two levels: (a) L1+2: containing reasoning information, and (b) L1+2+3+U containing reasoning and projection *with* projection uncertainty information. Communication framing was tested as two contrasting attitudes from the agent: (a) *Critical:* highlighting a parameter of the chosen plan that is not satisfied, and (b) *Complimentary:* highlighting a parameter of the chosen plan that is optimal. The user interface varied per condition by showing corresponding pieces of SAT-level information on a map display, in text, and on a sliding bar scale (Figure 5). Prior to the experimental trials, participants received about 1 hour of training. The experiment was divided into 2 blocks of 8 missions. Transparency order and communication framing were counterbalanced within sets of four participants, within which the scenarios where the agent's recommendations were correct and incorrect were held constant. The choice of correct and incorrect scenarios was randomized for each set but kept the 5 correct and 3 incorrect ratio described below.



(a)

(b)

Figure 5. Multiple-heterogeneous UxV control mission plan decision (a) and mission update decision (b) screenshots.

Using the interface presented in Figure 5a, participants were presented with two plan options to complete each mission. Both plans were viable and were balanced in their fulfillment of the mission needs, but optimal plans prioritized the commander's intent and addressed more mission goals and needs. The agent always recommended one plan over the other, and the participant's task was to select the best plan based on current mission requirements. After the participant made the initial selection, an update informed the participant and agent of an aspect of the scenario which had evolved after the initial plans were generated (Figure 5b). Once the update was acknowledged, the agent re-assessed the plans and informed the participant of a parameter that had either *remained or become suboptimal* (criticizing condition) or had *remained or become optimal* (complimentary condition). The participant confirmed or changed the plan based on the new information. While the transparency manipulation applied throughout the entire interaction, the communication framing manipulation only applied post-update. Reliability of the agent was held constant such that its initial recommendation was correct for 5 of every 8 scenarios. The agent's post-update assessment was correct in all scenarios (though no actual recommendation was made for the post-update plan decision). Participants assessed their trust and perception of the agent after each block of scenarios. Within each block, participants

also filled out items for usefulness and reliance following each of the eight initial mission plan decisions.

3.2 Results

To investigate the influence of both transparency and framing, as well as to account for order effects, a series of 2 (low & high transparency) x 2 (complimentary & critical framing) x 2 (transparency block order) mixed model ANOVAs were run with transparency as the within-subjects variable. We reported here only main effects and interactions for transparency and framing; including order in the analysis is important to account for order effects, but is not of specific interest here. Correlations were then run to help with interpretation. In light of relatively low statistical power, we reported trends with a *p*-value greater than .05 but less than .10.

3.2.1 Operator Performance

There were no significant task performances differences. However, for agreement with the agent, there was a main effect for transparency, $F(1,26) = 4.28$, $p = .049$, $\eta^2_p = .141$, as well as an interaction between transparency and framing, $F(1,26) = 5.35$, $p = .029$, $\eta^2_p = .171$. Agreement with the complimentary agent was consistent between transparency conditions. In contrast, agreement with the critical agent was higher in the low transparency automation condition than in the high transparency condition (Figure 6).

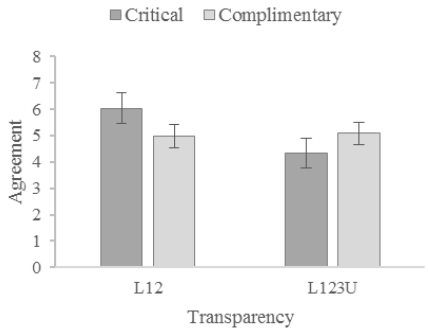


Figure 6. Mean number of scenarios per block participants agreed with the agent.

3.2.2 Operator Trust in Agent

Trust results were mixed. Trust Survey 1 [26] revealed no significant findings, but a main effect trend for framing, $F(1,26) = 3.97$, $p = .057$, $\eta^2_p = .132$. Participants were more trusting when the agent was critical (Figure 7). With regard to trust in the agent's ability to integrate and display analyzed information, Trust Survey 2 [23] revealed a significant main effect for transparency, $F(1,26) = 6.5$, $p = .017$, $\eta^2_p = .200$. There were no other significant findings, but a trend suggests the possibility of an interaction between transparency and framing were there more statistical power, $F(1,26) = 3.4$, $p = .077$, $\eta^2_p = .116$. Participants were relatively distrustful of the low transparency complimentary agent (Figure 8).

3.2.3 Perceptions of Agent

There were main effects of both transparency, $F(1,26) = 4.9$, $p = .036$, $\eta^2_p = .16$, and a non-significant interaction trend between the variables, $F(1,26) = 3.9$, $p = .058$, $\eta^2_p = .13$, for perceptions of agent aptitude with regard to integrating and displaying information [27]. Regarding perceptions of agent aptitude in suggesting plan decisions, there were also main effects for transparency, $F(1,26) = 5.4$, $p = .028$, $\eta^2_p = .17$, and framing, $F(1,26) = 4.8$, $p = .038$, $\eta^2_p = .16$, but an interaction was not significant $p > .10$. In both cases, participants perceived the agent to be more apt when transparency was high. Perceptions of agent aptitude

when integrating and displaying information were higher when the agent framed the update plan critically (Figure 9). For suggesting decisions, the interaction was driven by the difference between the low transparency complimentary condition and the other three conditions. There were no significant (or trends of) differences between condition in perceived automation reliability ($p > .10$).

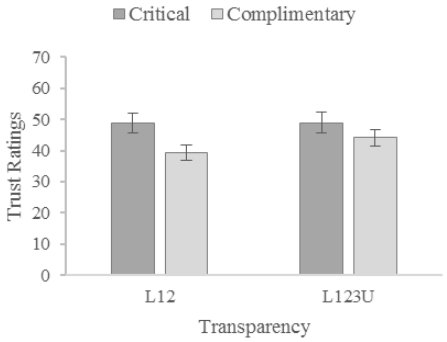


Figure 7. Trust Survey 1 ratings

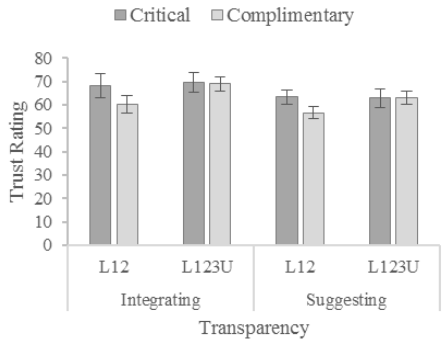


Figure 8. Trust Survey 2 ratings

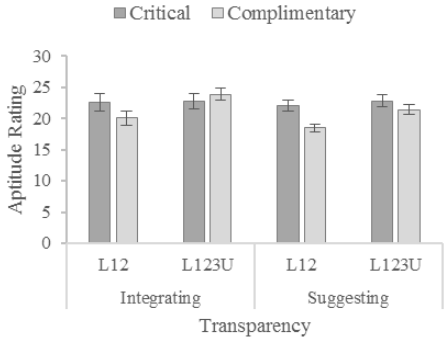


Figure 9. Automation aptitude ratings

3.2.4 Post Scenario Ratings

There was a strong main effect of transparency, $F(1,26) = 49.59$, $p < .001$, $\eta^2_p = .656$, and of framing $F(1,26) = 11.87$, $p = .002$, $\eta^2_p = .314$, for ratings of automation interface usefulness; participants found the high transparency and critical agent configurations to be more useful, respectively (Figure 10). For reliance, there were also main effects for transparency, $F(1,26) = 46.82$, $p < .001$, $\eta^2_p = .643$, and framing, $F(1,26) = 12.56$, $p = .002$, $\eta^2_p = .326$. Participants reported more reliance when transparency was high and the update assessment was critical in nature (Figure 11).

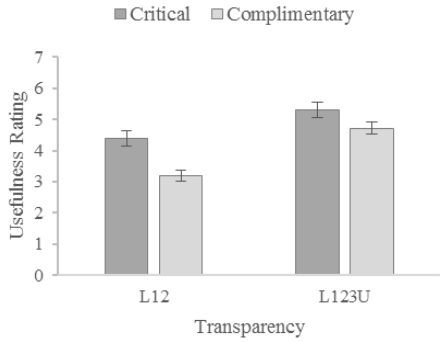


Figure 10. Ratings of usefulness of the automation's transparency information interface.

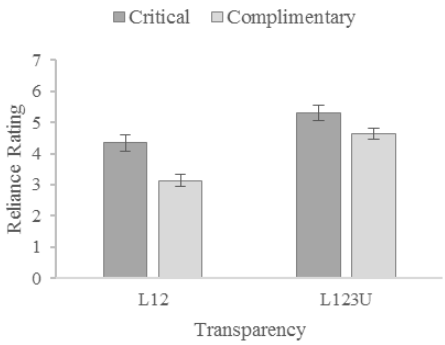


Figure 11. Reliance on the automation's transparency information interface.

4.0 CONCLUSIONS

Results from these studies completed as part of the IMPACT program yielded several insights to the utility of transparency, as well as best practices for implementing information transparency as part of an intelligent agent's interface. Primarily, it was found that agent transparency, as operationalized and implemented according to Chen et al.'s SAT model, is useful for improving performance in complex decision making such as that done in multi-UxV management tasks. Additionally, this performance is increased without a cost to workload. However, it should be noted that response time does increase for a few seconds, which may or may not create an issue depending on the mission environments. Overall, the results showed that perceptions of the IA (its aptitude and usefulness) increased with agent transparency. Trust survey results showed that the complimentary IA was rated as less trustworthy compared with the critical IA, especially in low transparency conditions. Participants tended to agree more with the critical IA when it was more opaque than when it was

more transparent; on the other hand, participants' agreements with the complimentary IA did not differ regardless of the IA's transparency.

Additional research is needed which explores human-agent teaming in a more bidirectional manner. Future studies will examine human-agent teaming with bidirectional communications to evaluate the utility of SAT-based interfaces in a dynamic manner [4]. This can be used to inform the design of field research being done with finalized intelligent agents that are capable of behaving independently. Additional efforts will also focus on developing a repository of user interface design elements (e.g., visualizations) to support SAT-based interfaces.

5.0 REFERENCES

[1] Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence.* Knopf, New York.

[2] Husain, A. (2017). *The Sentient Machine: The Coming Age of Artificial Intelligence.* Simon & Schuster, New York.

[3] Chen, J. Y. C., & Barnes, M. J. (2014). Human-agent teaming for multi-robot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems, 44,* 13-29.

[4] Chen, J., Lakhmani, S., Stowers, K., Selkowitz, A., Wright, J., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomic Science, 19,* 259-282.

[5] Gunning, D. (2016). Explainable Artificial Intelligence. http://www.darpa.mil/program/explainable-artificial-intelligence.

[6] Defense Science Board. (2016). Defense Science Board Summer Study on Autonomy. *Washington, D.C.: Under Secretary of Defense.*

[7] J. Y. C. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia, & M. Barnes. (2014). *Situation awareness-based Agent Transparency* (Technical Report: ARL-TR-6905). Aberdeen Proving Ground MD: US Army Research Laboratory.

[8] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37,* 32–64.

[9] Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *J. Cog. Eng. & Decision Making, 9,* 4–32.

[10] Mercado, J., Rupp, M., Chen, J., Barber, D., Procci, K., & Barnes, M. (2016). Intelligent agent transparency in human-agent teaming for multi-UxV management. *Human Factors, 58,* 401–415.

[11] Bass, E. J., Baumgart, L. A., & Shepley, K. K. (2013). The effect of information analysis automation display content on human judgment performance in noisy environments. *J. Cog. Eng. and Decision Making, 7,* 49–65

[12] Helldin, T. (2014). *Transparency for Future Semi-Automated Systems* (Doctoral Dissertation). Orebro University.

[13] Meyer, J., & Lee, J. (2013). Trust, reliance, compliance. In J. Lee and A. Kirlik (Eds.), *The Oxford Handbook of Cognitive Engineering*. Oxford, UK: Oxford University Press.

[14] McGuirl J. M., & Sarter, N. B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human Factors, 48,* 656–665.

[15] Beller, J. Heesen, M., & Vollrath, M. (2013). Improving the driver–automation interaction: An approach using automation uncertainty. *Human Factors, 55,* 1130–1141.

[16] Calhoun G.L., Ruff, H.A., Behymer, K.J., & Frost, E.M. (2018) Human-autonomy teaming interface design considerations for multi-unmanned vehicle control, *Theoretical Issues in Ergonomics Science, 19,* 321-352.

[17] Fern, L., & Shively, R. J. (2009). A comparison of varying levels of automation on the supervisory control of multiple UASs. *Proc. AUVSIs Unmanned Systems North America* (pp. 10-13).

[18] Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors, 49,* 57-75.

[19] Draper, M. et al. (2018). Realizing autonomy via intelligent adaptive hybrid control: Adaptable autonomy for achieving UxV RSTA team decision superiority (also known as Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT)) (Tech Rep. AFRL-RH-WP-TR-2018-0005). Wright-Patterson Air Force Base, OH: U.S. Air Force Research Laboratory.

[20] Wickens, C., & Dixon, S. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8,* 201-212.

[21] Dzindolet, M., Peterson, S., Pomranky, R., Pierce, L., & Beck, H. (2003). The role of trust in automation reliance," *Int. J. Human-Computer Studies, 58,* 697-718.

[22] Hart, S., & Staveland, L. (1988). Development of NASA TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: Elsevier.

[23] Jian, J., Bisantz, A., & Drury, C. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4,* 53-71.

[24] Parasuraman, R. Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Trans. Syst., Man, Cybern A, Syst., Hum., 30,* 286–297.

[25] Tanner W. P. Jr, & Swets, J.A. (1954). A decision-making theory of visual detection. *Psychological review, 61,* 401-409.

[26] Lyons, J. B., Koltai, K. S., Ho, N. T., Johnson, W. B., Smith, D. E., and Shively, R. J. (2016). Engineering trust in complex automated systems. *Ergon. Des. Q. Hum. Factors Appl., 24,* 13–17.

[27] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q., 13,* 319,

HFM-300 Symposium on Human Autonomy Teaming (HAT)
Trust Me I'm Artificial Intelligence

Fiona Butcher
Defence Science and Technology Laboratory (Dstl)
UNITED KINGDOM

fdbutcher@dstl.gov.uk

Dstl Publication, DSTL/CP110706 (2018). Crown Copyright © 2018 Dstl. This information is licensed under the UK Open Government Licence v3 (to view this licence, visit http://www.nationalarchives.gov.uk/doc/ open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk). The contents of this document should not be interpreted as representing the views of the UK MOD, nor should it be assumed that they reflect any current or future UK MOD policy. The information contained in this document cannot supersede any statutory or contractual requirements or liabilities and is offered without prejudice or commitment.

ABSTRACT

The presentation will explain an ongoing research study that addresses the symposium's objective to consider the human factors of trust in autonomous systems and, specifically, the topic of the ethical and social aspects of HAT. The research aims to validate a conceptual model that has been developed to understand how a person's intention to initially trust an artificial intelligence (AI) system may be influenced by a) prior beliefs and attitudes about AI technology b) social influences and organisational norms and c) a belief in the level of control an individual has when deciding to initially trust an AI technology. Using an experimental design, an on-line survey will be administered to consenting UK military and Defence civilian participants who will complete measures that assess their decision making preferences, knowledge and attitude towards technology in general, as well as AI specifically. In addition, the participants will provide responses to an imaginary scenario to assess why and when they would trust an AI system. The scenario topic reflects a typical example of the trust challenge that humans may face when using new AI decision aides to support nuanced human decision making. The presentation will provide an explanation of the conceptual model, the empirical research method and preliminary research findings.

1.0 BACKGROUND

This empirical study is part of Dstl's Autonomous Systems research programme and sits within a project that is exploring the social, legal and ethical considerations of MoD using these systems in the future. The research aims to validate a conceptual model that has been developed to understand how a person's intention to initially trust an artificial intelligence (AI) system may be influenced by a) prior beliefs and attitudes about AI technology b) social influences and organisational norms and c) a belief in the level of control an individual has when deciding to initially trust an AI technology.

As part of an on-line survey, consenting UK participants will complete measures that assess their decision making preferences, knowledge and attitude towards technology in general, as well as AI specifically. In addition, the participants will be provided with information about how a proposed new AI system will be used in Defence and a scenario in which they will be asked to use the system before then being asked a series of questions about why and when they will trust the AI system.

The scenario is based on a recruitment AI system as this is a typical example of the trust challenge that humans may face when using new AI decision aides to support nuanced human decision making. This topic has been chosen as recruitment is considered a strategic priority for MOD and there are a number of commercial providers who are beginning to offer AI systems to aid recruitment decisions (Jeffery, 2017).

It is hoped that the model being tested in this study will provide evidence to inform future MOD assurance policy as well as military training and education policy.

1.1 What is Artificial Intelligence (AI)?

AI has been listed as one of the 'Eight Great Technologies' (GoScience, 2017, p. 5) for the future economy and is defined as "more than the simple automation of existing processes: it involves, to greater or lesser degrees, setting an outcome and letting a computer program find its own way there." (GoScience, 2016, p.5). Rather than an automated system following prescribed routines the AI is coded with probabilistic equations known as algorithms that autonomously seek out different possible routes to reach a goal.

1.2 Why does AI matter to Defence?

The Defence Concepts and Doctrine Centre describes a future in which MOD personnel will be required to make quick decisions in an increasingly complex and cluttered operational environment, managing large volumes of ever changing information (UK MOD, 2014a). In this new environment it is anticipated that Defence will continue to be a significant recruiter of people in the UK. To meet the demands of this new operating environment it is predicted that over the next two decades military operations will increasingly rely on AI and advanced automated systems as part of military planning and operational decision-making processes (UK MOD, 2014b). Figure 1 below illustrates how it is anticipated future military commanders' decision making will be augmented by the use of AI technologies, automating elements of the decision making process. One of the challenges with using this technology will be to ensure that the way the information has been filtered and processed by the autonomous system is sufficiently trusted by the decision-maker.

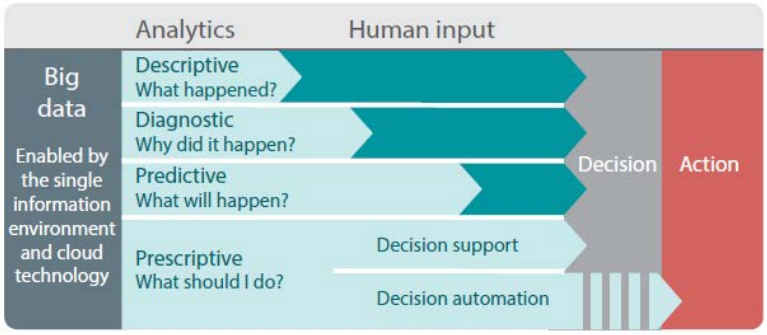


Figure 1: Types of analytics capability[1]

1.3 How is trust in AI being defined?

There is no single agreed definition of trust, however when the technology is an AI that is learning and adapting it is suggested that the interaction becomes more like a human-human interaction. Therefore, interpersonal trust theories (Lewicki & Bunker, 1995; Lewicki, McAllister, & Bies, 1998) have been drawn upon in this study to define 'trust'. These theories suggest a person may initially trust in a new AI system due to affective and social factors. When to trust will be influenced by the context in which the AI system is being used (Hoffman, 2017). The proposition that these factors influence initial trust is supported by Li, Hess, & Valacich, (2008) in which they found when a new information system is proposed initial trust in the

[1] See: UK MOD (2017), Development Concepts Doctrine Centre, Joint Concept Note 2/17, Future of Command and Control www.gov.uk/mod/dcdc

technology is influenced by the reputation of the technology and the belief norms about technology in an organisation.

2.0 STUDY

Following an exploratory literature review an Initial Trust in AI Model (ITAIM) has been developed for this study and is shown in Figure 2. The model is an adaptation of Ajzen's Theory of Planned Behaviour (TPB) (Ajzen, 1991; Ajzen, 2011; Cooke & French, 2011; Yan et al., 2013) linking a person's beliefs with an intention to perform a particular behaviour. The model focuses on initial trust as AI is a new technology that is just beginning to be introduced into military organisations. It will be for future research to investigate how trust develops and changes over time once this new technology has become more established.

The ITAIM describes an intention to trust a new AI system based upon three factors: 1) prior beliefs and attitudes about AI technology, 2) the influence of organisational social norms (this will include the reputation of the AI system), and 3) the perceived level of control in deciding whether to trust (this will include decision making style preferences).

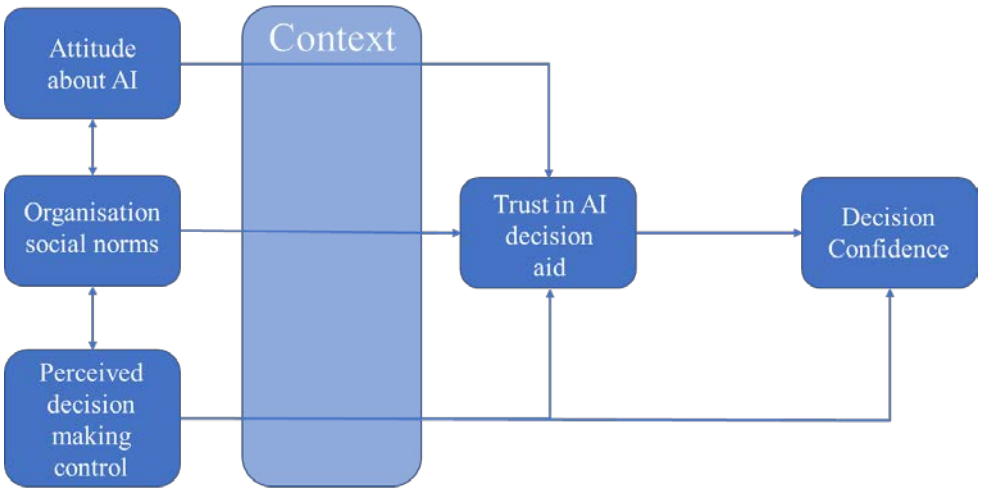


Figure 2: Initial Trust in AI Model (ITAIM)

In applying the model this study aims to answer the following research questions:

- Can knowledge, beliefs and attitudes about AI influence a person's level of initial trust in an AI decision aid?
- Can decision-making preferences (such as a preference for maximum information before making a decision) influence the level of initial trust a person has in an AI system?
- Does the level of initial trust in an AI decision aid influence the level of confidence a person has in their decision to trust?
- Does the reputation of the AI tool influence a person's initial trust in AI tools?

2.1 Overall Methodology

This study will carry out a between-subjects randomised control experiment. The experiment will be completed on-line using the SurveyMonkey platform. Consenting MoD participants will be asked to complete pre-test measures that capture their (a) biographical data (b) decision making preferences and (c)

their knowledge and attitude towards AI. Post-test measures will assess (a) attitudes and beliefs about the AI system described in the scenario (b) level of confidence in their decision to trust the AI system and (c) levels of trust in technology generally. These self-report measures will be combined for administration purposes into single pre and post-test questionnaires.

The dependant variables are the level of initial trust in AI system and confidence in the decision to initially trust. The independent variable is information on the reputation of the tool. The covariates are: knowledge and attitudes about AI, and decision-making preferences.

Listed below are potential hypotheses that will be tested:

H0(null): There is no effect on an individual's intention to initial trust advice from an AI decision system as a result of the individual's beliefs, attitudes or decision making preferences, nor the reputation of the AI system.

H1: Those individuals who have positive attitudes and beliefs about AI technology are likely to initially trust AI decision aids.

H2: Those who seek less information will perceive themselves as having higher levels of control in deciding to initially trust.

H3: Individuals who are likely to initially trust an AI tool are also more confident in the decisions they make using an AI decision aid.

H4: Individuals with high initial trust will also be more confident in their decision to trust.

H5: The reputation of the AI system will influence an individual's initial trust in the system.

An overview of the experimental process is shown in Figure 3 below. The design has been piloted with Dstl employees to test the online survey items suitability before being administered.

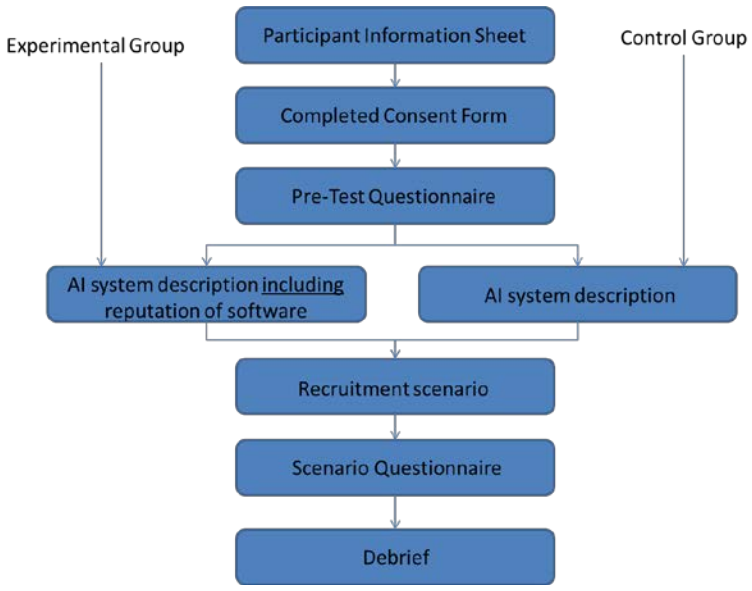


Figure 3: Overview of experimental design for data collection

2.2 Study Procedure

Participants will be recruited from those who have either recently completed or are part of the current intake of 300 personnel on the Advanced Command Staff Course (ACSC) course delivered by the Defence Academy at Shrivenham. This group of participants have a high degree of homogeneity and is a post-graduate level course for mid-seniority military and civilian personnel who will have similar levels of experience of working in Defence. By choosing this group it is hoped to control for factors such as experience and rank.

Participants will be randomly allocated to either the experimental or control group and provided with a unique identifier so that the numbers in each condition can be monitored to ensure that they are equal. All participant information and questionnaire responses will be collected via SurveyMonkey. All consenting participants will be asked to complete the following online questionnaires:

Demographic questionnaire. Participants will be requested to complete a series of demographic questions that will be used to ascertain the extent to which the sample is representative of a mid-seniority military and civilian Defence population. The demographic questions will ask participants to report their gender, whether they are a UK citizen, educational level, role (military or civilian), rank and length of service.

Decision-making tendencies questionnaire. Participants will also be asked to complete a questionnaire on their decision-making tendencies to assess whether their tendencies influence their belief in the level of self-control to decide to trust AI.

Schwartz (2000) suggests that some individuals consistently attempt to seek out the best solution before making a decision (which demands an exhaustive search of the options), while others consistently attempt to find a solution that is satisfactory or good enough given their standards (which can be met by a non-exhaustive search). To measure the degree to which a decision-maker is a 'maximizer' versus a 'satisficer' Schwartz et al. (2002) developed a 13-item Maximization Scale, with maximizing and satisficing at polar opposites of a continuum.

More recently, Misuraca et al. (2015) developed a new measure that incorporates Schwartz et al.'s items and this new measure of decision-making tendencies has been shown to have good construct validity across six dimensions. The 29-item Decision Making Tendency Inventory (DMTI) measures the tendency to maximize, to satisfice, and a third new construct: to 'minimize', which is the tendency to minimize the amount of resources in order to get the minimum of the possible results. Participants respond to each item using a behaviourally anchored 7-point, Likert-type scale (1 = completely disagree,7 = completely agree).

The DMTI questionnaire items will be used and will extend the DMTI research further by investigating whether there is a correlation between the tendency to maximize, satisfice and minimize and actual decision making behaviour (i.e. intention to trust an AI system).

Survey about AI and Machine Learning perceptions. This survey has been adapted from the Ipsos MORI 2017 survey used in the UK public research and engagement conducted on behalf of the Royal Society . The survey was developed in accordance with the requirements of the international quality standard for Market Research, ISO 20252:2012 and the task for the research was to create an evidence base about public perceptions of the potential opportunities and risks of machine learning. By administering items from this survey it will be possible to begin to explore the perceptions of the ACSC participants compared to the UK public.

Once the questionnaires are completed participants will read a short scenario. A recruitment scenario has been chosen as this is a typical example of the trust challenge that humans may face when using new AI decision aides to support nuanced human decision making.

The experimental group scenario description includes persuading statements about the reputation of the AI recruitment system and includes the following words:

…with more than 20 years' experience … Tai systems is considered an ethical system and is used by some of the world's leading companies and in 2017 was listed as one of the top five recruitment software developers in the UK…. reducing the time spent by companies in reviewing candidates by up to 70%...Recruitment policymakers believe we can save time and money by using such technology to accurately find candidates who meet our exact job requirements at an early stage.

The statements outlined above have been designed around three of Cialdini's principles of persuasion; which are scarcity, authority and consensus. After observing real-world influence techniques and reviewing related empirical research, Cialdini (2009) has defined six principles of social influence routinely employed to persuade potential consumers. Cialdini's research shows that people tend to want more of the things they can have less of (e.g. scarcity of time to recruit new personnel), will follow the lead of knowledgeable credible experts (e.g. those software developers who are world leaders in their field) and will also look to the actions of others to determine their own (e.g. the views of other recruitment policymakers).

The **control group scenario description** does not include these persuading statements (see below):

Tai Systems a company that has developed technology that uses Machine Learning (a form of Artificial Intelligence) to help employers to identify the best-suited candidates in the shortest amount of time. A Tai system supplies a broad range of companies that are interested in recruiting new personnel. Tai Systems is on a mission to make recruitment more efficient, engaging and inclusive, applying their technology, helping companies recruit people that will adhere to their values, and thrive in their culture.

Once the participant has read the scenario they will be asked to complete an online questionnaire about the recruitment system. The first seven items in the questionnaire have been developed using Hoffman's 'trust in automation' taxonomy (Hoffman, 2017). This taxonomy is a nuanced classification of human trust in machines as a process that is dynamic over different contexts leading to a variety of trusting relationships. In developing this taxonomy Hoffman has considered the various factors, dimensions and varieties of trust that have been proposed by previous theories and empirical research and recognises the complications and complexities of trust.

The remaining scenario response items are designed to assess the participant's behavioural intentions, attitudes, subjective norms and perceived behavioural control. The design of these items has been informed by Francis, Eccles, Johnston, Walker, Grimshaw, Foy, Kaner, Smith, & Bonetti,(2004) Theory of Planned Behaviour Questionnaires: Manual for Researchers which was developed for researchers throughout the European Union involved in the Research-Based Education and Quality Improvement (ReBEQI) project. The manual has been subjected to comprehensive review and trialling procedures as part of the ReBEQI project and is therefore considered an authoritative guide on the theory of planned behaviour questionnaire development. There is one item (with four behaviourally anchored 7-point, Likert-type scales) that directly measure the participant's attitude towards initially trusting the Tai system. This is followed by four items that measure subjective norms using incomplete sentences that the participant indicates a response to on a 7-point likert-style scale. The final four items are statements with 7-point Likert-style responses that assess the participant's self-efficacy i.e. how confident they are in trusting Tai and how much control they have over whether they can decide whether to trust Tai. Once the participant has completed the questions about the scenario they will be asked about their general attitude and trust in technology.

2.3 Data Analysis Method

A between groups analysis of variance will be conducted to test the individual performance hypotheses with

the Independent Variable:

- Information on the reputation of the tool

and Dependent Variables:

- Level of initial trust in the Tai system
- Confidence in the decision to initially trust the Tai system.

An analysis of covariance will be conducted to check for possible confounding effects due to knowledge and attitudes about AI, and decision-making preferences.

A factorial model (ANOVA) will be used to assess the magnitude and significance of fixed effects on a dependent variable. Accordingly, multiple hypotheses will be tested. To avoid the familywise error an adjustment will be made to significance values using Tukey's method. Averages from the questions in the AI trust questionnaire for each participant will be used to generate the dependent variable; between 1 and 7 representing an individual's tendency to trust or reject AI decisions. Prior to this questionnaire the participants will have also answered two questionnaires; firstly, regarding their knowledge and attitude toward AI, and secondly regarding their decision making processes. The first questionnaire, knowledge about AI, will be processed into three distinct outcomes (high, medium, low). The second will be treated similarly to the dependent variable in that an average score between 1 and 7 will be calculated. Following this, participants will be assigned to two groups given differential information, a control and an experimental group. The fixed effects model will include three independent main-effects variables, and all feasible interactions; the experimental group (control or experimental) and the outcomes of the knowledge and attitudes toward AI and decision making questionnaires (categorical and continuous respectively).

Table 1: Data Analysis Plan

Variable Name	Type	Model position
AI trust questionnaire	(1 to 7)	Dependent variable
Experimental group	Categorical (2 levels)	Fixed effect
Knowledge & attitudes	Categorical (3 levels)	Fixed effect
Decision making preference	Categorical (3 levels)	Fixed effect

The exact nature of the multivariate model used will depend upon the distribution of the response variable.

Qualitative data analysis on the free response question 'why?' the participant will or will not trust will be analysed using thematic analysis to identify themes between the two groups of the participants. Braun and Clarke's (2006) approach to thematic analysis will be performed through the process of coding in six phases to create established, meaningful patterns. These phases are: familiarization with data, generating initial codes, searching for themes among codes, reviewing themes, defining and naming themes, and producing the final report

3.0 NEXT STEPS

Data collection is underway and it hoped that preliminary analysis will be completed by early October so that early results can be shared with the HFM-300 HAT Symposium. The trust and social implications for the future development and operationalisation of AI systems that support human decision making and this

research will provide MoD with insights that should be considered in the policy, doctrine and use of this technology.

4.0 ACKNOWLEDGEMENTS

The author would like to thank colleagues within Dstl and her supervisor at University of Leicester for numerous spirited discussions that informed this paper. In addition, the paper benefitted significantly from comments received during the Dstl technical review process.

5.0 REFERENCES

[1] Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50, 179–211.

[2] Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. Psychology & Health, 26(9), 1113–1127.

[3] Braun, V. & Clarke V. (2006). "Using thematic analysis in psychology". Qualitative Research in Psychology. 3 (2): 93.

[4] Cialdini, R. B. (2009). Influence: Science and practice (5th ed.). Boston: Allyn & Bacon.

[5] Cooke, R., & French, D. P. (2011). The role of context and timeframe in moderating relationships within the theory of planned behaviour. Psychology & Health, 26(9), 1225–1240.

[6] Francis, Eccles, Johnston, Walker, Grimshaw, Foy, Kaner, Smith, & Bonetti,(2004) Constructing questionnaires based on the theory of planned behaviour: A manual for health services researchers. Newcastle upon Tyne, UK: Centre for Health Services Research, University of Newcastle upon Tyne.

[7] GoScience (2016). Artificial intelligence: opportunities and implications for the future of decision making. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf

[8] GoScience (2017). Technology and Innovation Futures 2017. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/584219/technology-innovation-futures-2017.pdf

[9] Hoffman, R.R. (2017). A taxonomy of emergent trusting in the human-machine relationship. In P. Smith & R.R. Hoffman (Eds.) (2017). Cognitive systems engineering: The future for a changing world. Boca Raton, FL: Taylor & Francis.

[10] Jeffery, R., (2018). Don't hire the candidate who presses this key. How algorithms are rewriting the rules of recruitment (And what that means for your job). People Management, December 2017 / January 2018.

[11] Lewicki, R. J., & Bunker, B. B. (1995). Trust in relationships: A model of development and decline. Jossey-Bass.

[12] Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. Academy of Management Review, 23(3), 438–458.

[13] Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. Journal of Strategic Information Systems, 17(1), 39-71.

[14] Misuraca, R, Faraci, P, Gangemi A., Carmeci F.A., Miceli S. (2015). The Decision Making Tendency Inventory: A new measure to assess maximizing, satisficing, and minimizing, Journal of Personality and Individual Differences, 85 (2015), 111–116.

[15] Schwartz, B. (2000). Self-determination: The tyranny of freedom. American Psychologist, 55, 79–88.

[16] Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. Journal of Personality and Social Psychology, 83, 1178–1197.

[17] UK MOD (2014a). Defence Concepts and Doctrine Centre. Strategic Trends Programme: Future Operating Environment 2035, (First Issue). Retrived from https://www.gov.uk/government/publications/future-operating-environment-2035

[18] UK MOD (2014b). Defence Concepts and Doctrine Centre. Strategic Trends Programme Global Strategic Trends - Out to 2045 Fifth Edition. Retrieved from https://www.gov.uk/government/publications/global-strategic-trends-out-to-2045

[19] Yan, Z., Dong, Y., Niemi, V., & Yu, G. (2013). Exploring trust of mobile applications based on user behaviors: An empirical study. Journal of Applied Social Psychology, 43(3), 638–659.

Transparency over Time: The Importance of Pre- and Post-Mission Interactions in Human Autonomy Teaming

Christopher A. Miller
Smart Information Flow Technologies, LLC
319 First Ave N., Suite 400
Minneapolis, MN 55401
UNITED STATES

cmiller@sift.net

ABSTRACT

In this paper, I briefly review recent work on the concept of "transparency" in Human Autonomy Teaming (HAT) and, while agreeing that with findings that it is a desirable property for human interaction with complex autonomous systems, nevertheless I point out a fundamental problem: that the presentation of transparent information at the time of its use in the system is at odds with the workload savings that is, frequently, the goal of including autonomy in the first place. A potential solution to this problem stems from the opportunity to "displace" the conveying of transparent information from the time of use to periods (frequently of lower workload) before and after execution time. I explore this opportunity and draw connections to the desirability of pre-mission planning and training and post-mission explanation and debriefing in human-human teaming, as well as the role of such transparency displacement in situation awareness and trust tuning. Mental Model Synchronization or Reconciliation is suggested as a mechanism by which benefits from displaced transparency might achieve these benefits. Finally, an experimental paradigm is suggested for exploring displaced transparency along with some specific predictions for its benefits.

1.0 INTRODUCTION

As Human-Autonomy Teaming (HAT) becomes more pervasive and complex, the nature and extent of that 'teaming' must be examined more closely. Two or more agents teaming to perform a function necessitate some communication and cooperation to minimize conflict and maximize joint performance. This is the "team work" which adds to "task work" (the work which directly accomplishes desired functions) and increases workload for the agents, though in effective teaming the task work accomplished exceeds the added workload from team work needs.

Trying to perform this added team work—making communication, behaviours, plans and their associated rationales "transparent"— in the moment of task execution seems counterproductive. After all, this "moment of execution" is the time frame of most critical performance constraints. Human performance limitations are, frequently, what drives the inclusion of additional team members (human or autonomous). Yet our focus in developing transparent automation systems has generally been on conveying information during execution, via user interface (e.g., a ground control station) which enables real or near real time awareness of the autonomous system and control inputs to it. While a worthy goal, transparency "in the moment" will inevitably run up against—and may exacerbate—the human workload and attentional constraints that motivated the development and inclusion of the autonomy in the first place. In short, the transparency required to enable HAT may not be possible in the same time frame where the team must perform its functions.

In this paper, I will examine this problem and potential solutions to it from Human-Human teaming approaches. These solutions centre around the concept of "displacing" the conveying of transparent information over time. Perhaps somewhat unexpectedly, this points to the link close linkage between a variety of concepts in Human Factors and Human-Autonomy Teaming: Trust, Mental Model sharing, Team Cohesion as well as the utility of training, after action reviews, explanation and overall Situation Awareness.

2.0 THE PROBLEM OF "TRANSPARENCY"

Autonomous systems, largely by definition, operate at times and in contexts in which the human supervisor/controller is not actively monitoring or issuing control inputs. This may be because inputs are impossible (such as in loss of communications due to jamming or weather or deep space communication time lags) or because humans are not able, willing or expected to intervene (e.g., due to workload constraints, inferior performance or simply to expectations of autonomous functioning). Yet humans still want such systems to behave for our benefit and within the instructions we provide. This puts the human operator in the role of a supervisor or, when further removed, a planner or even a designer

This displacement of control means that behaviour shaping inputs and interactions will, almost inevitably, be displaced relative to action execution. The displacement may well be geographical, but it will always be temporal. We necessarily task automation via commands that precede action—though the duration of the gap may range from micro-seconds to years. As automation becomes more complex and as humans are further temporally displaced from the immediate context of the automation's behaviour, however, such tasking will increasingly be via abstract guidelines, plans, and policies well before actions are taken and decisions are made, and we will be reviewing those actions and decisions (and, ideally, explanations for them) after they have been completed and their effects are known. That is, the conveying of information and the issuing of commands becomes increasingly "displaced" in time and perhaps geography from the locus of action.

This emerging situation with more complex and autonomous systems demands that we re-examined some of what we know about human machine teaming for "displaced interactions". We must ask questions such as the effect of varying degrees of temporal and geographic displacement on maintaining Situation Awareness (and what that means when communications are not possible or expected), what skills or personal traits are needed to accurately, effectively and safely interact with automated systems over increasing degrees of displacement, and what kinds of decision aids and training may make such interactions easier and more effective.

We have begun thinking about one such "displacement effect"—the role of transparency in displaced interactions. "Transparency", here, refers to the ability for the automation to be inspectable or viewable in the sense that its mechanisms and rationale can be readily known. Researchers since at least Charles Billings [1] and Sarter and Woods, [2, 3] have called for greater transparency in automation. Furthermore, recent research has

[1] Billings, C.: Aviation Automation: The Search for a Human-Centered Approach. Erlbaum; Mahwah, NJ (1997)

[2] Sarter, N. B., & Woods, D. D.: How in the world did we ever get into that mode? Mode er-ror and awareness in supervisory control. Human Factors, 37(1), pp. 5-19 (1995)

[3] Sarter, N. B., Woods, D. D., & Billings, C. E: Automation surprises. In: Handbook of hu-man factors and ergonomics, 2, pp. 1926-1943 (1997)

generally confirmed [4,5,6,7] that such transparency yields better human situation awareness, trust and, frequently, better overall human-machine performance than systems which include less or no transparency.

But there is a problem inherent in transparency, especially in displaced interactions, that begs for an answer to the displacement question. As noted previously [8] not all information about what an "autonomous" agent is doing and why can be shared if there is to be any workload savings in a multi-agent team. Indeed, for the human to attend to and process any information about what an "autonomous" agent is doing will come at the expense of the human's attention devoted to perceiving and understanding other aspects of their world and performing actions within it. We should expect this to produce additional workload and potential loss of situation awareness of other aspects of the work context if attention is oversubscribed. This will, of course, be particularly problematic in high tempo operations and critical response conditions. Finally, worse still, there are frequent situations in military contexts, especially those involving geographic and temporal displacement of team members, where the communication of information is simply not possible—perhaps due to jamming, distances, terrain and/or the need to remain undetectable.

So, we can conclude that transparency is valuable in human-automation interaction, but information to support it frequently cannot be communicated—at least in the moment when it is most needed--for a variety of reasons. Is there a way out of this dilemma?

3.0 THE BEGINNINGS OF A SOLUTION

This "problem of transparency" is hardly new in human experience; we have encountered it in human-human interactions since we began to collaborate and delegate for complex, organized activities, possibly beginning with pre-human hunting in groups. We may now be running into this problem with human-machine interactions precisely because machines are beginning to be complex and sophisticated enough that we are willing to use them in significant autonomous roles.

Humans interacting with other humans have encountered and wrestled with this problem throughout time. Human supervisors attempting to increase their capabilities through organizing and administering (human) subordinates have almost identical requirements: the need to maintain awareness of the performance of their subordinates so as to ensure that the supervisor's intent is accomplished as accurately as possible—even when that subordinate must necessarily act at a geographic or temporal distance, with lags or gaps in communication, from the supervisor.

---

[4] Mercado, J. E., Rupp, M. A., Chen, J. Y., Barnes, M. J., Barber, D., & Procci, K.: Intelli-gent agent transparency in human–agent teaming for Multi-UxV management. Human fac-tors, 58(3), pp. 401-415 (2016)

[5] Lyons, J. B., & Havig, P. R.: Transparency in a human-machine context: approaches for fostering shared awareness/intent. In: Proc. International Conference on Virtual, Aug-mented and Mixed Reality, Springer, pp. 181-190 (2014)

[6] Ososky, S., Sanders, T., Jentsch, F., Hancock, P., & Chen, J.: Y.: Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. In: SPIE Defense+Security, ISOP, pp. 90840E-90840E (2014)

[7] de Visser, E. J., Cohen, M., Freedy, A., & Parasuraman, R.: A design methodology for trust cue calibration in cognitive agents. In: International Conference on Virtual, Aug-mented and Mixed Reality, Springer, pp. 251-262 (2014)

[8] Miller, C. A.: Delegation and transparency: Coordinating interactions so information ex-change is no surprise. In: International Conference on Virtual, Augmented and Mixed Re-ality, Springer, pp. 191-202. (2014)

As I have argued elsewhere [9], acting through a human subordinate is a process of delegation. Delegation inherently involves the communication of intent with oversight (including inspection and correction) of the subsequent performance of that intent by the subordinate. Furthermore, the communication of intent itself forms an "intentional frame" which can serve to increase situation awareness and decrease communication and cognitive processing demands in the future. Sheridan's original definition of supervisory control [10] included the provision that supervisors had to communicate their intent to subordinates or, in Sheridan's words, to "teach" subordinate automation what it should do.

Not surprisingly, humans have already developed approaches to this mitigating problem, at least in human-human tasking and delegation. These involve what might be termed "temporally shifted" or "displaced" transparency—providing transparent information at times other than the critical time of use. Transparency in human and human-automation work is, roughly, the provision of information about what the independent human or automated system is doing and why, potentially at multiple levels of abstraction. Thus, knowledge about what the system will or might do in various circumstances can be provided in pre-mission interactions which we might call planning or training. Similarly, knowledge about what the system has done and why can be provided post-execution in explanations or post-mission debriefings. In all cases, as with in-the-moment transparency, the goal is the synchronization of mental models about what is/will/should be done during missions.

Human to human communication in and near the moment of execution is an extraordinarily rich tool, especially when it is deployed against the backdrop of common cultural and professional understanding of the domain and its goals and methods, and even moreso when it is augmented by mutually-understood professional jargon. It serves to make the communication of intent from supervisor to subordinate, and status from subordinate to supervisor, extraordinarily efficient and rich.

But there is evidence that high-performing human-human teams, especially in high criticality domains, frequently exhibit *less* (and less explicit) communication than do less well-integrated teams. Entin and Serfaty [11] compared teams trained to adopt terse and implicit communication styles during periods of high operational tempo and stress with a control group that had not received such training and found that performance improved significantly with this training. They also found that multiple measures of team behaviour were significantly affected by the training, including particularly an increase in anticipations (and, presumably, a corresponding reduction in the need to explicitly request information) between teammates). Together, they took these findings as indicating that efficient teams were sharing and making more efficient use of mental models which had the effect of both reducing the need for communications and speeding those communications which were necessary.

This finding implies that part of the reason human natural language communication can be so effective, and a large part of the reason well-trained and experienced teams can be so sparse with their communications, is that such team members share an understanding of the domain and of work within it. This understanding is certainly acquired during training and experience, but on a day by day (or mission by mission) basis, it is also acquired through mechanisms such as training and planning before execution, and explanation and debriefing (or after-action reviews) after execution. Below, I will argue that these mechanisms provide "transparent" information to the human team-members—that is, the same kind of information that has been shown to improve performance

[9] Miller, C. & Parasuraman, R.: Designing for flexible interaction between humans and au-tomation. Human Factors, 49(1), pp. 57-75 (2007)

[10] Sheridan, T.: Supervisory Control. In: Salvendy, G., (ed.), Handbook of Human Factors pp.1244-1268. John Wiley & Sons, New York (1987)

[11] Entin, E., & Serfaty, D.: Adaptive team coordination. Human Factors, 41, pp. 312–325 (1999)

from “transparent” displays. They just do so at times other than the time of execution. That is, they provide temporally (and potentially, geo-graphically) displaced information about how an agent is, will, or should behave, or how and why he/she/it did behave, but do so at a time when workload and attentional demands are lower. In short, they provide ***Displaced Transparency***. Displaced Transparency may serve as a solution to the dilemma of transparent information in HAT contexts.

4.0 TRANSPARENT INFORMATION AND WHY DISPLACING IT WORKS

What is transparent information and why should we believe that it could be transferred before or after the temporal context of use? Similarly, why should we believe that there are frameworks that might mitigate the need for or improve the efficiency of real time transfer of transparent information?

Chen, et al., [12] have defined transparency as “… the descriptive quality of an interface pertaining to its abilities to afford an operator’s comprehension about an intelligent agent’s intent, performance, future plans, and reasoning process.” But the emphasis on “the interface” in the above definition puts a, perhaps undue, focus on information that is conveyed *during* execution—when an interface is typically used. I contend that much information which achieves the goals of transparency (i.e., affording “an operator’s comprehension about an intelligent agent’s intent, performance, future plans and reasoning process”) need not be provided only, or even primarily, at that workload-intensive time. This relationship is illustrated in Figure 1.

Indeed, Chen [12] also defines a scale or model for transparency, the Situation Awareness-based Transparency (SAT) scale, which in turn leverages Endsley’s [13] scale for Situation Awareness. Chen’s SAT levels are summarized below along with a summary of the transparency information intended to support them:

[12] Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M.: Situation awareness-based agent transparency (No. ARL-TR-6905). ARL/HRED Aberdeen Proving Ground, MD (2014)

[13] Endsley, M. R.: Toward a theory of Situation Awareness in dynamic systems. Human Factors, 37(1), pp. 32–64 (1995)

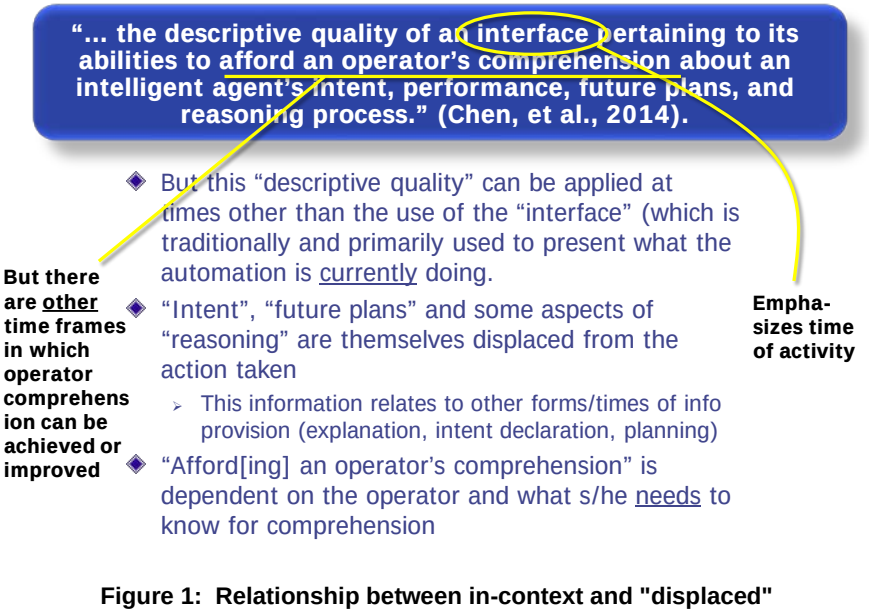


Figure 1: Relationship between in-context and "displaced" transparency in Chen et al.'s (2014) definition.

1. Level 1 SA (What's going on and what the agent is trying to achieve) which is satisfied by providing information about the agent's:
   - Purpose or Desire and Goal selection
   - Process and Intentions (including Planning and Execution) and Progress along that process
   - Performance of both the process and in general.
2. Level 2 SA (Why does the agent do what it does?) which is satisfied by providing information about the agent's:
   - Reasoning process for planning or decision making, including the agent's beliefs and broader purpose, including agent's current beliefs about the environmental and other factors which constrain it
3. Level 3 SA (What should the operator expect to happen?) which is satisfied by information about the agent's:
   - Projection to Future/End state
   - Potential Limitations including likelihood of error and history of performance.

If we take this information content as what is required for effective transparency, then it is worth noting that much of it could be—and probably is in much effective human-human teaming—provided either before or after the time of action execution. "What an agent is trying to achieve?" is something that can and generally should be worked out, at least at a high level, before the subordinate agent is deployed. Intent expressions (which may include goals, purposes, methods and priorities [14]), by their nature, occur before action, while an understanding of why an agent does what it does and what its belief were that might have influenced decisions and actions are

[14] Miller, C.: Delegation for Single Pilot Operation. In: 2014 HCI-Aero ACM, New York (2014)

precisely the focus of the explanations that occur in effective after-action reviews and debriefings [15]. Figure 2 provides a hypothesized annotation of Chen's SAT levels into items which can occur before or after the moment of action.

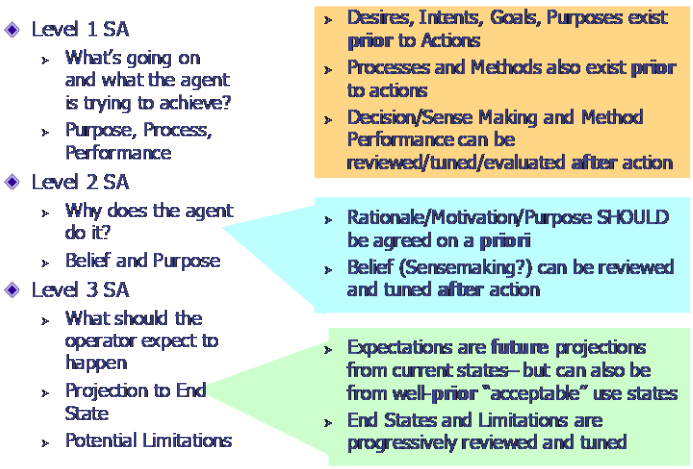


Figure 2: Chen et al's, (2014) SAT levels annotated for their ability to be conveyed either before or after the time of action execution.

We can illustrate the effects and power of pre- and post-mission transparent information transfer with a simple thought experiment. Figure 3 provides a simple timeline for a pre-planned workflow (e.g., for a military reconnaissance mission). Imagine that this timeline is worked out and agreed to by a supervisor and subordinates (e.g., a mission commander and his/her subordinate pilots) before a mission. The plan says that this mission will consist of a reconnaissance task consisting of an Ingress Phase to being at time T and end at time T+1:00 hour, to be followed by a Search Phase to run from the end of the Ingress Phase for another hour. These will then be followed by a decision point whose agreed-upon logic is that if a target has been detected it will be Monitored for another hour and if not, the recon aircraft will Traverse to an egress point.

The simple fact of having made this a prior plan affords substantial situation awareness to the supervisor *even if s/he has no further communication* with the subordinate. For example:

- Given that it is 45 minutes into the mission, the supervisor knows that the subordinate is (supposed to be) engaged in Ingress and even, approximately, where the subordinate is. This is "What's going on?" knowledge—Level 1 SAT.

[15] Tannenbaum, S. I., & Cerasoli, C. P.: Do team and individual debriefs enhance performance? A meta-analysis. Human factors, 55(1), pp. 231-245 (2013)

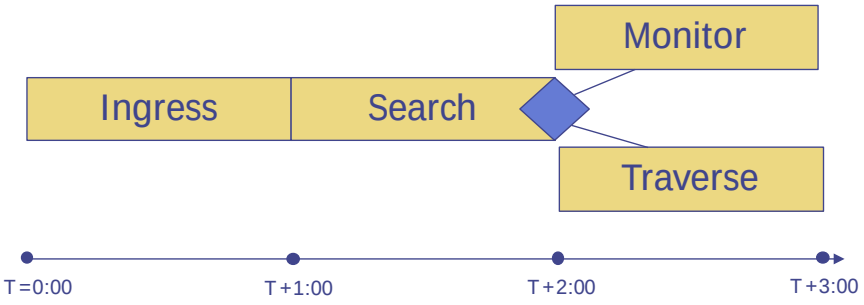


Figure 3: Simple timeline example to illustrate conveyance of "transparent" information before use via planning.

- Furthermore, the supervisor knows why the subordinate is ingressing: to get to the search area and begin search—Level 2 SAT.
- Finally, the supervisor knows that, at time 1:00, the subordinate will transition to Search. This is "What the operator should expect to happen?" knowledge—Level 3 SAT.

Granted, this awareness may be erroneous, if no further communication is available, because the plan has deviated from reality, but the deviations themselves will be probabilistic (i.e., they may not happen) and, if they do happen, they will represent partial deviations from the baseline plan. Note too that having an a priori plan in place makes communication more efficient. Instead of having to report all three levels of SA (what is going on, why, and what to expect next), the subordinate will generally have to report only current status or, perhaps, only deviations. The presence of pre-planned alternatives and acceptable error bounds (as represented partially by the pre-planned decision point in Figure 2) makes it possible to further reduce communication needs. If, anywhere from time 1:00 to 2:00 the subordinate reports finding a target, then the behaviour at the time 2:00 decision point becomes predictable. Even more, if at some time after 2:00 the subordinate reports that it is Monitoring, the superior is entitled to conclude that the reason is because a target was found during Search. Even in the event where the entire plan is made impracticable, having had a shared plan makes communication about future behaviours more efficient by giving all participants a shared ground to build or deviate from. If, for example, a fuel leak makes sustained Monitoring impossible, both supervisor and subordinate will know (at least approximately) where the agent is, how much range is required to get home, that a target has been detected, that the monitoring objective will have to be abandoned, etc. And all of this knowledge can be shared with little or no "in the moment" information exchange or processing.

5.0 TWO TYPES OF DISPLACED TRANSPARENCY

The above example emphasizes the role of pre-mission planning in establishing "displaced transparency", but that is not the only way transparent information can be displaced. After action reviews, debriefings and explanations also provide after-the-fact transparent information as well. Admittedly, this information is not provided in a timely fashion to enable a supervisor to override or correct behaviour which may not be desired for the current mission, but insofar as work with this particular subordinate continues in the future, it does play a mutual learning and trust building/tuning role. By learning how the subordinate thought and behaved in a specific situation, and potentially by offering advice or instruction about how future instances should be handled, the supervisor can shape future behaviour in much the same way that planning shapes behaviour (and information interpretation) before execution. As Tannenbaum and Cerasoli [15] have said, after-action debriefs

are an effective and efficient means of improving team performance and team cohesion. Their meta-analysis of debriefing studies (covering 46 samples and 2,136 individual participants) indicated that on average, debriefs improve effectiveness in future team collaborations and performance over a control group by approximately 25% (d = .67).

Figure 4 illustrates the "displaced transparency" relationship we posit. Most transparency research to date focuses on information which is conveyed in a narrow temporal window around an action or event of interest, perhaps because of the focus on the design of user interfaces to provide "transparent" information. Such research has generally shown improvements in performance, situation awareness and trust when transparency information is provided. But such research has rarely examined workload effects, especially in time critical and overloaded periods and/or has examined it with subjective and coarse-grained tools such as NASA TLX [16]. We posit that there are periods in human-automation interaction, just as there are in human-human interaction, where the human's attention, processing and comprehension capabilities are so sparse and/or over-subscribed, that the attention to transparent information will, at best, be incomplete, and at worst may provide a disastrous distraction.

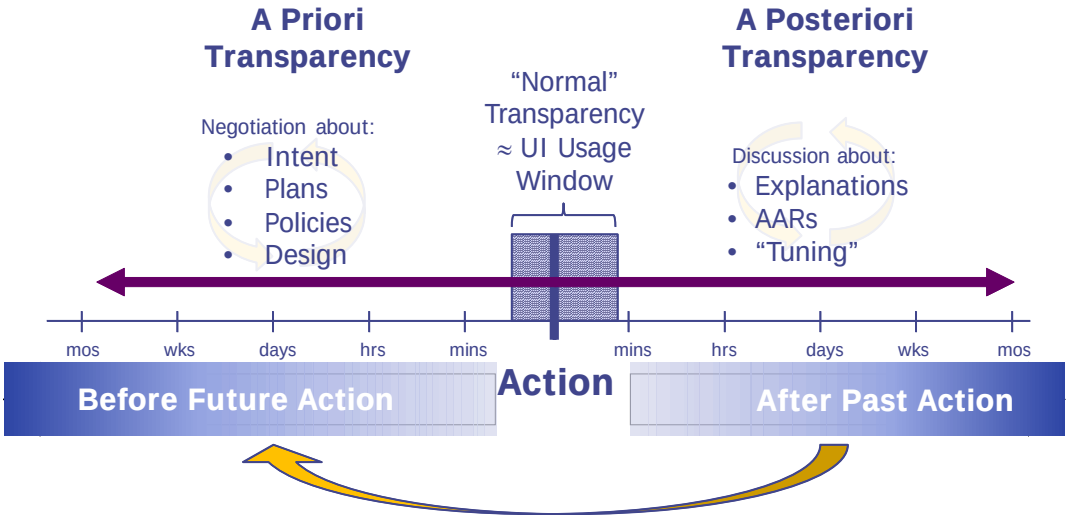


Figure 4. Proposed role of displaced transparency and its effects in team communication and interactions.

We suspect, however, that it will be possible to displace much of this transparent information transfer in time into periods which are substantially less overloaded and/or to provide frameworks and protocols for information transfer which greatly reduce the amount of information which needs to be transferred at critical times in order to achieve the same level of situation awareness. These periods may be anywhere from months or years before an action, or weeks to months after an action. This gives us, conceptually at least, two different periods in which displaced transparency information might be transferred: "***a priori transparency***" (transferring transparent information before the time and context of use) and "***a posteriori transparency***" (transferring it after the context

[16] Hart, S. G., & Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Advances in psychology, 52, pp. 139-183, North-Holland (1988)

of use). In a priori transparency, the processes by which this information transfer occurs are called training, discussion, planning, etc. and they occur before an action and the communication it. The focus of a priori transparency is intent, plans, policies, alternatives and how to decide between them. At extreme durations, the "planning" process becomes one of design of the team, its concept of operations or its equipment. For a posteriori transparency, where discussion occurs after the action, it is called explanation, debriefing, after-action review or various forms of "tuning" (including continuous quality improvement). Iterative cycles through either pre- or post-action discussion will likely only help the understanding of transparent information among team members. It is worth noting, in addition, that the process of pre-action planning and post-action review itself forms a virtuous cycle that can build team understanding for future actions. Although one might reasonably ask how and why providing transparent information *after* an action is helpful, the answer is that it will be helpful in providing team coordination and understanding for future actions, assuming that team must interact again in the future.

This may be why Entin and Serfaty [11] found that verbal communication of explicit information is actually lower among high-performing human-human teams than for less well-integrated teams: because such teams already know what each member will and should do, and trust each other to make good, coordinated decisions. It may also be because communication vocabulary has been streamlined a priori—a benefit we have claimed for playbooks [17]. This is also why Dwight Eisenhower [18] can say "Plans are worthless, but planning is everything"—because planning activities enabled his teams to develop a shared understanding of missions and each other. Explanations and post-mission debriefings work similarly to enable inspection, challenge, tuning and coordination of mental models, though such interactions only have benefit if the team members are able to work together (or work with similarly trained individuals) in the future.

6.0 A MECHANISM FOR DISPLACED TRANSPARENCY EFFECTS

The claim that displacing transparent information can serve much the same purpose as presenting it at the time of need begs the question of how such a process might work. The work of Kambhampati, Chakraborti, Talamadupula and others [19, 20] may provide at least the beginnings of an answer. Although working to provide explanations from machine or robot planners to humans, they nevertheless begin with a human social and cognitive model of the role of explanation itself. They take issue with many past approaches to providing machine explanations as rooted in presenting the machine's reasoning in its own terms—a process they call "soliloquy." [19] They say "Such soliloquy is wholly inadequate in most realistic scenarios where the humans have domain and task models that differ significantly from that used by the AI system." Instead, they say, effective explanation between actors must be a "Model Reconciliation Problem" [19]—that is, a process which is entered into only when a difference between "models" (roughly, mental models) is detected by one or both actors, and which proceeds until this difference is understood and repaired, at least to the level of affording translation between the two models.

---

[17] Miller, C. & Parasuraman, R.: Designing for flexible interaction between humans and au-tomation. Human Factors, 49(1), pp. 57-75 (2007)

[18] Eisenhower, D.D.: From a speech to the National Defense Executive Reserve Conference in Washington, D.C. (November 14, 1957); in Public Papers of the Presidents of the United States, Dwight D. Eisenhower, 1957, National Archives and Records Service, Government Printing Office, p. 818. Downloaded from en.wikiquote.org/wiki/Dwight_D._Eisenhower on March 3 (2018)

[19] Chakraborti, T., Sreedharan, S., Zhang, Y., & Kambhampati, S.: Plan explanations as model reconciliation: Moving beyond explanation as soliloquy. In Proceedings of IJCAI, pp. 156-163 (2017)

[20] Talamadupula, K., Briggs, G., Chakraborti, T., Scheutz, M., & Kambhampati, S.: Coordination in human-robot teams using mental modeling and plan recognition. In: IROS 2014, pp. 2957-2962, IEEE (2014)

Kambhampati has pointed our [21] that except perhaps in testing and evaluation circumstances, explanations are generally neither sought nor provided in circumstances where individuals believe their reasoning and behaviours are compatible or "reconciled". In other words, if I believe that I understand your mental model of the situation sufficiently that you and I would arrive at the same decision I have just arrived at about a course of action to pursue, then no explanation of my decision is necessary for you. It is only when our models diverge (or are believed to diverge) that explanations are invoked—either because you do something which is not compatible with my model of what should be done in the situation, or because you intend to do something that you believe will differ from my decisions or sensemaking in the situation.

Explanation is thus, in their view, largely a process of *Mental Model Reconciliation*. It is provided in order to synchronize the models of those who must work together. It is sought when models do not synch—and it may not be sought when models are assumed to synch, even if they do not.

While explanations can certainly deviate from our actual methods of decision making [22,23], they nevertheless represent how we are trained and acculturated to providing rationalizations for our decision making. Thus, it represents a form of team or group synchronization of thinking in its own right.

This viewpoint explains many phenomena in both explanation and transparency. For example, Entin and Serfaty's [11] finding that well-performing, efficient teams require less, not more, explicit communication can be seen as arising from the fact that such teams are likely to have trained and worked together extensively in the past or (as in their experiment) have been trained to attend to and to anticipate each other's information needs. Thus, their mental models of task and domain are likely to be well-synchronized—meaning that explanatory interactions are less likely to be required, commands can be abbreviated with contextual modifications presumed, and even task-based status information can be abbreviated and given with reduced contextual information because all parties are likely to understand what is needed when.

Similarly, team debriefing after a mission or project, with most of its associated benefits [15] can be attributed to the fact that effective debriefings involve team members interacting about their decisions and their behavioural processes. Simply knowing who knew what when and how they made decisions on the basis of that knowledge serves to educated other team members about their teammates' mental models, while group discussion about how things might be done more effectively in the future creates shared mental models going forward. Group discussion and shared rationalization of events and processes can also serve an educative function, effectively causing the group to converge on a series of broad thought processes and values that, within the "culture" formed by the group, count as valid and reasonable "ways of thinking" (or, at least, of explaining) [24]. Note, though, that this can cut both ways—resulting in adverse cases in "groupthink" [25] and "normalization of deviance" [26]. Finally, explanations offered within a group or to superiors play a social function as well [27],

---

[21] Kambhampati, S.: Personal communication. August 1, Arlington., VA (2017)

[22] Tversky, A., & Kahneman, D.: Judgment under uncertainty: Heuristics and biases. Science, 185(4157), pp. 1124-1131 (1974)

[23] Klein, G.: Naturalistic decision making. Human factors, 50(3), pp. 456-460 (2008)

[24] Miller, C.: Learning to Disagree: Argumentative Reasoning Skill in Development. Ph.D. Thesis. University of Chicago, August (1991).

[25] Turner, M. E., & Pratkanis, A. R.: Twenty-five years of groupthink theory and research: Lessons from the evaluation of a theory. Organizational behavior and human decision processes, 73(2-3), pp. 105-115 (1998)

[26] Vaughan, D.: The Challenger launch decision: Risky technology, culture, and deviance at NASA. University of Chicago Press; Chicago, IL (1997)

serving to reduce social distance and establish or reinforce power structures. While these functions may not directly contribute to a shared mental model about how decisions should be made within the group, they will serve to enhance team cohesion and affiliation.

One sense in which Eisenhower's "planning is everything" [18] may be true is that the activity of planning itself actively promotes model reconciliation. Extensively pouring over plans, rehearsing what could go wrong and what might be needed in contingent situations is a process which conveys and affords opportunities for model synchronization in much the same way that post-mission debriefing does. Participants are likely to emerge from such a process with a richer understanding of each other's models—and having had many opportunities to confront and refine those models in a process which tends toward synchronization. While this, too, can engender groupthink and the suppression of some voices, it also allows team members to "get inside each other's heads" and, thereby, increase their chances of knowing how each other will behave even in unanticipated situations.

Note that I am, in no way, claiming that measuring and computing model reconciliation needs and effects will be easy, especially between humans and machines. Chakraborti and Kambhampati and their colleagues have largely sidestepped this problem by using machine readable symbolic models for "simulated" humans to illustrate efficiency gains in symbolically characterized explanation content. By contrast, humans have evolved cultural, semantic and pragmatic markers rooted in natural language and "body language" for interpreting the need for model reconciliation and then effecting it—a process that, though complex and rich, is far from error proof (cf., [28] for examples of human-human and human-machine model mismatch where communication failed to avoid misinterpretations and, therefore, accidents).

It is worth noting that the process of building and "reconciling" mental models is clearly related to the process of building and tuning "analogic" and "analytic" trust as described by Lee and See [29]. Lee and See define three alternate routes to achieving and tuning trust: affective, analogic and analytic. Affective trust is based on emotion and intuitive pleasure in interacting with the system (or, presumably, the person): we tend to more quickly trust people and experiences that are pleasing to us. Analogic trust is trust based on similarity to individuals or situations we trust or accreditation by individuals or organizations we trust. Analytic trust involves determining how a person or system arrives at its decisions and/or behaviours. It is the slowest and most data-driven of the methods, but also the approach that results in the deepest awareness and most accurate predictions of future attitudes and behaviours. As such, mental model reconciliation is at least one process by which we can learn about (and revise as necessary) the thought processes of another agent, whether human or machine. Initial learning will help to establish whether or not the agent "thinks like it should"—that is, like one with expertise in the relevant domain—and is thus pertinent to accurate tuning of trust through analogic methods. Deeper and more extensive interaction and discussion (through planning, explanation, after action reviews, etc.) will provide the information necessary for analytic trust tuning.

---

[27] Brown, P., & Levinson, S. C.: Politeness: Some universals in language usage, 4. Cambridge University Press; Cambridge, UK (1987)

[28] Miller, C.: Social Relationships and Etiquette with Technical Systems. In B. Withworth and A. de Moor (Eds.): Handbook of Research on Socio-Technical Design and Social Networking Systems, Information Science Reference; Hershey, PA. pp. 472-486 (2009)

[29] Lee, J. and See, K.: Trust in computer technology: Designing for appropriate reliance. Human Factors, 46, pp. 50–80 (2004)

7.0 A SIMPLE ILLUSTRATION

It is reasonably straightforward to show how a process of model reconciliation which occurs either before or after a time frame in which the model is used to make a decision can lead to a reduction in the need to provide "transparency" information in that time frame. Consider again the simple mission sketched in Figure 3, along with an "Observing Teammate" (OT) who, let's assume, can observe and know everything that is happening in the region and to the vehicles involved in the mission, but who is completely unaware of the mission and the intentions of an "Enacting Teammate" (ET). Thus, OT has all and only the "What's going on?" portion of Level 1 SA in Figure 2. Note that this is not even all of the "transparent" information required for Level 1 SAT; the ""What the agent is trying to achieve" portion is missing, as well as essentially all of the information for Level 2 and 3 SAT.

In this scenario, any intent-related interpretation of the world and vehicle states and behaviours from ET will certainly increase OT's SA, especially at levels 2 and 3, but that is because OT's knowledge is essentially nil without such reports. Even observable events in the world (e.g., a headwind) will need to be interpreted for OT in order to provide him/her and understanding of the headwind on "What the agent is trying to achieve" as well as "Why the agent does what it does" and "What will happen next". Without such interpretation from ET, OT's ability to infer this information will be near zero.

OT has none of ET's mental model of the mission (though they may share a model of the world state in this example). All three levels of Chen's SAT [12] must be communicated for OT's full SA. On the other hand, if OT and ET shared the same mental model of the mission and can perceive the same world events, then ET might need to communicate nothing since events would be interpreted similarly and decision making would be identical between ET and OT.

Even when OT can't observe everything that ET can (e.g., in the case of a remote supervisor), the burden of communication is substantially reduced. When events are unfolding as planned, at most ET might need to communicate confirmations that the plan is proceeding as expected. Even when unexpected events occur (e.g., the headwind above), reporting them may be all that is necessary to synchronize ET and OT's models of the impact and revisions necessary and desirable to the mission (revisions in knowledge at Level 3 SA that will *not* have to be communicated since both sides will make them concurrently—though confirmation might still represent useful redundancy). As before, anticipated variations (e.g., the decision point about whether to remain and monitor or simply to traverse) can be communicated much more tersely since both parties know, a priori, what the significance of a detected target will be on this decision point, and/or what the valid reasons are for remaining to monitor vs. traversing.

Even mental model mismatches become easier to detect given this prior planning. Let's say that OT failed to notice that ET detected a target and thus, doesn't understand why ET is transitioning to Monitor rather than Traverse. Simply posing the question in the context of what was expected to be a shared model of the mission identifies the mismatch. "Why are you Monitoring?" conveys a violation of expectations for Monitoring (i.e., the prior detection of a target) and hones in on the piece of information which is needed to repair model mismatch.

This relationship is depicted in Figure 5, illustrating that much less information needs to be conveyed at execution time if full knowledge of plan, purpose and procedures are shared ahead of time. Note that the relative size of the blocks of information associated with Level 1, 2 and 3 SAT are purely speculative. Note too that while it may be true that the higher levels of transparency information may be omitted under conditions of shared a priori transparency information, I do not mean to claim that having this information available is useless.

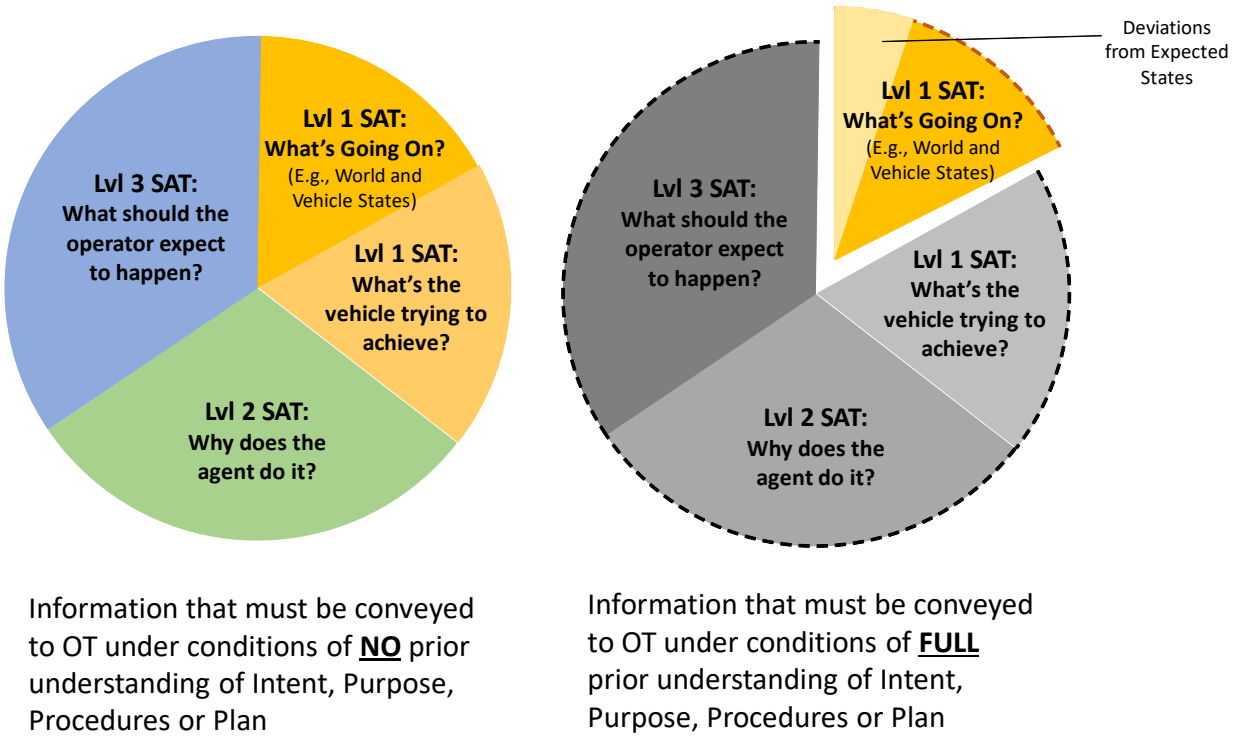


Figure 5. Types of information that need to be conveyed under conditions of no vs. full a priori sharing of Transparency-related information.

Since any sharing of mental models between supervisor and subordinate (even an automated subordinate) will necessarily be subject to uncertainty, continued sharing of higher level information types will help to trap potential mismatches, as well as helping to continue reminding the human members of the team what they *should* be aware of even without reminding.

Finally, although harder to quantify, post-mission debriefings which are later followed by subsequent missions can have similar effects. If, for example, OT learns that ET has a tendency toward speedy completion of missions, s/he might assume a bias or preference in ET for Traversing vs. Monitoring and, in the presence of an ambiguous target detection signal, make more nearly accurate predictions about what ET will do. This represents a variation in the mental model (specifically, in values or priorities) between ET and OT, but insofar as OT understands this about ET (that is, OT's model of ET contains it) it will be accurately factored in to OT's SA and result in accurate understanding of the situation.

8.0 PREDICTIONS AND NEXT STEPS FOR DISPLACED TRANSPARENCY

In this paper, I have presented a fundamental problem in providing "transparent" information in the context of high workload and high complexity. But I have also presented an argument for the desirability and effects of displacing the presentation of transparency information into a priori mission planning interactions and a posteriori explanations and debriefings, along with a hypothesized mechanism for why these effects might be

obtained. We know, from the sources cited above and others, that transparency information frequently provides detectable mission performance benefits. We also know (again from sources sited) that prior mission planning and explanation and debriefings also provide benefits for team cohesion, team satisfaction and team performance. It seems likely that these benefits obtain because they are making use of the same underlying mechanism: the communication of information which promotes situation awareness through mental model synchronization and reconciliation at the time of use. The fact that this information doesn't have to all be transmitted at the time of use, but instead can be spread into lower workload periods before and after usage, is a feature we should use more extensively in design—for human-human and also for human-machine interactions.

Some simple, testable predictions from this model are provided below. While we have not yet been able to conduct experiments to validate these predictions, a simple laboratory test seems eminently plausible. A relevant yet simple scenario is sketched in [20] where a human and a robot are located in a building with a long corridor and multiple side rooms. The human tasks the robot to fetch "a med kit"—one or more of which may be located in the side room(s). Which med kit is desired, expected, and provided is a function of elements of context (e.g., where robot, human, med kit(s) and other potential humans using med kits are located) as well as the robot's decision making algorithm. Of course, mismatches in elements of mental models (e.g., awareness of the physical context and of the robot's decision making process) are exactly what is required to provide SAT knowledge—and can be manipulated in an experimental design. The human may expect the robot to go to a different side room if s/he erroneously believes the med kit to be located there, or the robot may take longer and travel further than necessary if it has an erroneous model of where the human will be located. In this or a similar paradigm, we would predict:

With mental model synchronization between teammates, reduced time, workload effort and even communication bandwidth will be necessary to achieve a similar level of situation awareness compared to conditions without mental model synchronization.

Shifting the communication of transparent information into other time frames (before or after execution) will yield improved situation awareness (with reduced workload) even under conditions of communications restriction or constrained workload for the human recipient, given that model synchronization makes that information comprehensible.

A priori mental model reconciliation will produce more accurate inferences by team members even in unanticipated situations, even with little or no explicit communication of transparent information.

Particularly with regard to post-mission debriefing and explanations, effects of a posteriori model reconciliation will produce increased awareness and ability to predict teammates behaviour even in unanticipated situations in subsequent missions.

We note with interest that [20] reports that the inclusion of mental modeling capabilities in the reasoning of a robot agent, where the robot was modeling the expected reasoning of a human operator and reacting accordingly, produced a 44-75% improvement in robot decision making in terms of avoiding resulting resource conflicts in one analytic experiment they performed.

What is less well documented is the tradeoffs involved in shifting transparency information into time frames before and after it is needed. Somewhere between "no plan survives first contact with the enemy" (implying that "overplanning" is wasteful) and "Plans are worthless but planning is everything" (implying that planning activities are very valuable), there must lie a (probably context-dependent) happy medium. Where is that medium, and what parameters characterize it? It is likely that information theoretic models can provide us with

boundary conditions for this claim, but their relationship to actual human-human (or human-machine) interaction remains to be determined.

Finally, as automation becomes more capable, omnipresent and more "autonomous" in complex work domains, it is becoming clearer that it cannot provide all sufficient transparency information *in the moment* of action execution. Even if the human is capable of understanding it given time, it will all too frequently be the case that s/he will be engaged in other tasks and will be unable to devote sufficient attention and cognitive processing capability in a timely fashion. Instead, we need to strive to enable automation to participate in pre-mission planning and in post-mission debriefing and explanations in order to develop and accurately tune human trust and comprehension frameworks so that available capacity in the moment of use will be sufficient.

I make no claims that the recommendation and predictions provided above are particularly novel. As I said above, human-human interaction has wrestled with the problem of communicating intent and task status in communications-restricted domains throughout much of human history. Therefore, it is not particularly surprising that many of these methods of pre-mission planning and 1post-mission debriefing, along with other techniques for mental model reconciliation and improving the efficiency and predictability of intent communications have long been known and at least occasionally practiced in human-human teaming. It is also not particularly novel as applied to human-machine teaming. There have been calls for improved explanation capabilities for decades and my own work (e.g., [17,[30]]) has been specifically focused on improved human-automation intent communications for nearly 20 years now. I hope, however, that by showing the links between transparency, trust, situation awareness, explanation, debriefing, planning and mental model synchronization I may foster greater sensitivity to how these phenomena inter-relate and that that may, in turn, serve to make the development and use of human-automation systems more effective within the socio-organizational and macro-ergonomic systems where they are deployed.

Acknowledgments. This paper is entirely the work of the author and was not directly funded by any project or agency. I am indebted to Dr. Jesse Chen for providing a forum for initial thoughts on this topic, and to Rao Kambhampati for the insight that explanations generally need to focus only on mismatches in mental models. A prior version of this paper was presented at HCI International in Las Vegas on July 20, 2018 and appeared in the proceedings of that event.

[30] Miller, C. The FireFox fallacy: Why intent should be an explicit part of the external world in human automation interaction. In P. Smith and R. Hoffman, R. (Eds.): Cognitive Systems Engineering: A Future for a Changing World. CRC Press; Boca Raton, FL. Pp. 269-294 (2017).

Autonomic Plan Monitoring

Braulio Coronado, Dr. Douglas S. Lange
Space and Naval Warfare Systems Center - Pacific
San Diego, California
UNITED STATES OF AMERICA

{braulio.coronado, doug.lange}@navy.mil

ABSTRACT

Autonomous agents have their own plans and monitoring their plans is among the essential tasks of command and control. As we move towards forces of large numbers of autonomous entities, we must decide at what level of detail we will monitor and alter plans. Part of that decision will depend on how well we can provide monitoring tools to the human-autonomy team conducting command and control (C2). In the Intelligent Multi-UxV Planner with Adaptive Collaborative Control Technologies (IMPACT) project, we employed an autonomics framework to manage the plan monitoring task. Fundamental to this approach was the realization that plans could be represented as networks, which are the target of the framework's capabilities. Useful to human-autonomy teaming, the strategies employed to respond to performance issues within a network form abstractions of the situation that can be acted upon or communicated to human operators.

1.0 INTRODUCTION

Future C2 operations will involve large numbers of heterogenous autonomous entities executing detailed mission plans towards mission objectives. As battlespace complexity increases, it will become increasingly difficult for the human-autonomy team conducting C2 operations to track the progress and performance of entities executing plans. A human C2 operator will have high-level supervisory tasks to perform including the monitoring of entity progress. Depending on the operational situation, the human will need to prioritize tasks through traditional manual means or with the help of task-managing automated assistants [1]. The cognitive load on the human will be increased as larger numbers of autonomous entities will need their performance monitored which will draw attention away from other mission critical operations.

Decision-making in C2 is the selection of an action aimed at improving the tactical situation, and is achieved by the human through past experience, judgment, and intuition; or by autonomous Intelligent Agents (IA) through automated reasoning. In both cases, decision-making is performed based on information obtained through tools available to the human-autonomy team. Optimal Situational Awareness (SA) is critical for effective decision-making, as the human-autonomy team can only make decisions to address issues they are aware of and understand.

In this paper we describe our approach to improve SA and decision-making by utilizing autonomic computing techniques to track the performance of autonomous entities executing mission plans. An enabling factor towards this approach is the selection of an autonomics framework that can model mission plans as networks. By modeling plans as networks, we can provide continuous plan quality evaluation for all performing autonomous entities. In addition to evaluating mission plans, we can monitor changes in the environment outside of mission plans to offer courses of action. The goal is to enhance the capabilities of the human-autonomy team by

providing continuous feedback on the current tactical posture, and providing mission plan corrections as required by the operational situation.

2.0 BACKGROUND

Highly automated unmanned systems have shifted the role of the human conducting C2 from platform-level control to supervisory control of teams of unmanned systems. Advances in physical systems, their capabilities, and onboard automation have abstracted manual control of vehicle sensors and actuators to high-level commands that human supervisors can issue remotely. Research conducted in the Intelligent Multi-UxV Planner with Adaptive Collaborative Control Technologies (IMPACT) prototype testbed explores the use of a "playbook" control paradigm approach towards C2 aimed at inverting the human to unmanned system staffing ratio [2]. An IMPACT operator can call a "play", a high-level control abstraction, that encapsulates the constraints and behaviors in a mission plan for unmanned systems to perform. Supervisory control of large numbers of unmanned systems will require increased automation to help manage complexity in the C2 environment. Control of multiple unmanned systems can be represented by hierarchical control loops [3].

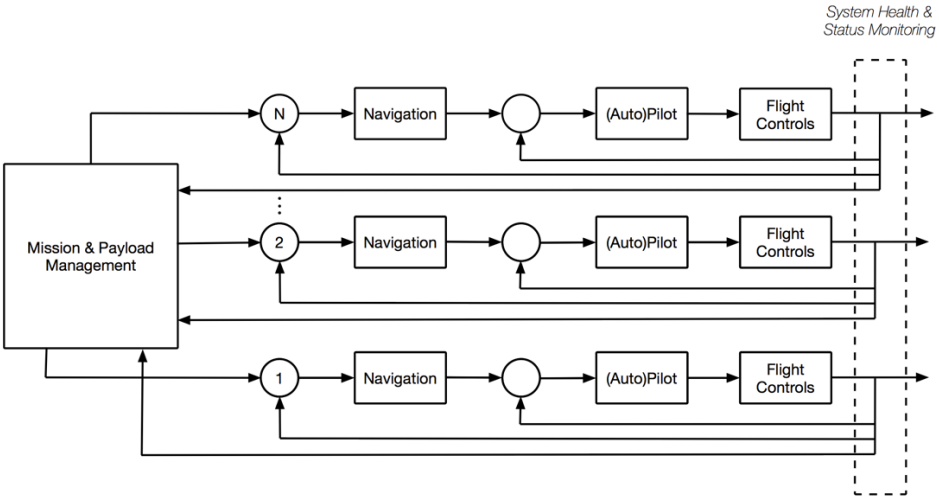


Figure 1: Hierarchical Control for Multiple Unmanned Vehicles. (Adapted from [3])

In Figure 1, we observe a hierarchy of control loops towards C2 of multiple unmanned systems. The Mission & Payload Management loop is shared across all systems, and is the control loop that orchestrates plans and tasks towards mission objectives. The Navigation control loop concerns route waypoints and tasking which set the constraints that define a successful mission. The Piloting and Flight Controls control loop manages the physical system's motion through control of actuators. Finally, the System Health & Monitoring control loop ensures the overall system operates normally. This control loop is a dashed line because it represents a highly intermittent loop from the human's perspective. The human will interact with this control loop the least, as the other loops take priority. In general, the outermost control loops will require the most attention by the human, as they require human experience, judgment, and intuition. The inner control loops concern more technical details of physical system capabilities, and are increasingly amenable to control by automation.

While monitoring of system health and status can be done by the human, it is an opportunity to investigate how

the autonomy can provide this functionality. Our plan monitoring agent, the Plan Monitor, has been implemented in the IMPACT testbed to help manage the monitoring of unmanned system plans.

3.0 THE RAINBOW AUTONOMICS FRAMEWORK

A major problem with many autonomics approaches is they are tied to particular system implementations making scalability an issue for complex systems. For example, being tied to a particular architecture means a lack of flexibility to express new additions that may have not been designed for. Rainbow, the method in this paper, addresses this problem through abstraction of architecture into a formal architecture description language, Acme [4]. The Acme language allows Rainbow to express any system that can be expressed formally.

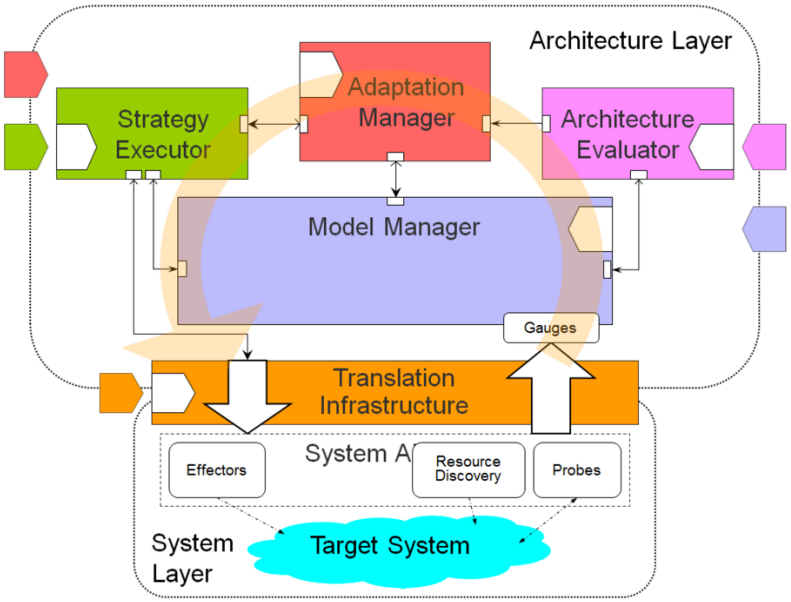


Figure 2: The Rainbow autonomics framework. (From [5], used with author's permission)

An Architecture Evaluator analyzes the model to determine violations of constraints that were defined for the target system. If constraints are violated, then an Adaptation Manager is triggered, which evaluates various strategies defined in a logic-based language (Stitch [6]) to determine the best approach to solving the problem that triggered the need for adaptation. The Adaptation Manager sends the chosen strategy to the Strategy Executor that is tasked with programmatically carrying out the strategy and effecting changes to the target system. The Rainbow framework interacts with the system through Probes, which send data to Rainbow, Gauges, which receive data from Probes to manipulate the model, and Effectors, which are actionable items that cause changes on the target system [5].

4.0 IMPLEMENTATION IN IMPACT

IMPACT allows for a human-autonomy team to supervise multiple heterogenous UxVs through high level goal-oriented plays, which Intelligent Agents (IA) translate into plans to achieve mission goals [2]. By modeling these plans as networks, we can use Rainbow's evaluation and adaptation tools and languages to effect plans and the

systems that are executing them. To utilize Rainbow, we must supply certain software:

- Probes into the tactical data.
- Gauges that aggregate tactical data into useful information for the evaluation of performance.
- Translation of plans into network models.
- Strategies for repairing plans or communicating critical data to the operator.
- Effectors for executing strategies.

Among the most interesting pieces are the strategies. Strategies can range from simply notifying the operator to autonomously re-planning a mission. When notifying an operator, the very fact that a particular strategy was selected is a useful abstraction of the tactical situation. Rather than providing details about movement, the strategy is essentially communicating the final classification of the situation by the system. That the monitor may be permitted to autonomously make changes provides the other end of the full spectrum of contributions that the monitor makes as a teammate of the human operator.

4.1 Probes

An IMPACT instantiation features software modules communicating through a central messaging hub [2]. Data flowing through the hub composes the environment where the human autonomy team conduct C2. Entity types such as vehicle states, plans, tasks, and areas of interest are published to the hub and are used by IAs to handle route planning, resource allocation, and task management [7]. In order for the plan monitor to be an effective teammate, it needs information about the environment; from single events to data streams. Rainbow probes provide a means to sense the environment and feed information to gauges for modeling. In the plan monitor, we employ the ZeroMQ distributed messaging framework within probes to subscribe to messages that can help model the environment. Probes and the messages they subscribe to are:

- Vehicle Probe – Air, Ground, Surface and Track vehicle state messages. Includes vehicle telemetry and is generally frequent with multiple states per vehicle per second.
- Task Probe – Task-related messages which are important when monitoring plan quality for vehicles performing a task as part of a plan.
- Geo Probe – Messages that describe a location in the environment. Areas of interest, lines of interest, points of interest, and restricted zones can potentially be a target or obstacle in a plan and thus important for measuring performance.
- Plan Probe – Messages related to IMPACT play calls. Play automation requests and responses include critical information such as route waypoints that the plan monitor ingests to establish baseline plan quality. This probe will also process messages announcing play activation, cancellation, pause, resume, and completion.
- Critical Area Probe – Messages with information relating to vehicle ability to reach a critical area in the environment. For IMPACT scenarios, a critical area may be a flightline that vehicles must be able to access within a set timeframe.
- Policy Probe – Messages reported by policy-enforcing agents. Policies such as vehicle proximity to no-fly zones discovered by a policy-checker can potentially influence plan performance.

With the exception of the vehicle and critical area, there is little processing performed by probes. In most cases,

probes will simply forward relevant information to interested gauges.

4.1.1 Vehicle Probe Additional Processing

Vehicle state updates are reported several times a second. For the plan monitoring task, updates of this frequency are not necessary for two reasons. One, the turnaround time from a probe report to an effector triggered by information from that report is analyzed through several independent modules in the autonomics framework and generally takes longer than one second. Two, telemetry data is incremental; that is, in most cases there is a small difference between a vehicle state and its next state. To that end, the vehicle probe records the latest vehicle states and forwards only the latest information to gauges every one second. This drastically reduces processing of extra data that likely does not offer much towards effective plan execution monitoring. Furthermore, this probe performs additional processing of vehicle speed in order to better handle telemetry of real vehicles which is further explained in a later section.

4.1.2 Critical Area Probe Additional Processing

Mission requirements may dictate coverage of certain areas of the environment by vehicles; not necessarily live coverage but the ability for a vehicle to be able to reach the area within a certain time. Since the operator can be managing multiple UxVs, their routes, and other tasks simultaneously, maintaining sufficient SA to keep this mission requirement in mind can be a challenge. This is especially true since there are few cues for the operator to recognize that at any point in time, at least one vehicle is able to reach the area quickly. The critical area probe offers to help the operator by handling the tracking of this requirement. This probe is a special case because it is the only active probe used by the plan monitor. As opposed to passively collecting data through subscriptions, the critical area probe actively queries the IMPACT system. This is done by publishing a request to the messaging hub with information on vehicles and an intended target. The route planning IA ingests these requests, calculates a route for each requested vehicle to the target and responds with estimated times for each vehicle to arrive at the destination. The probe then forwards the information to gauges. We set the period of critical area probe requests to one-minute intervals in order to reduce the traffic we generate in the hub and to avoid overburdening the route planner with requests for vehicle timings.

4.2 Gauges and Network Models

The primary function of gauges is to aggregate probe reports into useful data and translate this data into Acme elements to be analyzed by the Architecture Evaluator. Challenges towards this goal are:

- Translating IMPACT entity types into Acme models.
- Deciding the level of model detail necessary for effective monitoring.
- Representing IMPACT plans as networks.

In our previous work [8], we outlined the technical details for automatic Acme element generation through the use of software introspection. Briefly, native system data models represented in a type-introspection capable language, Java in the case of IMPACT, are mapped to abstract Acme element types. At runtime, concrete Acme elements are instantiated per unique data model instance. Data model properties, e.g. runtime Java object field values, are converted to Acme element properties to keep Acme models up-to-date with corresponding data models. The motivation for this model discovery mechanism is to free the system designer from the burden of explicitly declaring detailed model templates at design time; and because it is not always known which properties will be useful at some future time. Most gauges utilize this approach to generate Acme components from data gathered through probe reports. In most cases, components map one-to-one with native data models.

However, the plan monitor uses several Acme software constructs [4] towards the plan monitoring task:

- Components – The primary computation elements.
- Connectors – Interactions between components.
- Groups – Collections of the above.

To represent mission plans as networks, vehicle components are linked, though connectors, to other elements in the plan such as other vehicles, routes, and associated tasks. A vehicle connected to another vehicle matters when considering multi-vehicle plans. A plan can be thought of as a collection of vehicles, routes, and tasks encompassing the plan's network. Therefore, plans can be represented as Acme Groups of connected vehicle, waypoint, and task components. The Acme element generation discussed above converts independent data models such as a vehicle, a task, or an area but does not automatically convert complex data models such as plans. For that, a special plan gauge has been developed to handle plan related messages.

4.2.1 Plan Gauge

The plan gauge ingests plan-related probe reports with information on new plans, or modifications to currently executing plans. Probe reports on new plans include vehicle, route, and task information. Route waypoint metadata includes the expected vehicle speed and any tasks the vehicle performs while on the waypoint. Upon receipt of a new plan, the plan gauge first calculates plan quality baseline parameters up front, effectively caching the expected performance of vehicles to be compared with actual values during plan execution. Waypoint metadata is augmented with these calculated values:

- Distance to next waypoint – Great circle distance of a waypoint's location to the next waypoint's location.
- Expected time in transit to next waypoint – Distance to next waypoint divided by the waypoint's expected vehicle speed.
- Distance to task – Aggregate distance to next waypoint from the first waypoint through waypoints without associated tasks.
- Expected time in transit to task – Aggregate expected time in transit to next waypoint from the first waypoint through waypoints without associated tasks.

Assumptions are that a route leg is a straight line and any obstacles have been considered at plan instantiation, and in the latter two cases that a waypoint without associated tasks is in transit. Next, the plan gauge instantiates an Acme group element with placeholder properties for plan quality to be modified during plan execution. Associated vehicle, route and task Acme components are added to this Acme group. Finally, Acme connectors are created to link vehicle components with vehicle, task and route components in the group. At this point, the plan is essentially modeled as a network in Rainbow with components representing nodes in the network, and connectors representing edges. Grouping plan elements in such a network allows for a simple query of connected elements while performing plan quality evaluation, as opposed to fetching the desired elements from the entire model through a look-up or filter. This technique has the potential to help scale the plan monitoring task as we move towards larger forces of autonomous vehicles.

Probe reports on currently executing plans direct the plan gauge to modify plan models in several ways. An IMPACT operator is able to pause, resume, or cancel plays while the IMPACT system is able to activate or mark plans as completed. The plan gauge responds to these reports as follows:

- Pause – Marks a plan as inactive in the model. Halts processing of plan quality.
- Resume – Marks a plan as active in the model. Resumes processing of plan quality.
- Activate – Marks a plan as active in the model. Initiates processing of plan quality. In this case, a plan has just been created with the default setting of inactive until this message is received.
- Cancel and Complete –
  - Marks plan as inactive to immediately remove the plan from evaluation logic.
  - Deletes Acme connectors generated through the creation of this plan.
  - Deletes task and route Acme components. Vehicle components are not deleted.
  - Deletes the Acme group representing the plan.

We model the environment of the IMPACT system at runtime by maintaining network models of plans and modifying them as needed. This enables us to evaluate the quality of all plans as the scenario and tactical situation evolve over time.

4.3 Plan Quality

When an IMPACT play is called, the operator is essentially directing a vehicle to perform a task at a location to achieve some goal. The plan monitor in this case needs to consider what is being monitored. We can think of monitoring a plan as having two phases: route monitoring and task monitoring. Route monitoring considers a vehicle's real-time performance while traveling towards its destination. E.g. Will the vehicle arrive on time? Is it moving slower than expected? Task monitoring, on the other hand, considers how well the vehicle is performing its task once it has arrived at its destination. When performing plan monitoring, all plans will experience a transition from route monitoring to task monitoring. In the case of multi-vehicle plans, each performing vehicle will transition at a different point along the execution of the plan.

A consideration for plan monitoring, specifically through autonomics, is how and when to communicate plan quality to the operator. The autonomic control loop responds with adaptations when design time constraints are violated. Since a primary goal of the plan monitor is to evaluate the quality of all plans and communicate quality to the operator, we need to consider the case when a plan is in a good state. It would be a misuse of the autonomics control loop to trigger adaptation strategies to communicate a good plan state; since a good plan state does not violate constraints and the system is not in need of adaptation. However, a plan quality report with nominal quality is still useful to the operator's SA so we do not want to restrict plan quality reports to only those in bad states. To that end, the plan monitor delegates plan quality evaluation and communication to a helper module that exists outside the autonomic control loop. This helper module is a delegate of the plan gauge and as delegate has access to gauge functionality such as the ability to access and modify model properties. Normally, a gauge will act on information from a probe report but in the case of plan quality, we process all plans continuously rather than waiting for probe reports. If a plan is in a bad state, the delegate will modify the properties that will be noticed by the Architecture Evaluator to later trigger adaptation for the plan. During plan execution, the delegate queries the Rainbow model every second to processes the quality of all active plans. Quality parameters we communicate to the operator depends on the stage in the plan's execution; that is, whether we perform route monitoring or task monitoring.

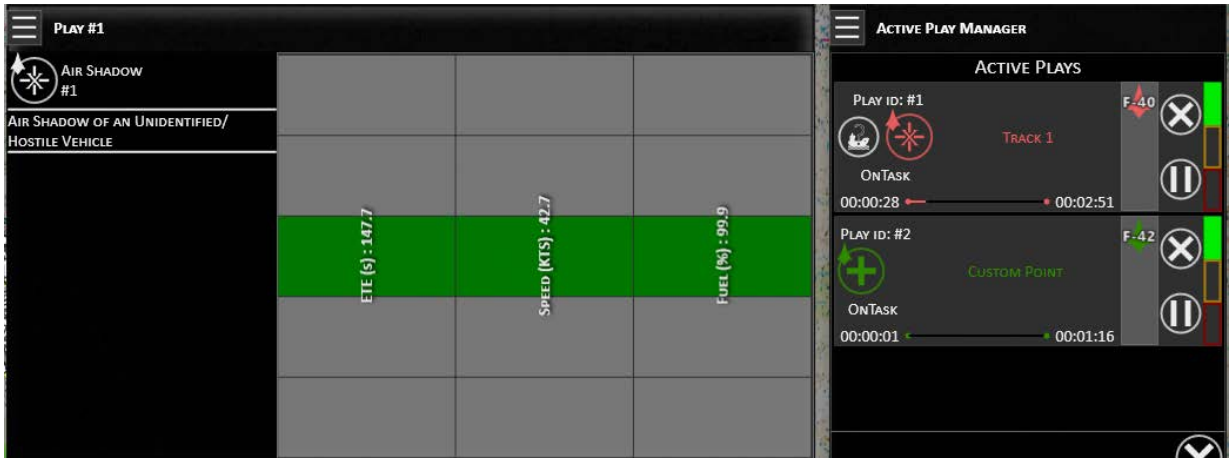


Figure 3: Plan quality for two mission plans communicated to the IMPACT operator.

Quality parameters always include speed and fuel, as these are universal to all plans and need to be monitored at all times; including when a vehicle is not actively executing a mission plan. Since there may be multiple ongoing plans, and each plan may be in a different stage in its execution, the plan monitor process each plan sequentially through the following:

1. Evaluate speed quality parameter.
2. Evaluate fuel quality parameter.
3. Determine if vehicle is currently in transit or on task.
    a. If vehicle is in transit, evaluate expected time to execution (ETE) quality parameter.
    b. If vehicle is on task, evaluate task quality parameter.

We can determine if a vehicle is in transit by comparing the vehicle's currentWaypointID property to the connected route component's metadata. If the waypoint is not associated to a task, the vehicle is currently in transit. Otherwise, the vehicle has arrived and is performing a task. To calculate ETE, the plan monitor compares the expected ETE, which was cached by the Plan Gauge at plan instantiation (section 4.2.1), with actual ETE. If the vehicle is projected to arrive outside of a configurable threshold, then the ETE quality parameter will be of low quality. ETE thresholds can be configured at design time as a constraint in the model. The task quality parameter will depend on the type of task being performed. In general, all tasks will have meta-data that can be used to establish task quality. Some tasks like a point inspect task specify a location for surveillance, while others such as an escort task will have a requirement to maintain a minimum distance [7]. Figure 3 shows how plan quality is represented to the operator. The active play manager features a simple color-coded (Green, Yellow, Red) indicator that provides continuous feedback on the quality of the play. The active play manager tile is always available on the screen, and the operator only needs to glance at the queue of plays to determine if their plans are being performed as expected. This approach enhances SA as plan quality for all plans is summarized into a single metric that can be reviewed at any time. If more detail on the performance of a play is necessary, the operator can click the play in the active play manager to open a plan quality tile (Figure 3). This tile will

display Speed, Fuel and ETE or Task quality parameter values. In the case of multi-vehicle plays, the least performing vehicle quality parameters are shown.

4.4 Strategies and Effectors

Automated adaptation towards improving the state of the system is the goal in autonomic computing [9]. As an instantiation of an autonomics framework, the most interesting part of the plan monitor is the adaptation. Strategies are the method by which the plan monitor offers utility as a team member in the human-autonomy team. By sensing changes in the environment through probes and gauges, and establishing constraints in the model of the world at design time, the plan monitor can offer varying levels of automation. Strategies range from simply notifying the operator of an event, to autonomously re-planning a mission if permitted. Effectors in the plan monitor publish to the messaging hub as it is the primary way of communicating within the IMPACT system.

4.4.1 Notification Strategies

An IMPACT operator will receive aural feedback on actions and events while performing C2. The IMPACT system will announce information on play-related events such as starting a new play, or the state of the vehicles performing background behaviors. Similarly, the IMPACT system will display a text message at the bottom of the screen to draw attention from the operator. The plan monitor can communicate information to the operator through these methods. Notifications by the plan monitor are as follows:

- Fuel – The plan monitor will track fuel levels for all vehicles at all times. Thresholds set in the model represent two fuel levels: warning and critical. A message will be published to the hub once per threshold breach announcing fuel level and vehicle call sign. For example: “Fuel State Warning: Devonian12 has 32% fuel remaining.”
- Flightline – The critical area probe (section 4.1.2) tracks vehicle times to arrive at a specified area such as a flightline. Constraints in the model specify thresholds per vehicle type. For example, air vehicle thresholds can be configured to be 120 seconds from flightline, while ground vehicles can be configured to 360 seconds. As long as there is one vehicle able to reach the flightline within the specified time, the constraint is satisfied. If there are no vehicles satisfying the constraint, the following message is displayed to the operator: “Flightline warning: There are no vehicles in range of the flightline.”.
- Restricted Area – IMPACT provides tools to the operator to designate restricted areas. By drawing a restricted area on the map through the user interface, an entity is generated and its details published to the messaging hub. The plan monitor is aware of restricted areas through the geo probe (section 4.1). Restricted areas are checked for vehicle presence every one second. If the presence of a vehicle is detected, the plan monitor will announce that information to the operator. For example: “Restricted Area Warning: Devonian12 violates area No Fly 2.” This vehicle area pair is tracked for the duration of the violation and a notification is published when the vehicle exits the area; e.g. “Restricted Area Exit: Devonian12 no longer violates area No Fly 2.”.

In a complex IMPACT environment where the operator experiences a heavy workload, notifications can help to maintain SA. Monitoring fuel, flightline and restricted area violations manually through visual checks would require a cognitive load that may not be available depending on the operational situation.

4.4.2 Actionable Strategies

By communicating with other IAs in IMPACT, the plan monitor can execute strategies that cause changes in the

environment. The policy probe (section 4.1) monitors messages issued by the Configurable Operating Model Policy Automation for Control of Tasks (COMPACT) agent that offers compliance checking of policies to IMPACT [10]. COMPACT provides monitoring of vehicle communication ranges and is able to issue a policy summary when a vehicle is projected to lose communications. COMPACT has knowledge of restricted areas that may not exist in the IMPACT environment, and will communicate policy summaries when vehicles are projected to enter these restricted areas. The plan monitor can then introduce information on new restricted areas from COMPACT into the IMPACT system. Through the task manager [11], the plan monitor can initiate the calling of a new play, or re-routing of an existing play to address policy violations. The task manager, if configured to allow human-on-the-loop principles through working agreements [12], will execute the changes autonomously.

- Call Play – Upon detection of a communications policy violation, a strategy will initiate a request to call a new Communications Relay play. Meta-data in the policy violation specifies the affected vehicle's callsign. The strategy will look-up the required play calling data in the model, such as the vehicle's ID to compose a message published to the hub. The route-planner and resource-allocation IAs will select a vehicle to act as relay, if one is available, and the play will be executed without human intervention.
- Introduce restricted areas – A strategy to address a restricted area policy summary will initiate a dialogue with the COMPACT system. The initial policy summary will describe the vehicle and name of the restricted area but will omit details such as the coordinates, shape and size of the area. The strategy will request area details from COMPACT to compose a message that will generate the area and introduce this new knowledge into the IMPACT system. IAs listening on the hub become aware of this new area and will take it into consideration when performing route-planning or resource-allocation.
- Re-route play – Upon receipt of a restricted area policy summary, and generating the area in the IMPACT system through the strategy above. The next step is to re-route the affected vehicle's plan. This is done by a strategy through a simple "ReroutePlay" message with information on the vehicle and play ID. The route-planning IA will adjust the vehicle's route as necessary to avoid the area.

Human-in-the-loop principles can also be applied to these three strategies through working agreement configuration. Instead of the effect happening autonomously, the task manager will add a supervisory task in the task list for the human operator to perform manually.

5.0 REAL WORLD TELEMETRY

The Maritime Autonomous PLatform Exploitation framework (MAPLE) is one of the agents within the IMPACT system [10]. MAPLE is a Ground Control Station (GCS) that allows IMPACT operators control of real unmanned systems. Real world telemetry of actual vehicles is injected into an IMPACT instantiation to practice supervisory control of a combination of real and simulated vehicles. The use of real data introduces unique issues not found while working solely with simulated vehicles. The nature of the physical systems used to report telemetry can involve a level of signal noise. In particular, vehicle speed is a fundamental quality parameter for the plan monitor because it affects other quality parameters such as ETE. To that end, we have implemented an exponential moving average (EMA) algorithm to smooth out vehicle speeds. The vehicle probe (section 4.1.1) tracks the EMA of vehicle speeds by providing more weighting to more recent vehicle speeds captured from the hub. Plan quality evaluation then considers this average speed when reporting speed quality, and when calculating ETE.

6.0 FUTURE WORK

Related to the signal noise in vehicle speeds discovered while working with real vehicles through MAPLE, we intend to investigate the use of machine learning algorithms to learn trends in vehicle types. We expect to find that real vehicles will behave in certain ways under certain conditions. By learning trends in vehicle types, we may be able to adjust the thresholds towards triggering adaptation. Currently, these thresholds are configured at design time but it is technically possible to update thresholds at runtime. This could ultimately improve the plan monitor's ability to determine when adaptation should be used.

Currently, the plan monitor features adaptation strategies to re-route active plays when influenced by a restricted area. This is done by publishing a "ReroutePlay" message to the messaging hub which will invoke re-planning by the route-planning IA. We intend to experiment with strategies that can perform re-planning rather than delegating re-planning to another agent. This would effectively make the plan monitor a planner, and would allow greater control of the environment through strategies.

7.0 SUMMARY

In this paper, we described our approach towards plan execution monitoring of unmanned vehicle mission plans using an autonomics framework. Our approach uses the autonomic framework's probes to sense the C2 environment, and gauges to build and update a model that is maintained at runtime. By modeling mission plans as networks, we are able to provide continuous plan quality evaluation to the C2 operator and discuss how modeling plans as networks can allow for plan monitoring of large numbers of unmanned vehicles. Quality parameters such as vehicle fuel, speed, ETE and task quality are displayed to the operator in a simple and intuitive way that makes it easy to evaluate plan quality at any time. Strategies to notify the operator of important events as they occur can help towards maintaining SA. Finally, strategies to make changes in plans and the environment if permissible through working agreements complete the plan monitor's range of contributions as a team member in the human-autonomy team.

8.0 REFERENCES

[1] Gutzwiller, Robert S., Douglas S. Lange, John Reeder, Rob L. Morris, and Olinda Rodas. "Human-computer collaboration in adaptive supervisory control and function allocation of autonomous system teams." In *International Conference on Virtual, Augmented and Mixed Reality*, pp. 447-456. Springer, Cham, 2015.

[2] Behymer, Kyle, Clayton Rothwell, Heath Ruff, Michael Patzek, Gloria Calhoun, Mark Draper, Scott Douglass, Derek Kingston, and Doug Lange. *Initial evaluation of the intelligent multi-uxv planner with adaptive collaborative/control technologies (IMPACT)*. Infoscitex Corp. Beavercreek, 2017.

[3] Cummings, Mary L., Sylvain Bruni, S. Mercier, and P. J. Mitchell. *Automation architecture for single operator, multiple UAV command and control*. Massachusetts Inst Of Tech Cambridge, 2007.

[4] Garlan, David, Robert T. Monroe, and David Wile. "Acme: Architectural description of component-based systems." *Foundations of component-based systems* 68 (2000): 47-68.

[5] Garlan, David, S-W. Cheng, A-C. Huang, Bradley Schmerl, and Peter Steenkiste. "Rainbow: Architecture-

based self-adaptation with reusable infrastructure." *Computer* 37, no. 10 (2004): 46-54.

[6] Cheng, Shang-Wen, and David Garlan. "Stitch: A language for architecture-based self-adaptation." *Journal of Systems and Software* 85, no. 12 (2012): 2860-2875.

[7] Dzieciuch, Iryna, John Reeder, Robert Gutzwiller, Eric Gustafson, Braulio Coronado, Luis Martinez, Bryan Croft, and Douglas S. Lange. "Amplifying human ability through autonomics and machine learning in IMPACT." In *Micro-and Nanotechnology Sensors, Systems, and Applications IX*, vol. 10194, p. 101941Y. International Society for Optics and Photonics, 2017.

[8] Coronado, Braulio, Eric Gustafson, John Reeder, and Douglas S. Lange. "Mixing Formal Methods, Machine Learning, and Human Interaction Through an Autonomics Framework." In *2016 AAAI Fall Symposium Series*. 2016.

[9] Kephart, Jeffrey O., and David M. Chess. "The vision of autonomic computing." *Computer* 1 (2003): 41-50.

[10] Lucero, Crisrael, Braulio Coronado, Eric Gustafson, and Douglas S. Lange. "Tracking Provenance in Decision Making Between the Human and Autonomy." In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pp. 171-180. Springer, Cham, 2018.

[11] Lange, Douglas S., Robert S. Gutzwiller, Phillip Verbancsics, and Terence Sin. "Task models for human-computer collaboration in supervisory control of teams of autonomous systems." In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2014 IEEE International Inter-Disciplinary Conference on*, pp. 97-102. IEEE, 2014.

[12] Lange, Douglas S., and Robert S. Gutzwiller. "Human-autonomy teaming patterns in the command and control of teams of autonomous systems." In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pp. 179-188. Springer, Cham, 2016.

Discretizing and Managing the Task Environment

Braulio Coronado, Crisrael Lucero, Dr. Douglas S. Lange
Space and Naval Warfare Systems Center - Pacific
San Diego, California
UNITED STATES OF AMERICA

{braulio.coronado, crisrael.lucero, doug.lange}@navy.mil

ABSTRACT

Partners usually need to know what the other partners are doing. Players on football teams, executives in a company, pilot and co-pilot must know what tasks their teammates are taking responsibility for, so they can each do their part. Even when we say that two agents are working together on a task, one can decompose their roles. For instance, if two people are carrying a large plank, one has the job of carrying the right side of the plank and the other takes the task of carrying the left side of the plank. But the two know the other individual has the other task. As humans we know the task assignments through communication (e.g., "I've got it"), doctrine, visual cues, etc. We don't have a direct connection into another human's mind in order to see their task queue, but we might get to look at their to-do list on their office white board. We then keep a model of our partners' task lists in our mind, on a piece of paper, or by some other means. How will autonomous agents know what their human partners are doing, and how will we communicate task lists and negotiate divisions of labor with such agents? While there is research being conducted on activity recognition through robotic vision, this is not an approach that will work for all human-autonomy teaming (HAT) environments.

1.0 INTRODUCTION

The evolution of Command and Control (C2) systems to match a perpetually evolving battlespace provides several challenges for human operators. The introduction of Artificial Intelligence (AI) and autonomous unmanned vehicles increases C2 system capabilities by offloading work from operators, but it is difficult to have complete trust in fully autonomous systems within the context of military operations. Additionally, increased use of autonomy and AI can abstract too much information from human operators, leaving them less engaged in certain mission critical areas of the battlespace. However, without autonomy, C2 operators can experience stressful workloads of processing information, managing tasks, and accomplishing goals in a timely manner.

Successful HAT systems involve a level of interactivity and information-sharing to keep the operator engaged while also balancing workload and stress. Within the context of task management in a C2 environment, this would involve discretizing and digitizing the task environment to allow tasks to be understood by computers and not just humans. The computer-based agents would, in turn, model human performance, track urgency and importance, balance workloads, provide visibility and control, and have the ability to take control when permissible.

In this paper we discuss an approach for discretizing the task environment to be understood by humans and machine agents. We then demonstrate different task management techniques to autonomously aid the operator. Our system also practices human-in-the-loop and human-on-the-loop principles as to not remove the operator from the equation in order to build trust through human-autonomy teaming.

2.0 BACKGROUND AND RELATED WORK

Successful Intelligent Decision Support (IDS) systems contain high-level characteristics that make them useful for humans [1]. We intend to implement some principles of successful human-autonomy collaboration systems into our task management agent. *Interactivity* is important so that the system can work well with other databases, systems, and human users; preferably, the agent would explore the space of possibilities as opposed to just the optimal solution. In this fashion, our task management agent must be able to represent new tasks in a human-centered way while also having the capability of processing requests from other agents in a multitude of different structures. *Event and Change Detection* is necessary to effectively communicate important activities. The task management agent can respond to tasks immediately and switch from a human-in-the-loop to a human-on-the-loop approach in order to avoid potential mission failures. *Predictive Capabilities* are used for predicting the effect of actions on future performance. The task management agent will be able to predict both the workload balance of the user as well as urgency and importance of tasks.

Cognitive assistants in support of task management is being investigated for different military and administrative tasks [2]. This work describes implications for designing a task manager list. *Time constraints should be captured* in order to determine urgency, and provide a clear idea of the overall picture. A task manager should *offer more benefits than just task reminders*, which we expand upon by using an autonomous agent. *Entire tasks can be viewed, but with different perspectives for different kinds of planning.* Our task management agent is capable of showing tasks to be done, but also allows different approaches and plans to approach each task.

The Intelligent Multi-UxV Planner with Adaptive Collaborative Control Technologies (IMPACT) system is a prototype C2 platform for centralized supervised control of simulated autonomous unmanned vehicles [3, 4, 5]. The underlying goal of the IMPACT system is to allow an inversion of the unmanned vehicle staffing ratio and allow a small number of human operators to control a large number of unmanned heterogenous assets. IMPACT uses a "playbook" approach, in which an operator calls a "play" to task teams of vehicles with disparate mission objectives [3]. We implemented our agent, known as the Task Manager, into the IMPACT system to practice discretizing and managing the task environment.

3.0 APPROACH

Providing autonomous assistance for task management requires a few fundamental capabilities:

1. Digitizing the task environment. Tasks not held explicitly in digital form are difficult to manage by computer-based agents. Many tasks are now held only within an operator’s mind.

2. Modeling the performance of agents. This includes building models of human performance on tasks.

3. Tracking the urgency and importance of tasks as the operational situation changes.

4. Balancing workloads to prevent under- and over-loading of agents.

5. Providing visibility and control over the task management process to human operators to act as the ultimate authority. This is done through working agreements with the automation as well as through the user interface.

6. Links to tools for execution of tasks to allow operators to perform tasks from the task manager allowing it to better track completion and performance.

4.0 IMPLEMENTATION IN ALLIED IMPACT

A task manager that begins to address all of the key capabilities described under the approach above was developed and utilized as part of IMPACT. In this section, we discuss the features representing steps along the way to the key capabilities described above.

The Allied IMPACT (AIM) system is a joint international effort and the next evolution of IMPACT. The AIM system integrates core technologies from IMPACT with several autonomous systems from the United States, United Kingdom, Australia, and Canada. Some international systems working directly with our task manager are the COMPACT [6] policy checking system and the Narrative virtual human [7].

The allied capabilities were integrated in preparation for Autonomous Warrior 2018. The goal of the task manager is to aid the operator by accomplishing the fundamental capabilities addressed above within the context of a C2 application controlling several autonomous unmanned vehicles. The AIM system also has other agents working on route planning, resource allocation, constraint solving, and capability enhancements for the operator. The task manager aids with high level tasking, as the vehicles have platform autonomy and abstracts maneuvering from the operator. Due to a high volume of tasks, tasking can easily become overwhelming and disorganized, leading the user to become frustrated and less effective in completing assigned tasks. The task manager seeks to address these issues by determining the user tasks, dividing tasks into a hierarchy, presenting the tasks to the user, and providing a mechanism for their execution [8].

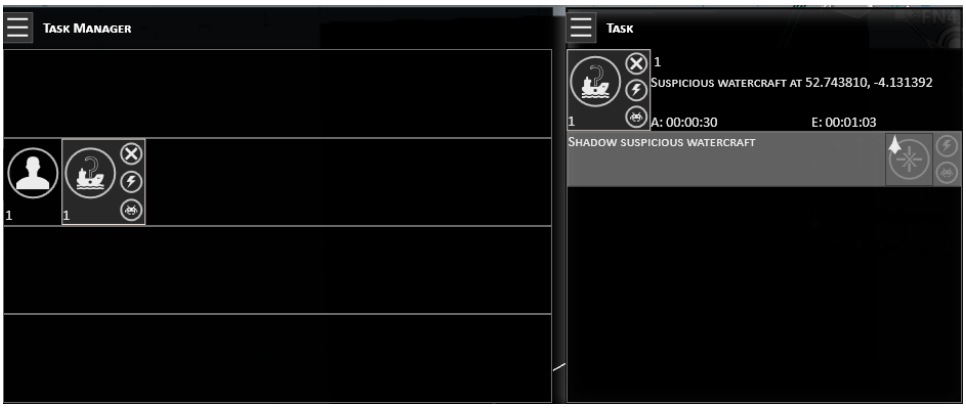


Figure 1: The Task Manager interface showcasing current supervisory tasks.

4.1 Digitizing the task environment

A human operator will have goals and objectives depending on the tactical situation while performing supervisory C2 in AIM. For instance, a scenario in AIM may involve an objective to provide base defense. In this case, the operator's goal is to manage the autonomous assets in order to protect the base. Tasks in the operator's mind may be to distribute assets throughout the environment in order to maintain coverage of the base as a whole, or to execute periodic patrols of specific critical areas of the base. These tasks are known only to the operator, and unless they are written down or recorded in some fashion there is no way of tracking their progress. AIM features a chat messaging system that allows operators to communicate with other operators or commanders. An operator controlling sensors may notice an important event that should be looked at more closely by the operator performing C2, and will send a chat message with information on the event. Similarly, a commander with information not available to the C2 operator may direct the operator to perform a high-level supervisory task. These messages are read, understood and added to a checklist in the operator's mind for a task to be performed later. AIM features agents that can evaluate the tactical situation in real time, through policy checks and environment monitoring to communicate important events that may

require operator intervention. In general, supervisory tasking for the C2 operator is realized through these two methods; chat messages, and changes in the operational situation through events. We employ a method to discretize tasks into digital form that can be understood by both the human and the autonomy, and discuss the different ways that supervisory tasks are generated by the task manager.

4.1.1 Task structure and generation

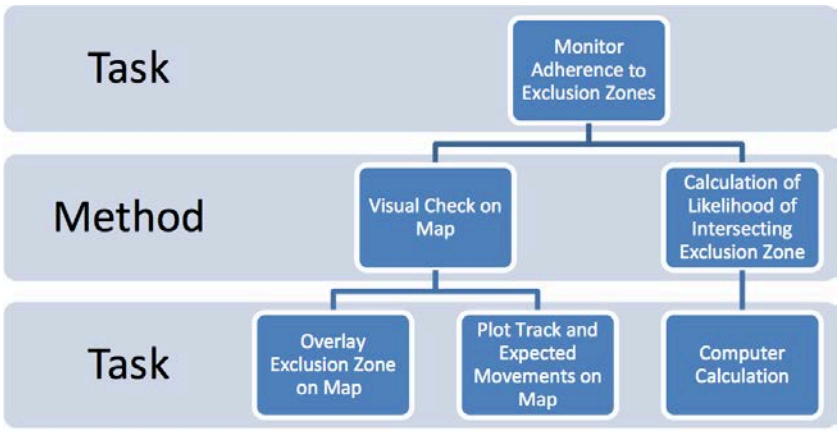


Figure 2: Directed Acyclic Graph (DAG) showcasing the bipartite nature of task method structure. From [8].

Our approach towards discretizing tasks uses a bipartite DAG to decompose tasks into methods, which are in turn composed of tasks, and can be applied recursively [8, 9]. This approach allows us to decompose tasks into a hierarchy of subtasks, where a subtask can be a simple action to perform; and to categorize tasks by the suitability for execution by either the human or the autonomy. Methods such as a visual check on a map are easily performed by a human, while data analytic methods like calculating the probability of a vehicle intersecting a restricted area is more suitable for the autonomy. By decomposing tasks in this way, we can track progress of a task in segments, with completion of each subtask essentially a check on a checklist towards completion of the task as a whole. There are two methods for generating tasks in the task manager; through chat analysis, and through event response.

4.1.1.1 Chat Analysis

An AIM setup will have a C2 operator manning a station that can communicate with other stations; e.g., a sensor operator (SO) station with an AIM configuration specifically tailored towards monitoring vehicle sensor feeds. A SO can control vehicle cameras, switch camera modes, and change camera zoom among other useful methods towards improving quality of surveillance. A SO can then relay information that may be of tactical use for the C2 operator. For example, a SO could notice an area that may be experiencing unexpected activity; and could communicate that information to the C2 operator for action. A commander can communicate with the C2 operator to provide critical data that is normally not obtainable through the operator's tool suite, or to provide supervisory tasking. In both cases, this communication can be done through the AIM chat system. A C2 operator can read the chat, determine if anything needs to be done, and execute or modify plays towards resolution of the purpose of the chat message. To that end, the task manager features a repository of sample chat text that is directly mapped to supervisory tasks. As chat messages arrive through subscriptions to AIM's central messaging hub, the task manager will compare a new chat message to its repository of sample text. Regular expressions are used to identify whether a chat message will invoke the instantiation of a supervisory task in the task manager.

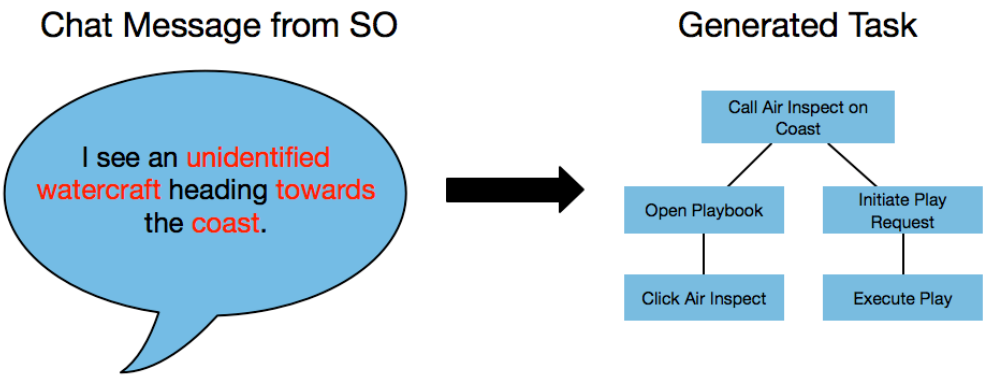


Figure 3: Generation of supervisory task from chat message.

A supervisory task in the task manager is a data structure that represents a high-level task, and is composed of actions towards accomplishing that task. In Figure 3, the generated supervisory task will have Call Air Inspect as the root task, and a subtask to call an air inspect by clicking a button, which the Open Playbook method dictates can be achieved through the playbook. The right side of the DAG is amenable to execution by autonomy as it involves dialogue through request protocols with the resource allocation and route-planning agents. The decision to select the appropriate method for completion of the task, that is, for the human to open the workbook, or for the autonomy to initiate an automated play will be discussed in the Working Agreements section (4.5) of this paper.

4.1.1.2 Events

Events in an AIM instantiation may involve changes in the environment that need to be addressed by the C2 operator. Some events are triggered through the policy-checker in AIM; the Configurable Operating Model Policy Automation for Control of Tasks (COMPACT) [6]. COMPACT provides AIM the ability to perform compliance-checking of policies as the operational situation evolves. Among the policies monitored by COMPACT are projected asset intersection with restricted areas, and ensuring asset communications. COMPACT may have knowledge of restricted areas that AIM does not have, and will communicate policy summaries announcing a projected infraction. The plan monitoring agent [10], which provides runtime mission plan and environmental monitoring, ingests these policy summaries and initiates a dialog with COMPACT to introduce new knowledge of restricted areas known only to COMPACT into AIM. The plan monitor then publishes information on the affected assets that the task manager uses to generate new supervisory tasks directing the C2 operator, or the automated assistant, to re-route affected plans around restricted areas. Similarly, a communications policy summary is processed by the plan monitor in order for the task manager to generate a task towards establishing a communications relay for an asset at risk of losing communications.

4.2 Modeling agent performance

The ability to track the performance of agents may help optimize task management. While some tasks are more suitable for human operators, e.g. those involving decision making, other tasks such as data collection or analysis are better suited for autonomous assistants [8]. Repetitive or tedious tasks may introduce the risk of lowering human operator attention but such tasks are perfectly suitable for execution by autonomy. Human operators are unique and can have preferences towards the types of tasks they prefer to prioritize. An operator's ability to perform tasks can change due to stress, fatigue, or through changes in the environment. Therefore, modeling the human's ability to perform tasks is an important step towards useful task

management. Modeling the autonomous agent's task performance can help to compare and determine which tasks should be assigned to whom.

One approach to model agent performance is through autonomics. Autonomic computing aims to grant systems the ability to self-manage through high-level administrator goals. Inspired by the autonomic nervous system, autonomics-enabled systems exhibit self-healing properties through the Monitor, Analyze, Plan, Execute (MAPE) paradigm [11]. Key to autonomics is the ability to model the system in order to recognize when adaptation is needed. One such autonomics tool is the Rainbow Autonomics Framework, developed at Carnegie Mellon University [12]. Rainbow enables self-adaptation to implementing systems through special software. Rainbow probes sense the system and publish useful data to Rainbow gauges, which reflect the state of the system in models defined in an architectural description language, Acme [13]. An Acme model can be primed with design assumptions, in the form of rules which are analyzed and compared against the model at runtime by an Architecture Evaluator. Upon detection of a rule violation, an Adaptation Manager selects a strategy that is carried out by a Strategy Executor through effectors, which make changes in the target system to bring the system back to a desirable state. Although Rainbow is primarily used to manage computer networks, we have implemented a Rainbow instantiation in the task manager for the purpose of managing supervisory tasking in AIM. To build a model of a task-executing agent, we have developed probes to collect information on the AIM system, and gauges to build models at runtime.

4.2.1 Probes and Gauges

To begin building a performance profile for an agent, we must first be able to uniquely identify an agent; whether they are a human operator, or an autonomous assistant. Meta-data such as user name, login, timestamps, and operating system information can help determine the identity of a human operator. Similarly, a model of an autonomous assistant can be tied to system data. Such data is reachable by probes into operating system properties, and databases backing an AIM instantiation. Aside from agent data, the task manager needs information comprising the AIM environment where the agents perform supervisory tasking. This is done through probe subscriptions to AIM's central messaging hub where entities such as vehicle states, vehicle tasks, and mission plans are published. These entities help provide context for the supervisory tasks to be performed by agents. Probes into the task generation described in the Task structure and generation section (4.1.1) complete the information required for agent modeling. Rainbow gauges process probe reports to maintain a model of the target system. In this case, models for agents, vehicles, tasks, and mission plans are used to represent the environment. Gauges update agent models as supervisory tasks are performed at runtime. Parameters towards measuring agent performance include timestamps when a task is initiated and completed, and its duration. Supervisory tasks can be categorized by the type of the task such as a task to call a play, or a task to respond to a query in chat. Therefore, performance metrics could be established per task type. Rate of cancellation per task type may provide insights for agent performance profiling, and since an operator can transfer tasks to the automated assistant; we can record such actions as parameters as well. The end goal is to be able to predict agent performance on a particular task based on past performance on similar tasks [14].

4.3 Tracking task urgency

Critical events can occur while the operator performs C2 operations in AIM. Such events may require immediate attention and should be addressed in a timely manner. Supervisory tasks generated by the task manager towards resolution of these events are configured with a higher priority than tasks that may not be as urgent. Other tasks have deadlines where failure to complete the task in the specified timeframe has a detrimental effect on the tactical situation. Within the task manager, tasks utilize a priority attribute that is configured by task type, and can be updated over time. For example, query tasks have a low priority by default, as they involve simply answering a query in the chat system. The identity or situation of the request source could affect its priority. While communicating information is an important aspect of C2, there may be more urgent tasks than responding to a chat message. A critical event such as detection of a gate-runner

requires an immediate response by the C2 operator. Therefore, the intruder response supervisory task is by default a high priority task. Tasks generated through chat messages may contain time information. A chat message such as “Alpha company is sending out a foot patrol, with callsign Devonian12, at 1430Z” can generate a task with a priority that is initially low, but will increase as the time approaches 1430Z. The task manager UI could sort tasks in the task queue based on priority, so that the more urgent tasks appear at the top of the queue and elicit more of the operator’s attention.

4.4 Balancing workload

The task manager is capable of balancing the workload of the operator by tasking an autonomous assistant with extraneous tasks. The assistant provides services to receive tasks and execute them, which alleviates operator tasking on lower priority or repetitive tasks. Tasking the automated assistant through Task Manager can be done in two ways: manually assigning tasks with the click of a button, and automatic assignment through load balancing. The operator has the authority to give or take tasks from the autonomous assistant if deemed necessary for control of mission parameters [5]. An approach for task assignment towards enhancing the performance of the human-autonomy team is to balance operator attention with risk [8]. If both the operator and autonomy are available to perform a task, then we want to assign the task to the agent that will complete the task with the least risk. During high workloads, the operator’s available attention is reduced and there is a higher risk of tasks failing; therefore, we should offload more tasks to the autonomy. On the other hand, if the workload is low, we want to assign more tasks to the operator; otherwise we risk reducing the operator’s situational awareness. This work is currently in progress in the task manager, and we are experimenting with ways of modeling operator attention in order to offer more sophisticated load balancing. Currently, we balance the workload based on how amenable a task is towards execution by the human or the autonomy; which was discussed in the Task structure and generation section (4.1.1) of this paper. Later we discuss future work towards load balancing when there are multiple human operators performing from a common task queue.

4.5 Working Agreements

A consideration in HAT is how to determine the tasks permissible for execution by the autonomy. In order to foster trust in the autonomy, and for the autonomy to be a dependable teammate, the human needs to understand what the autonomy can do. One approach towards enabling predictable behavior for the autonomy is through working agreements [15]. Through working agreements, human operators can configure the tasks or types of tasks permitted to the autonomy. In other words, the operator is the ultimate authority on task assignment and execution. Over time, the operator can choose to allow more work to be done by the autonomy to slowly build preferences as to the division of labor. A form of working agreements is in development in the task manager. In its current state, we have configured certain tasks to be executed by the autonomy. In Figure 3, we see a supervisory task to call a play that was generated from a chat message. The working agreement for this task is set to only be executed by the human operator. However, in a future version of the task manager, the operator will be free to change the configuration to a human-on-the-loop approach and always assign this task to the autonomy. In this case, the autonomy will immediately initiate dialogue with the AIM system to call a new play. The operator could instead choose a human-in-the-loop approach and assign this task to the autonomy but require the autonomy to request permission from the operator at certain steps along the play calling process.

4.6 Linked tools

An important milestone towards task management is helping to improve operator workflow. In complex situations, tasks may be streaming into the task manager and it may become difficult for the operator to keep up with the work. Productivity and performance can suffer as the backlog of tasks increases, which can ultimately lead to decreased supervisory C2 effectiveness. The task manager UI features a task queue with supervisory tasks to be completed by the operator, or the autonomy. Certain tasks can link to other UI

elements in the AIM system. In Figure 4, we observe the task manager linking a call play task to AIM's playbook tool, which is a completely separate module from the task manager within AIM. The task manager forwards task meta-data to the playbook in order to pre-populate play request settings towards calling of the play and completion of the task. In a demanding environment, the operator would need to manually open the playbook, select the location where the play is to be called, input the specific parameters to call the play, and finally accept the play. These actions could distract from other potentially more urgent tasks, and since this involves manually copying data, is prone to mistakes through the operator misreading, or mistyping a piece of data. Finally, task meta-data can have timestamps, that can help us track when a task was initiated, and completed which can help with performance profiling.

Figure 4: The Task Manager facilitating the calling of a play by opening a playbook with pre-populated settings from a supervisory task.

5.0 EXPECTATIONS FOR TEAM PERFORMANCE

IMPACT has gone through several evaluations during development and experimentation. With each iteration, the scenarios have become more complicated and the operator task load has increased and become more complex. In IMPACT's Spiral 1 evaluation, Air Force personnel with experience piloting UAVs, and subject matter experts in base defense operations were tasked with conducting C2 in IMPACT [3]. In general, the IMPACT system was rated positively with high scores in potential for use in future C2 operations. The task manager was not included in this evaluation. However, in the Spiral 2 evaluation, the task manager was one of the autonomous agents enabling HAT. In this evaluation, a baseline IMPACT version with minimal autonomy was compared to a full version of IMPACT, with all autonomy enabled [4]. Results showed participants preferred the version of IMPACT with full autonomy enabled, and again positively rated the system for its potential in future C2 operations. The upcoming evaluation of Allied IMPACT in Autonomous Warrior 2018 will feature the latest evolution of the task manager, with new features. One such feature will be providing the human with explanations of AIM scenario or activities with a provenance service [16] through the Narrative virtual human [7] which we believe will be useful for situational awareness. Another feature is the automatic re-routing and play-calling of new plays through COMPACT policy checking [6], which should increase the performance of the human-autonomy team.

6.0 FUTURE WORK

We plan to enhance task generation through chat with Natural Language Processing (NLP) methods. Currently, our regular expression-based method has limitations, as we use direct mappings of text to tasks. This approach is somewhat brittle as a human is able to write a chat message in several ways, while keeping semantics of the message intact. NLP could help provide a more robust chat generation technique.

A future evolution of C2 operations could enable multi-human autonomy teams where multiple operators, aided by autonomy, work together to provide supervisory C2. Multi-Armed Bandit (MAB) problems are basic examples of sequential decision problems using reinforcement learning [17]. MAB is used for regret analysis and maximizing total payoff obtained in a sequence of actions. We intend to use treat task management problems as MAB problems for multi-operator task assignment. This will enhance our task manager's load balancing abilities.

Deep Q-Learning from Demonstration is a deep reinforcement learning approach that leverages human demonstration data to massively accelerate the training process [18]. We plan to train the Task Manager's autonomous agent using data from the operator in other C2 systems to learn how to accomplish complicated tasks and optimize with reinforcement learning. The Task Manager will then be able to recommend and execute more complex strategies learned through human demonstration.

7.0 SUMMARY

In this paper we approached task management in a complex C2 environment. The challenges that come with task management for humans are difficult in an ever-changing and evolving battlespace. Autonomy within C2 has offloaded some work to computers, but a human-in-the-loop or human-on-the-loop approach is necessary to maintain operator engagement. Our task management system follows principles of both intelligent decision support systems and classical task management. Our approach takes an abstract command and converts it to a machine- and human-readable task and goal for completion. Additionally, an autonomous agent exists within our task manager in order to accomplish mission critical tasks that are repetitive, trivial, or unfavorable to the current operator, or the operator may be too overworked to notice. These key capabilities enable a more efficient human-autonomy team, as the human is able to manage supervisory tasking with greater control. The benefits of the task manager will become more pronounced as the complexity in the C2 environment increases.

8.0 REFERENCES

[1] Guerlain, Stephanie, Donald E. Brown, and Christina Mastrangelo. "Intelligent decision support systems." In *Systems, man, and cybernetics, 2000 ieee international conference on*, vol. 3, pp. 1934-1938. IEEE, 2000.

[2] Bellotti, Victoria, Brinda Dalal, Nathaniel Good, Peter Flynn, Daniel G. Bobrow, and Nicolas Ducheneaut. "What a to-do: studies of task management towards the design of a personal task list manager." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 735-742. ACM, 2004.

[3] Behymer, Kyle, Clayton Rothwell, Heath Ruff, Michael Patzek, Gloria Calhoun, Mark Draper, Scott Douglass, Derek Kingston, and Doug Lange. *Initial evaluation of the intelligent multi-uxv planner with adaptive collaborative/control technologies (IMPACT)*. Infoscitex Corp. Beavercreek, 2017.

[4] Draper, Mark, Allen Rowe, Scott Douglass, Gloria Calhoun, Sarah Spriggs, Derek Kingston, Laura Humphrey et al. Realizing Autonomy via Intelligent Hybrid Control: Adaptable Autonomy for Achieving UxV RSTA Team Decision Superiority (also known as Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies (IMPACT)). No. AFRL-RH-WP-TR-2018-0005. 711 Human Performance Wing Wright-Patterson Air Force Base United States, 2018.

[5] Dzieciuch, Iryna, John Reeder, Robert Gutzwiller, Eric Gustafson, Braulio Coronado, Luis Martinez, Bryan Croft, and Douglas S. Lange. "Amplifying human ability through autonomics and machine

learning in IMPACT." In *Micro-and Nanotechnology Sensors, Systems, and Applications IX*, vol. 10194, p. 101941Y. International Society for Optics and Photonics, 2017.

[6] Lucero, Crisrael, Braulio Coronado, Eric Gustafson, and Douglas S. Lange. "Tracking Provenance in Decision Making Between the Human and Autonomy." In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pp. 171-180. Springer, Cham, 2018.

[7] Nowina-Krowicki, Marcin, Peter Argent, Crisrael Lucero, Douglas Lange, and Steven Wark. "Now This Changes Everything: Managing Provenance using Multimedia Narrative."

[8] Gutzwiller, Robert S., Douglas S. Lange, John Reeder, Rob L. Morris, and Olinda Rodas. "Human-computer collaboration in adaptive supervisory control and function allocation of autonomous system teams." In *International Conference on Virtual, Augmented and Mixed Reality*, pp. 447-456. Springer, Cham, 2015.

[9] Chandrasekaran, Balakrishnan, Todd R. Johnson, and Jack W. Smith. "Task-structure analysis for knowledge modeling." *Communications of the ACM* 35, no. 9 (1992): 124-137.

[10] Coronado, Braulio, Eric Gustafson, John Reeder, and Douglas S. Lange. "Mixing Formal Methods, Machine Learning, and Human Interaction Through an Autonomics Framework." In *2016 AAAI Fall Symposium Series*. 2016.

[11] Kephart, Jeffrey O., and David M. Chess. "The vision of autonomic computing." *Computer* 1 (2003): 41-50.

[12] Garlan, David, S-W. Cheng, A-C. Huang, Bradley Schmerl, and Peter Steenkiste. "Rainbow: Architecture-based self-adaptation with reusable infrastructure." *Computer* 37, no. 10 (2004): 46-54.

[13] Garlan, David, Robert Monroe, and David Wile. "Acme: An architecture description interchange language." In *Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research*, p. 7. IBM Press, 1997.

[14] Lange, Douglas S., Robert S. Gutzwiller, Phillip Verbancsics, and Terence Sin. "Task models for human-computer collaboration in supervisory control of teams of autonomous systems." In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2014 IEEE International Inter-Disciplinary Conference on*, pp. 97-102. IEEE, 2014.

[15] Lange, Douglas S., and Robert S. Gutzwiller. "Human-autonomy teaming patterns in the command and control of teams of autonomous systems." In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pp. 179-188. Springer, Cham, 2016.

[16] Crisrael Lucero, Braulio Coronado, Oliver Hui, and Douglas S. Lange. Exploring Explainable Artificial Intelligence and Autonomy through Provenance. *IJCAI-ECAI 2018 2nd Workshop on Explainable Artificial Intelligence*, 2018.

[17] Bubeck, Sébastien, and Nicolo Cesa-Bianchi. "Regret analysis of stochastic and nonstochastic multi-armed bandit problems." *Foundations and Trends® in Machine Learning* 5, no. 1 (2012): 1-122.

[18] Hester, Todd, Matej Vecerik, Olivier Pietquin, Marc Lanctot, Tom Schaul, Bilal Piot, Dan Horgan et al. "Deep Q-learning from Demonstrations." *arXiv preprint arXiv:1704.03732*(2017).

Automated Aiding of Image Analysts as a Function of Sensor Type

Dr. Holly Zelnio and Dr. Mary Fendley
711th Human Performance Wing
Wright State University
UNITED STATES

holly.zelnio@us.af.mil

ABSTRACT

The use of an automated Decision Support System (DSS) to aid an image interpreter performing a target classification task is analyzed to ascertain performance improvement due to human automation cooperation. The performance of the image interpreter and DSS is analyzed from a human factors perspective. This research performs a human subject experiment measuring classification performance as a function of classification accuracy, interpretation time, confidence in the decision, and trust in the decision support system (DSS). In addition, multiple types of automated support were compared as were multiple types of imaging sensors. The imaging sensors that were used in this study were electro-optical (EO), infrared (IR), and synthetic aperture radar (SAR). The DSSs were developed using cognitive task analysis and previous experimentation. The results show that different types of sensors require different DSSs for optimal performance. Results suggest that EO requires less automated support while the more non-literal sensors (e.g., IR and SAR) need more automated support for highest performance. This research also investigated how a DSS itself can cause a bias influencing performance. The performance of the human and machine team can be enhanced by the presentation of key information. It was shown that the level of automated analyst aiding affects both the quality and timeliness of the decision as well as the trust in the DSS.

1.0 BACKGROUND

Trust has to be balanced for optimum performance. Either too much or too little trust can result in failures [1]. To maintain trust in automation, Sarter suggests that the human should understand the basis of the automated aids but should have the task responsibility. The human can use the automation for help, but the automation is now a means to the end of a human responsible task [2]. Increased transparency has led to increased trust, but higher decision times and workloads [3]. Increased trust and performance is suggested by findings if contextual information is provided by the automated DSS [4]. Evidence indicates that, over time, experienced workers can learn to calibrate their trust for the automated system used in their job [5].

Biases can appear when making judgments. It is difficult to fully eliminate a bias from a decision. However, two approaches to help mitigate bias includes: 1) additional training [6] or knowledge needs to be given to the individual or 2) the task needs to be structured to not create these biases. The task structuring can be designed to coincide with an individual's natural thought process to help inhibit bias [7]. This approach can be accomplished by examining where the bias is coming into play either via the person or the task/environment. Several different approaches have been studied to suppress bias including using structured processes [7], applying a three stage framework [8], improving the environment [9], and adopting the specific debiasing strategies of technical, motivational, and cognitive [10].

Cognitive bias happens when one deviates from good judgement and can occur or increase when introducing information to a decision maker. Arnott [11] has an extensive list of biases that influence people. Fendley [12] shortened this list to specifically biases that are found in object identification. The dominant biases that were identified by Zelnio and Fendley [13] when analysts perform object identification were confirmation and anchoring. Confirmation bias is present when the data and interpretation favors a single decision over alternatives [14-16]. Anchoring bias is being reliant on an initial value or given information [17, 18]. Overall, it has been suggested that a DSS should mitigate bias [19, 20]. The number of studies performed confirming bias exists heavily outweighs the number of studies that try to mitigate bias [21]. It is understood that bias cannot be eliminated. It is important to acknowledge that it exists and then try to design the support system to use the biases to our advantage or to try to make them negligible.

DSSs can be built in various ways including: a framework for the characterization of the decision situation [22], three factor function for designers to structure representation [23], the summary table of aiding requirements (STAR) table [24], and the operator function model (OFM) [25]. An OFM serves as a discrete control model for human decision making. An OFM is built through subject matter expert interviews and the Applied Cognitive Task Analysis (ACTA) process. OFM's have been used in image interpretation [26] and NASA ground control system interface [27]. The experiment in this paper builds on the image classification experimentation from Zelnio and Fendley [13].

2.0 METHODS

2.1 Participants

Thirty participants (21 male, 9 female) between the ages of 18-50, with an average age of 26.4, were recruited from the college community. Twenty-five participants answered yes to having experience with EO, IR, and SAR. Five participants responded to having knowledge of all three sensor types, but did not have experience with all three. Recruiting was conducted either through face-to-face contact or through personal e-mails. Study permission was obtained from the University IRB, and no compensation was provided.

2.2 Experiment Setup

The participant sat approximately 18 inches from the Tobii eye tracking monitor for this experiment. The different images and DSS sets were displayed on the screen, and the participants selected a response using the mouse. The participant responses and eye tracking data were collected using the Tobii program.

2.3 Stimuli

The equipment used in the study consisted of a computer display showing images drawn from publically available datasets. The Moving and Stationary Target Acquisition and Recognition (MSTAR) images consist of labeled military vehicles in SAR imagery and are available online. The electro-optical and infrared data were rendered using Meta-VR. The experiment was conducted in an access-controlled laboratory.

In the experiment, there were fifteen sets of images for the participant to classify. The images were from three sensor sources: electro-optical (EO), infrared (IR) and synthetic aperture radar (SAR). There were five sets of twelve images for each sensor. The participants classified whether there was a target in the image and if so, which target type was being shown. The targets in this experiment are military vehicles: T-72, BMP-2, and BTR-70. Four of the image sets for each sensor were displayed with an accompanying DSS. The other set did not include a decision aid to help the participant with their classification. For the four sets of images with a DSS,

the participant was asked why they made that decision. They had six choices that were designed to uncover the cognitive heuristics that were being used. All sensors' images were tested in this way. All participants were asked to provide a ranking of their confidence for each decision using a 5-point Likert scale. The tests were presented in random order within each sensor.

2.4 Experimental Design

The experiment was a fixed-effect experiment and a within subject design. The experiment was divided into five alternative designs: 1) no DSS, 2) DSS consisting of a recommended classification with a corresponding percent confidence, 3) DSS consisting of a recommended classification with a corresponding percent confidence and, in addition, reference images, 4) DSS consisting of the highest and second highest recommendations and corresponding percent confidences, 5) DSS consisting of the highest and second highest recommendations and corresponding percent confidences and, also, reference images. The participants had four classification options for each test image: T-72, BMP-2, BTR-70, and No Target. After each classification, the participant was asked to state their confidence in the correctness of their decision on a five-point Likert scale. Each classification was timed. The participant was also asked why they made that decision when the DSS was present by answering the following questions:

How did you use the Decision Aid in the last decision?

1a. Confirmation of what I already thought was correct using the PERCENTAGE more
1b. Confirmation of what I already thought was correct using the IMAGES more
1c. Confirmation of what I already thought was correct using the BOTH the percentage and the images equally
2a.Used the suggestion for decision using the PERCENTAGE more
2b. Used the suggestion for decision using the IMAGES more
2c. Used the suggestion for decision using BOTH the percentage and the images equally
3. A specific part of the image influenced your decision, solely using the test image and not using the decision aid
4. None of the above

After each set of twelve images in the decision aid portion of the test was completed, a trust rating was collected.

For this study, the independent variables were the sensor types and levels of DSS. The dependent variables were time, accuracy, confidence, bias, and trust rating. The order in which the participant viewed each sensor was randomly selected. In addition, the data set order with and without the DSS was randomly selected for each participant. Within each set, the images were randomized initially, and this presentation order remained the same for each participant.

3.0 ANALYSIS

The analysis is divided into three sections corresponding to the two independent variables: the sensor level (EO, IR, SAR) and the DSS levcl (corresponding to the 5 levels described above), and the analysis of the heuristic data. For the sensor level analysis, the performance is not broken out by DSS levels; however, the DSS level analysis does break out the performance as a function of sensors and the dependent variables of accuracy,

confidence, time, and trust. The use of heuristics by the image analysts considered both sensor and DSS effects. In addition, the possible biases were also investigated.

3.1 Sensor Level

Accuracy analysis: The accuracy is analyzed using the Marascuillo procedure. Accuracy is first analyzed for each sensor. It was found that each sensor's accuracy was statistically different than each other. EO had the highest accuracy (M=88.61). IR had the second highest accuracy (M=77.99), and SAR has the lowest accuracy (73.49).

Confidence analysis: Using matched pairs, the EO Confidence (M=4.11) was greater than IR (t=11.3759, p=<.0001) and SAR Confidence (t=17.9958, p<.0001) and the IR Confidence (M=3.547) was greater than the SAR Confidence (M=2.98) (t=9.59552, p=<.0001).

Trust Analysis: The EO trust (M=4.04) was greater than the IR trust (M=3.94) (t=2.12043, p=.018) and the SAR trust (M=3.625) (t=7.04389, p=<.0001). In addition, the IR trust was greater than the SAR trust (t=5.20151, p=<.0001).

Interpretation time analysis: The EO Time (M=5.04) to make a call was less than the IR time (M=6.37) and the SAR time (M=7.828) (t=-5.241772, t=-7.4516, both with p=<.0001). The IR time was less than SAR Time (t=-4.515577, p=<.0001).

Sensor level conclusion: All paired tests were significant at the alpha=0.05 level. Based on the Marascuillo procedure, EO Accuracy > IR Accuracy > SAR Accuracy. The results are depicted in Figure 1 below demonstrating the relationship of EO (most literal) to IR (less literal) to SAR (least literal) to the quantitative results reported in this section. These results support the hypothesis that performance correlates with the literalness of the sensor.

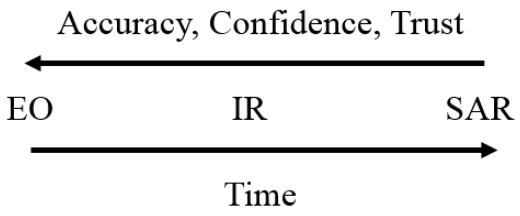


Figure 1: Spectrum of Sensor Results

3.2 DSS Level

Using the Marascuillo procedure to compare accuracy, DSS1 (M=63.70, no DSS) was clearly less accurate that the other DSS levels, and this difference was statistically significant. The only other statistically significant difference was between DSS 3 (M=80.73) and DSS 4 (M=86.54); however, this difference does not support any meaningful conclusion as it was grouped within the other DSS systems.

Accuracy analysis: For EO, DSS 4 (M=96.11) and DSS 5 (M=95.83) have the highest accuracy and are statistically different from the other DSS systems but were not different from each other. Both DSS 4 and 5

gave the top two confidence values. Also, for EO, DSS1 was not the poorest performing DSS system likely due to the literalness of EO. For IR and SAR, the only significant difference is between DSS 1 (IR DSS1 M=58.33 and SAR DSS1 M=45.55), the lowest accuracy system, and the other four DSS systems. The individual sensor analysis is consistent with the previous conclusion that the primary and only consistently statistically significant difference is between no DSS and any of the other DSS systems.

Confidence analysis: For all sensors, all decision aids were statistically higher in confidence than no decision aids (M=3.14). DSS 3 (M=3.74) and DSS 5 (M=3.72), the imagery examples, gave the highest confidence, and the difference between them was not statistically significant (t=.44762, p=.3278). Both were statistically greater than DSS 4 (M=3.47) (t=4.1511, p=<.0001; t=5.0696, p=<.0001 respectively). DSS 3 was statistically greater than DSS 2 (M=3.67) (t=1.7175, p=.0447), but DSS 5 was not statistically greater than DSS 2 (t=.9645, p=.1687).

For SAR confidence, DSS3 (M=3.31) was statistically significantly higher than the rest of the DSSs (for DSS1 t=8.422, p=<.0001; DSS2 t=2.0907, p=.0227; DSS4 t=3.2629, p=.0014; and DSS5 t=2.1593, p=.0199). The next highest were DSS 5 (3.11) and DSS 2 (M=3.16) which were not statistically different (t=-.70161, p=.2443). For IR, DSS 3 (M=3.92), DSS 2 (3.87), and DSS 5 (M=3.73) gave the highest confidence with DSS 3 and DSS 2 being statistically higher than DSS 5 (t=2.796, p=.0045 and t=2.0777, p=.0233). Also, DSS 3, DSS 2, and DSS 5 were statistically significantly higher than DSS 4 (M=3.11) (t=12.3566, p=<.0001; t=10.7366, p=<.0001; t=8.7364, p=<.0001). For EO, not all of the DSSs (DSS 2 (M=3.994) and DSS 3 (M=3.989)) were a statistical improvement over DSS 1 (M=3.97). However, DSS 4 (M=4.306) and DSS 5 (M=4.314) were statistically greater than the rest of the DSSs.

Trust analysis: Overall, the highest trust occurred with DSS 3 (M=4.02) and DSS 5 (M=3.88), which had imagery examples. DSS 3 was statistically significantly greater than all other DSSs whereas DSS5 was statistically greater than 4 (t=2.962, p=<.0001) but not DSS 2 (M=3.85) (t=.567653, p=.2858) at alpha=0.05.

For SAR, both DSS 3 (M=3.73) and DSS 5 (M=3.61) were again the highest trust, but the only significant difference was between DSS 3 and DSS 4 (M=3.54), DSS 4 < DSS 3 (t=-2.557, p=.008). Again, for IR, DSS 3 (M=4.14) and DSS 5 (M=3.97) gave the highest trust, but only the differences with DSS 3 were statistically significant (with DSS 2 t=3.13676, p=.0019; DSS 4 t=4.33296, p=<.0001; and DSS 5 t=2.40832, p=.0113). With EO, there was a similar story with DSS 3 (M=4.2) and DSS 5 (M=4.07), the ones with imagery examples giving the highest trust scores, again, differences of DSS 3 were statistically significant (with DSS 2 t=2.2685, p=.0155; DSS 4 t= 4.6697, p=<.0001; and DSS 5 t=1.8692, p=.0359). DSS 5 was statistically greater than DSS 4 (M=3.85) (t=2.641, p=.0066) at alpha=0.05.

Interpretation time analysis: Total time used with each DSS was ranked consistent with the amount of information presented. DSS 1 used the least amount of time (M=4.92), then DSS 2 (M=5.64) and DSS 4 (M=6.32) which were not statistically different but was the next in line of time, and DSS 3 (M=6.48) and DSS 5 (M=8.72) used the most amount of time. DSS 2 and DSS 4 showed no imagery whereas both DSS 3 and DSS 5 showed imagery and hence took the longest. Next, based on the eye tracking metrics, a deep dive was taken into where the time was spent when using the DSSs. The heat maps shown in Figure 2, 3, and 4 the duration of fixations and are a visual way of viewing fixations. It is noted that the heat maps depict the cumulative set of fixations of the 30 participants as they gazed individually at a single DSS test example. It is noted that when there are two suggestions, it appears most of the time is spent on the first or top suggestion.

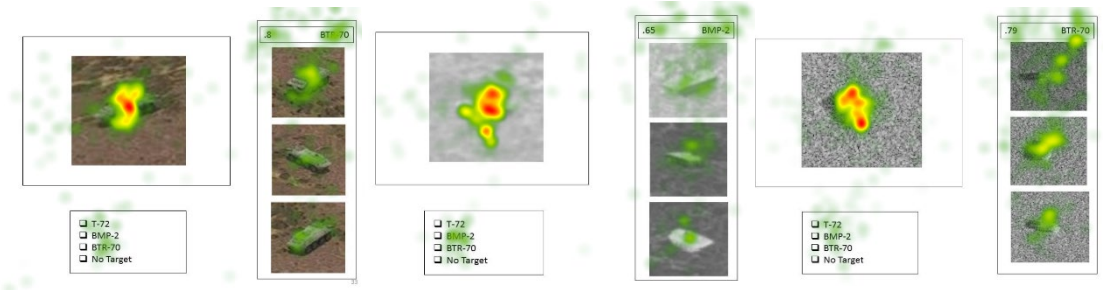


Figure 2: DSS 3 Heat Maps of Fixation Duration EO, IR, SAR

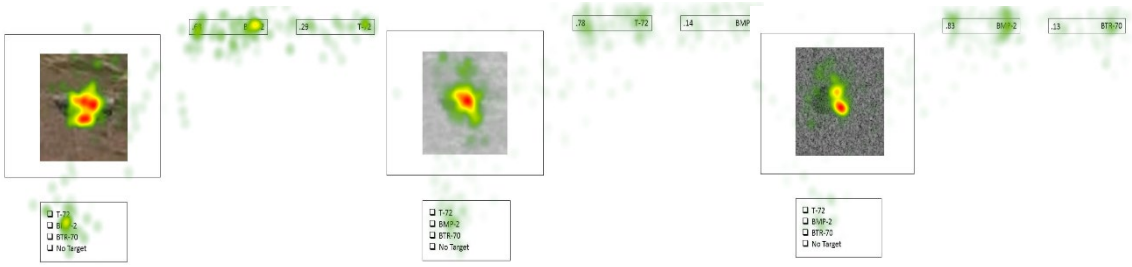


Figure 3: DSS 4 Heat Maps of Fixation Duration EO, IR, SAR

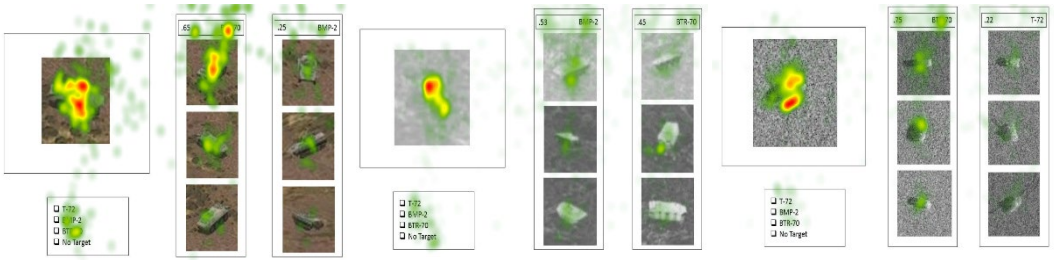


Figure 4: DSS 5 Heat Maps of Fixation Duration EO, IR, SAR

DSS level conclusion: The use of DSSs improved the confidence in the decision over no DSS provided, and these differences were statistically significant in all cases. In addition, the use of imagery in the DSS for all cases increased the trust (t=4.655146, p=<.0001) and confidence (t=4.96363, p=<.0001) over DSSs without imagery (trust with imagery M=3.95, trust with percentage M=3.78, confidence with imagery M=3.73, confidence with percentage M=3.57). It was interesting that DSS 3, the one with only one confidence value and one set of imagery examples was consistently the highest scoring DSS in both the trust and confidence category. The additional confidence and imagery information given in DSS 5 did not improve the confidence or trust overall but did slow the decision process. In fact, as the amount of information provided increased, the interpretation time increased.

3.3 Heuristic Data

We have established that cognitive heuristics are used to avoid mental workload and may have a negative impact on decision making. As discussed earlier, confirmation and anchoring were more clearly examined in

Experiment 2. For both confirmation and anchoring, there is a lack of adjustment in the decision making. For confirmation, there is no adjustment from the answer that the participant believes is true. The participant does this by only looking for confirming evidence of what they already believe is correct. When finding this confirming evidence, the affirmation can lead to higher confidence in the decision. For anchoring, there is lack of adjustment from the information given. The DSS provides a percentage and/or images that could be used by the participant. The participants could anchor on a piece or pieces of information and not adjust by using their own perception or secondary information. Overall, we are looking at the adjustment heuristic broken down into the subcategories of confirmation and anchoring. The question used to collect the heuristic and bias data with the explanation of the answers are shown in Table I.

Table 1: Heuristic/Bias Question and Explanation

Heuristic/Bias Question Used in Experiment 2	Heuristic Being Used
1a. Confirmation of what I already thought was correct using the PERCENTAGE more	Participant has made their decision and uses the percentage on the DSS to confirm their decision- Confirmation
1b. Confirmation of what I already thought was correct using the IMAGES more	Participant has made their decision and uses the images on the DSS to confirm their decision- Confirmation
1c. Confirmation of what I already thought was correct using the BOTH the percentage and the images equally	Participant has made their decision and uses the both the percentage and the images on the DSS to confirm their decision equally - Confirmation
2a. Used the suggestion for decision using the PERCENTAGE more	Participant is unsure of the answer and uses the suggestion of the DSS based off of the percentage- Anchoring
2b. Used the suggestion for decision using the IMAGES more	Participant is unsure of the answer and uses the suggestion of the DSS based off of the image- Anchoring
2c. Used the suggestion for decision using BOTH the percentage and the images equally	Participant is unsure of the answer and uses the suggestion of the DSS based off of the percentage and images equally- Anchoring
3. A specific part of the image influenced your decision, solely using the test image and not using the decision aid	Did not use DSS, no heuristic used
4. None of the above	No heuristic used

The impact of the use of cognitive heuristics was analyzed by looking at the responses by sensor type, DSS type and incorrect responses, indicating a cognitive bias.

First, the correct answer heuristic data was evaluated as a function of sensor type. The initial question addressed

was whether or not the participants used various heuristics dependent upon the sensor type. Whereas, for the least literal sensor, the participants relied on the decision aid as the basis for their decision. For EO, confirmation was used 75 percent of the time when the correct answer was selected and 62 percent of the time for IR. Anchoring was used 62 percent of the time for SAR when the correct answer was selected. This indicates that, for the more literal sensors, the participants used the DSS to confirm their answer. All the differences among the three sensor types were statistically significant, as shown in Figure 5. This follows the pattern that was set for all heuristic data for all responses.

Categorical

Responses(Heuristics of Correct Responses) By Sensor

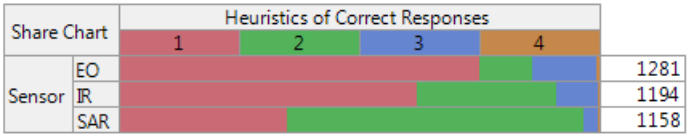
			Freq / Share	Heuristics of Correct Responses 1	2	3	4	Total Responses	Compare
Sensor	EO	A		958	140	171	12	1281	B,C
				74.8%	10.9%	13.3%	0.9%		
	IR	B		740	343	105	6	1194	C
				62.0%	28.7%	8.8%	0.5%		
	SAR	C		405	712	37	4	1158	
				35.0%	61.5%	3.2%	0.3%		

Default Comparison Groups: A/B/C

Shows letter of the category it is significantly different from at the higher share level

* Base count warning 100 Uppercase Alpha Level 0.05

** Base count minimum 30 Lowercase Alpha Level 0.1



Compare Each Sample

Letter comparisons use Pearson Chisq

Sensor, Heuristics of Correct Responses

LR PValues

LR Chi-square p-value on pairs

	A	B	C
A	1.0000	<.0001	<.0001
B	<.0001	1.0000	<.0001
C	<.0001	<.0001	1.0000

Pearson PValues

Pearson Chi-square p-value on pairs

	A	B	C
A	1.0000	<.0001	<.0001
B	<.0001	1.0000	<.0001
C	<.0001	<.0001	1.0000

Figure 5: Heuristic for Correct Responses by Sensor

The correct responses were next analyzed as a function of DSS type. The significant differences were seen between the following decision support levels: DSS 2 (confidence only) and DSS 5 (images and confidences), and DSS 4 (confidences only) and DSS 5 (images and confidences). This does follow a similar pattern with the statistical differences were between a DSS that used confidence(s) and a DSS that had both images and confidences. These differences are shown in Figure 6. Again, the differences as a function of DSS were subtle

as compared to the sensor differences. It would be hard to conclude that the DSS types had a significant effect on the heuristics that the analyst used.

Categorical

Responses(Heuristics of Correct Responses) By DSS

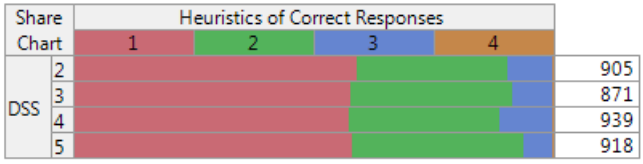
Freq / Share			Heuristics of Correct Responses 1	2	3	4	Total Responses	Compare
DSS	2	A	533	284	82	6	905	
			58.9%	31.4%	9.1%	0.7%		
	3	B	502	291	73	5	871	
			57.6%	33.4%	8.4%	0.6%		
	4	C	536	295	103	5	939	D
			57.1%	31.4%	11.0%	0.5%		
	5	D	532	325	55	6	918	A
			58.0%	35.4%	6.0%	0.7%		

Default Comparison Groups: A/B/C/D

Shows letter of the category it is significantly different from at the higher share level

* Base count warning 100 Uppercase Alpha Level 0.05

** Base count minimum 30 Lowercase Alpha Level 0.1



Compare Each Sample

Letter comparisons use Pearson Chisq LR PValues

DSS, Heuristics of Correct Responses

LR Chi-square p-value on pairs

	A	B	C	D
A	1.0000	0.8069	0.5579	0.0455
B	0.8069	1.0000	0.2928	0.2463
C	0.5579	0.2928	1.0000	0.0011
D	0.0455	0.2463	0.0011	1.0000

Pearson PValues

Pearson Chi-square p-value on pairs

	A	B	C	D
A	1.0000	0.8069	0.5588	0.0462
B	0.8069	1.0000	0.2951	0.2470
C	0.5588	0.2951	1.0000	0.0012
D	0.0462	0.2470	0.0012	1.0000

Figure 6: Heuristics for Correct Responses by DSS

Now that the use of the heuristics is established, the possible biases were analyzed by partitioning the data to contain only incorrect responses. As can be seen in Figure 7, the potential impact of the anchoring bias was demonstrated the most often when they answered incorrectly with all three sensors. The potential biases used were as follows: EO used anchoring 54 percent of the time while IR used anchoring 47 percent of the time. These differences were not found to be statistically significant; however, SAR used anchoring 63 percent of the time which was statistically different from both EO and IR.

Categorical

Responses(Bias) By Sensor

			Freq / Share	Bias 1	Bias 2	Bias 3	Bias 4	Total Responses	Compare
Sensor	EO	A	54	85	14	4	157	
			34.4%	54.1%	8.9%	2.5%		
	IR	B	89	112	32	8	241	
			36.9%	46.5%	13.3%	3.3%		
	SAR	C	56	176	40	7	279	A,B
			20.1%	63.1%	14.3%	2.5%		

Default Comparison Groups: A/B/C

Shows letter of the category it is significantly different from at the higher share level

* Base count warning 100 Uppercase Alpha Level 0.05

** Base count minimum 30 Lowercase Alpha Level 0.1

Share Chart		Bias 1	2	3	4	
Sensor	EO					157
	IR					241
	SAR					279

Compare Each Sample

Letter comparisons use Pearson Chisq LR PValues Pearson PValues

Sensor, Bias

LR Chi-square p-value on pairs

	A	B	C
A	1.0000	0.3768	0.0084
B	0.3768	1.0000	0.0002
C	0.0084	0.0002	1.0000

Pearson Chi-square p-value on pairs

	A	B	C
A	1.0000	0.3838	0.0078
B	0.3838	1.0000	0.0002
C	0.0078	0.0002	1.0000

Figure 7: Potential Bias by Sensor

When looking at the potential biases in terms of the DSSs, all DSS types had anchoring as the heuristic used most often. Secondarily, the potential biases were distributed differently among the four choices and these differences were statistically significant except for the difference between DSS 3 and DSS 5. These two DSS types both provided image examples as part of their decision systems which may explain their similar results. The statistical analysis is shown in Figure 8.

Categorical

Responses(Bias) By DSS

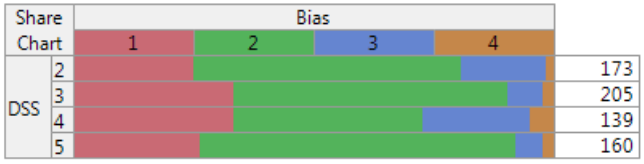
Freq / Share			Bias 1	Bias 2	Bias 3	Bias 4	Total Responses	Compare
DSS	2	A	43 24.9%	96 55.5%	31 17.9%	3 1.7%	173	C,D
	3	B	68 33.2%	117 57.1%	15 7.3%	5 2.4%	205	A,C
	4	C	46 33.1%	55 39.6%	31 22.3%	7 5.0%	139	
	5	D	42 26.3%	105 65.6%	9 5.6%	4 2.5%	160	C

Default Comparison Groups: A/B/C/D

Shows letter of the category it is significantly different from at the higher share level

* Base count warning 100 Uppercase Alpha Level 0.05

** Base count minimum 30 Lowercase Alpha Level 0.1



Compare Each Sample

Letter comparisons use Pearson Chisq LR PValues

DSS, Bias

LR Chi-square p-value on pairs

	A	B	C	D
A	1.0000	0.0105	0.0255	0.0050
B	0.0105	1.0000	0.0001	0.4044
C	0.0255	0.0001	1.0000	<.0001
D	0.0050	0.4044	<.0001	1.0000

Pearson PValues

Pearson Chi-square p-value on pairs

	A	B	C	D
A	1.0000	0.0110	0.0263	0.0068
B	0.0110	1.0000	0.0001	0.4071
C	0.0263	0.0001	1.0000	<.0001
D	0.0068	0.4071	<.0001	1.0000

Figure 8: Potential Bias by DSS

4.0 DISCUSSION

For classification accuracy, all the DSSs that were designed and tested outperformed the no DSS conditions. However, the interpretation time increased with additional information in the DSS. This finding suggests that for time critical tasks, careful attention must be paid to the amount of information in the decision aid as a more accurate decision that is not made within the necessary timeline would not meet mission objectives. From a sensor perspective, the classification accuracy of EO was greater than IR which, in turn, was greater than SAR which was in order of the literalness of the sensor. As may be anticipated, the timing numbers were reversed with SAR taking longer than IR which took longer than EO.

Both trust and confidence were enhanced with the DSSs; however, these subjective measures and eye tracking data showed that the reference imagery was the major contributor to the participant's trust and confidence, not the DSS algorithm probability. The study dug deeper in attempting to ascertain the dependency of the interpretation task on sensor type. To accomplish this goal, the heuristics that the participants used to make their

decisions were elicited. It was found that for EO and IR, confirmation (decision primarily driven by test imagery) was the heuristic that was used most often, and for SAR, anchoring (decision primarily driven by DSS) was used most often. Again, this is understandable based on the literalness of each sensor. Also, this finding was also supported by the eye tracking data as the fixation time on the reference imagery increased with the non-literalness of the sensor (i.e., SAR time > IR time > EO time). On the other hand, when considering possible biases (i.e., errors resulting from heuristic use), anchoring was the culprit for all sensor types.

5.0 FUTURE WORK

This experiment can be further analyzed by looking into the fixation count, visit count, and transitions of areas of interest within the DSS. The eye tracking data can be used to see if there are any other lessons learned, specifically, exploring the effect of the secondary suggestion. Did the secondary suggestion have a positive impact or was it too much information? Eye tracking will also be able to give a firm answer to whether the secondary suggestion was considered by determining if the participants looked at this area of the DSS.

REFERENCES

[1] Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5), 527-539.

[2] Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics*, 2, 1926-1943.

[3] Helldin, T., Falkman, G., Riveiro, M., Dahlbom, A., & Lebram, M. (2013). Transparency of military threat evaluation through visualizing uncertainty and system rationale. *Engineering Psychology and Cognitive Ergonomics*. Applications and Services, 8020, 263-272.

[4] Liebhaber, M. J., & Feher, B. (2002). Air threat assessment: Research, model, and display guidelines. SPACE AND NAVAL WARFARE SYSTEMS COMMAND SAN DIEGO CA.

[5] Hoffman, R. R., Johnson, M., Bradshaw, J. M., & Underbrink, A. (2013). Trust in automation. *Intelligent Systems*, IEEE, 28(1), 84-88.

[6] Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing Decisions Improved Decision Making With a Single Training Intervention. *Policy Insights from the Behavioral and Brain Sciences*, 2372732215600886.

[7] Fischoff, B. (1981). *Debiasing* (No. PTR-1092-81-3). DECISION RESEARCH EUGENE OR.

[8] Keren, G.; Caverni, J.P.; Fabre, J.M.; et al. *Cognitive Biases.* Amsterdam: North-Holland 1990.

[9] Soll, J. B., Milkman, K. L., & Payne, J. W. (2014). A user's guide to debiasing.

[10] Larrick R. P. (2004). “Debiasing,” in *Blackwell Handbook of Judgement and Decision Making* eds Koehler D. J., Harvey N., editors. (Malden, MA: Blackwell Publishing Ltd.) 316–337.

[11] Arnott, D. (2006). Cognitive biases and decision support systems development: a design science approach.

Information Systems Journal, *16*(1), 55-78.

[12] Fendley, M., & Narayanan, S. (2012). Decision aiding to overcome biases in object identification. *Advances in Human-Computer Interaction*, *2012*, 7.

[13] Zelnio, H., & Fendley, M. (2018). Human performance modelling for image analyst decision support design. *International Journal of Human Factors Modelling and Simulation*, *6*(2-3), 184-202.

[14] Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, *66*(1), 102-110.

[15] Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory.* Psychology Press.

[16] Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

[17] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

[18] Wall, E., Blaha, L. M., Paul, C. L., Cook, K., & Endert, A. (2017). Four perspectives on human bias in visual analytics. In *DECISIVe: Workshop on Dealing with Cognitive Biases in Visualizations*.

[19] George, J. F., Duffy, K., & Ahuja, M. (2000). Countering the anchoring and adjustment bias with decision support systems. *Decision Support Systems*,*29*(2), 195-206

[20] Bhandari, G., Hassanein, K., & Deaves, R. (2008). Debiasing investors with decision support systems: An experimental investigation. *Decision Support Systems*, *46*(1), 399-410.

[21] Zenker, F., Dahlman, C., & Sarwar, F. (2015). Reliable Debiasing Techniques in Legal Contexts? Weak Signals from a Darker Corner of the Social Science Universe. *The psychology of argument: cognitive approaches to argumentation and persuasion*.

[22] Zachary, W. W. (1980). *Decision aids for naval air ASW* (No. 1366-A). ANALYTICS INC WILLOW GROVE PA.

[23] Woods, D. D., & Roth, E. M. (1988). Cognitive Systems Engineering. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 3-43): Elsevier Science.

[24] Hopson, J., Zachary, W., & Lane, N. (1981). The intelligent use of intelligent systems: Problems in engineering man/machine symbiosis. *AGARD The Impact of New Guidance and Control Systems on Mil. Aircraft Cockpit Design 16 p(SEE N 82-13048 04-01)*.

[25] Mitchell, C.M. & Miller, R.A. (1986) A discrete control model of operator function: A methodology for information display design, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 16 No.3, pp.343-357.

[26] Muller, Amanda C., & S. Narayanan. Cognitively-engineered multisensor image fusion for military

applications. *Information Fusion* 10.2 (2009): 137-149.

[27] Mitchell, Christine M. "GT-MSOCC: A domain for research on human-computer interaction and decision aiding in supervisory control systems." Systems, Man and Cybernetics, IEEE Transactions on 17.4 (1987): 553-572.

Automated Aiding of Image Analysts as a Function of Sensor Type

Dr. Holly Zelnio and Dr. Mary Fendley
711th Human Performance Wing
Wright State University
UNITED STATES

holly.zelnio@us.af.mil

ABSTRACT

The use of an automated Decision Support System (DSS) to aid an image interpreter performing a target classification task is analyzed to ascertain performance improvement due to human automation cooperation. The performance of the image interpreter and DSS is analyzed from a human factors perspective. This research performs a human subject experiment measuring classification performance as a function of classification accuracy, interpretation time, confidence in the decision, and trust in the decision support system (DSS). In addition, multiple types of automated support were compared as were multiple types of imaging sensors. The imaging sensors that were used in this study were electro-optical (EO), infrared (IR), and synthetic aperture radar (SAR). The DSSs were developed using cognitive task analysis and previous experimentation. The results show that different types of sensors require different DSSs for optimal performance. Results suggest that EO requires less automated support while the more non-literal sensors (e.g., IR and SAR) need more automated support for highest performance. This research also investigated how a DSS itself can cause a bias influencing performance. The performance of the human and machine team can be enhanced by the presentation of key information. It was shown that the level of automated analyst aiding affects both the quality and timeliness of the decision as well as the trust in the DSS.

1.0 BACKGROUND

Trust has to be balanced for optimum performance. Either too much or too little trust can result in failures [1]. To maintain trust in automation, Sarter suggests that the human should understand the basis of the automated aids but should have the task responsibility. The human can use the automation for help, but the automation is now a means to the end of a human responsible task [2]. Increased transparency has led to increased trust, but higher decision times and workloads [3]. Increased trust and performance is suggested by findings if contextual information is provided by the automated DSS [4]. Evidence indicates that, over time, experienced workers can learn to calibrate their trust for the automated system used in their job [5].

Biases can appear when making judgments. It is difficult to fully eliminate a bias from a decision. However, two approaches to help mitigate bias includes: 1) additional training [6] or knowledge needs to be given to the individual or 2) the task needs to be structured to not create these biases. The task structuring can be designed to coincide with an individual's natural thought process to help inhibit bias [7]. This approach can be accomplished by examining where the bias is coming into play either via the person or the task/environment. Several different approaches have been studied to suppress bias including using structured processes [7], applying a three stage framework [8], improving the environment [9], and adopting the specific debiasing strategies of technical, motivational, and cognitive [10].

Cognitive bias happens when one deviates from good judgement and can occur or increase when introducing information to a decision maker. Arnott [11] has an extensive list of biases that influence people. Fendley [12] shortened this list to specifically biases that are found in object identification. The dominant biases that were identified by Zelnio and Fendley [13] when analysts perform object identification were confirmation and anchoring. Confirmation bias is present when the data and interpretation favors a single decision over alternatives [14-16]. Anchoring bias is being reliant on an initial value or given information [17, 18]. Overall, it has been suggested that a DSS should mitigate bias [19, 20]. The number of studies performed confirming bias exists heavily outweighs the number of studies that try to mitigate bias [21]. It is understood that bias cannot be eliminated. It is important to acknowledge that it exists and then try to design the support system to use the biases to our advantage or to try to make them negligible.

DSSs can be built in various ways including: a framework for the characterization of the decision situation [22], three factor function for designers to structure representation [23], the summary table of aiding requirements (STAR) table [24], and the operator function model (OFM) [25]. An OFM serves as a discrete control model for human decision making. An OFM is built through subject matter expert interviews and the Applied Cognitive Task Analysis (ACTA) process. OFM's have been used in image interpretation [26] and NASA ground control system interface [27]. The experiment in this paper builds on the image classification experimentation from Zelnio and Fendley [13].

2.0 METHODS

2.1 Participants

Thirty participants (21 male, 9 female) between the ages of 18-50, with an average age of 26.4, were recruited from the college community. Twenty-five participants answered yes to having experience with EO, IR, and SAR. Five participants responded to having knowledge of all three sensor types, but did not have experience with all three. Recruiting was conducted either through face-to-face contact or through personal e-mails. Study permission was obtained from the University IRB, and no compensation was provided.

2.2 Experiment Setup

The participant sat approximately 18 inches from the Tobii eye tracking monitor for this experiment. The different images and DSS sets were displayed on the screen, and the participants selected a response using the mouse. The participant responses and eye tracking data were collected using the Tobii program.

2.3 Stimuli

The equipment used in the study consisted of a computer display showing images drawn from publically available datasets. The Moving and Stationary Target Acquisition and Recognition (MSTAR) images consist of labeled military vehicles in SAR imagery and are available online. The electro-optical and infrared data were rendered using Meta-VR. The experiment was conducted in an access-controlled laboratory.

In the experiment, there were fifteen sets of images for the participant to classify. The images were from three sensor sources: electro-optical (EO), infrared (IR) and synthetic aperture radar (SAR). There were five sets of twelve images for each sensor. The participants classified whether there was a target in the image and if so, which target type was being shown. The targets in this experiment are military vehicles: T-72, BMP-2, and BTR-70. Four of the image sets for each sensor were displayed with an accompanying DSS. The other set did not include a decision aid to help the participant with their classification. For the four sets of images with a DSS,

the participant was asked why they made that decision. They had six choices that were designed to uncover the cognitive heuristics that were being used. All sensors' images were tested in this way. All participants were asked to provide a ranking of their confidence for each decision using a 5-point Likert scale. The tests were presented in random order within each sensor.

2.4 Experimental Design

The experiment was a fixed-effect experiment and a within subject design. The experiment was divided into five alternative designs: 1) no DSS, 2) DSS consisting of a recommended classification with a corresponding percent confidence, 3) DSS consisting of a recommended classification with a corresponding percent confidence and, in addition, reference images, 4) DSS consisting of the highest and second highest recommendations and corresponding percent confidences, 5) DSS consisting of the highest and second highest recommendations and corresponding percent confidences and, also, reference images. The participants had four classification options for each test image: T-72, BMP-2, BTR-70, and No Target. After each classification, the participant was asked to state their confidence in the correctness of their decision on a five-point Likert scale. Each classification was timed. The participant was also asked why they made that decision when the DSS was present by answering the following questions:

How did you use the Decision Aid in the last decision?

1a. Confirmation of what I already thought was correct using the PERCENTAGE more
1b. Confirmation of what I already thought was correct using the IMAGES more
1c. Confirmation of what I already thought was correct using the BOTH the percentage and the images equally
2a.Used the suggestion for decision using the PERCENTAGE more
2b. Used the suggestion for decision using the IMAGES more
2c. Used the suggestion for decision using BOTH the percentage and the images equally
3. A specific part of the image influenced your decision, solely using the test image and not using the decision aid
4. None of the above

After each set of twelve images in the decision aid portion of the test was completed, a trust rating was collected.

For this study, the independent variables were the sensor types and levels of DSS. The dependent variables were time, accuracy, confidence, bias, and trust rating. The order in which the participant viewed each sensor was randomly selected. In addition, the data set order with and without the DSS was randomly selected for each participant. Within each set, the images were randomized initially, and this presentation order remained the same for each participant.

3.0 ANALYSIS

The analysis is divided into three sections corresponding to the two independent variables: the sensor level (EO, IR, SAR) and the DSS levcl (corresponding to the 5 levels described above), and the analysis of the heuristic data. For the sensor level analysis, the performance is not broken out by DSS levels; however, the DSS level analysis does break out the performance as a function of sensors and the dependent variables of accuracy,

confidence, time, and trust. The use of heuristics by the image analysts considered both sensor and DSS effects. In addition, the possible biases were also investigated.

3.1 Sensor Level

Accuracy analysis: The accuracy is analyzed using the Marascuillo procedure. Accuracy is first analyzed for each sensor. It was found that each sensor's accuracy was statistically different than each other. EO had the highest accuracy (M=88.61). IR had the second highest accuracy (M=77.99), and SAR has the lowest accuracy (73.49).

Confidence analysis: Using matched pairs, the EO Confidence (M=4.11) was greater than IR (t=11.3759, p=<.0001) and SAR Confidence (t=17.9958, p<.0001) and the IR Confidence (M=3.547) was greater than the SAR Confidence (M=2.98) (t=9.59552, p=<.0001).

Trust Analysis: The EO trust (M=4.04) was greater than the IR trust (M=3.94) (t=2.12043, p=.018) and the SAR trust (M=3.625) (t=7.04389, p=<.0001). In addition, the IR trust was greater than the SAR trust (t=5.20151, p=<.0001).

Interpretation time analysis: The EO Time (M=5.04) to make a call was less than the IR time (M=6.37) and the SAR time (M=7.828) (t=-5.241772, t=-7.4516, both with p=<.0001). The IR time was less than SAR Time (t=-4.515577, p=<.0001).

Sensor level conclusion: All paired tests were significant at the alpha=0.05 level. Based on the Marascuillo procedure, EO Accuracy > IR Accuracy > SAR Accuracy. The results are depicted in Figure 1 below demonstrating the relationship of EO (most literal) to IR (less literal) to SAR (least literal) to the quantitative results reported in this section. These results support the hypothesis that performance correlates with the literalness of the sensor.

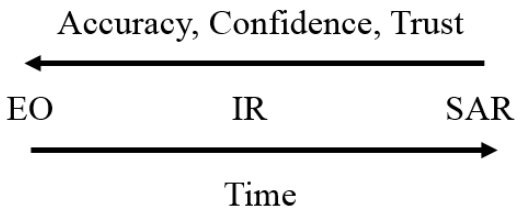


Figure 1: Spectrum of Sensor Results

3.2 DSS Level

Using the Marascuillo procedure to compare accuracy, DSS1 (M=63.70, no DSS) was clearly less accurate that the other DSS levels, and this difference was statistically significant. The only other statistically significant difference was between DSS 3 (M=80.73) and DSS 4 (M=86.54); however, this difference does not support any meaningful conclusion as it was grouped within the other DSS systems.

Accuracy analysis: For EO, DSS 4 (M=96.11) and DSS 5 (M=95.83) have the highest accuracy and are statistically different from the other DSS systems but were not different from each other. Both DSS 4 and 5

gave the top two confidence values. Also, for EO, DSS1 was not the poorest performing DSS system likely due to the literalness of EO. For IR and SAR, the only significant difference is between DSS 1 (IR DSS1 M=58.33 and SAR DSS1 M=45.55), the lowest accuracy system, and the other four DSS systems. The individual sensor analysis is consistent with the previous conclusion that the primary and only consistently statistically significant difference is between no DSS and any of the other DSS systems.

Confidence analysis: For all sensors, all decision aids were statistically higher in confidence than no decision aids (M=3.14). DSS 3 (M=3.74) and DSS 5 (M=3.72), the imagery examples, gave the highest confidence, and the difference between them was not statistically significant (t=.44762, p=.3278). Both were statistically greater than DSS 4 (M=3.47) (t=4.1511, p=<.0001; t=5.0696, p=<.0001 respectively). DSS 3 was statistically greater than DSS 2 (M=3.67) (t=1.7175, p=.0447), but DSS 5 was not statistically greater than DSS 2 (t=.9645, p=.1687).

For SAR confidence, DSS3 (M=3.31) was statistically significantly higher than the rest of the DSSs (for DSS1 t=8.422, p=<.0001; DSS2 t=2.0907, p=.0227; DSS4 t=3.2629, p=.0014; and DSS5 t=2.1593, p=.0199). The next highest were DSS 5 (3.11) and DSS 2 (M=3.16) which were not statistically different (t=-.70161, p=.2443). For IR, DSS 3 (M=3.92), DSS 2 (3.87), and DSS 5 (M=3.73) gave the highest confidence with DSS 3 and DSS 2 being statistically higher than DSS 5 (t=2.796, p=.0045 and t=2.0777, p=.0233). Also, DSS 3, DSS 2, and DSS 5 were statistically significantly higher than DSS 4 (M=3.11) (t=12.3566, p=<.0001; t=10.7366, p=<.0001; t=8.7364, p=<.0001). For EO, not all of the DSSs (DSS 2 (M=3.994) and DSS 3 (M=3.989)) were a statistical improvement over DSS 1 (M=3.97). However, DSS 4 (M=4.306) and DSS 5 (M=4.314) were statistically greater than the rest of the DSSs.

Trust analysis: Overall, the highest trust occurred with DSS 3 (M=4.02) and DSS 5 (M=3.88), which had imagery examples. DSS 3 was statistically significantly greater than all other DSSs whereas DSS5 was statistically greater than 4 (t=2.962, p=<.0001) but not DSS 2 (M=3.85) (t=.567653, p=.2858) at alpha=0.05.

For SAR, both DSS 3 (M=3.73) and DSS 5 (M=3.61) were again the highest trust, but the only significant difference was between DSS 3 and DSS 4 (M=3.54), DSS 4 < DSS 3 (t=-2.557, p=.008). Again, for IR, DSS 3 (M=4.14) and DSS 5 (M=3.97) gave the highest trust, but only the differences with DSS 3 were statistically significant (with DSS 2 t=3.13676, p=.0019; DSS 4 t=4.33296, p=<.0001; and DSS 5 t=2.40832, p=.0113). With EO, there was a similar story with DSS 3 (M=4.2) and DSS 5 (M=4.07), the ones with imagery examples giving the highest trust scores, again, differences of DSS 3 were statistically significant (with DSS 2 t=2.2685, p=.0155; DSS 4 t= 4.6697, p=<.0001; and DSS 5 t=1.8692, p=.0359). DSS 5 was statistically greater than DSS 4 (M=3.85) (t=2.641, p=.0066) at alpha=0.05.

Interpretation time analysis: Total time used with each DSS was ranked consistent with the amount of information presented. DSS 1 used the least amount of time (M=4.92), then DSS 2 (M=5.64) and DSS 4 (M=6.32) which were not statistically different but was the next in line of time, and DSS 3 (M=6.48) and DSS 5 (M=8.72) used the most amount of time. DSS 2 and DSS 4 showed no imagery whereas both DSS 3 and DSS 5 showed imagery and hence took the longest. Next, based on the eye tracking metrics, a deep dive was taken into where the time was spent when using the DSSs. The heat maps shown in Figure 2, 3, and 4 the duration of fixations and are a visual way of viewing fixations. It is noted that the heat maps depict the cumulative set of fixations of the 30 participants as they gazed individually at a single DSS test example. It is noted that when there are two suggestions, it appears most of the time is spent on the first or top suggestion.

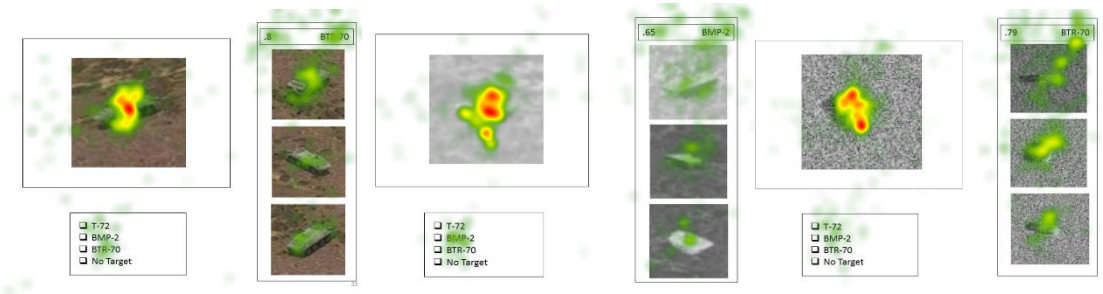


Figure 2: DSS 3 Heat Maps of Fixation Duration EO, IR, SAR

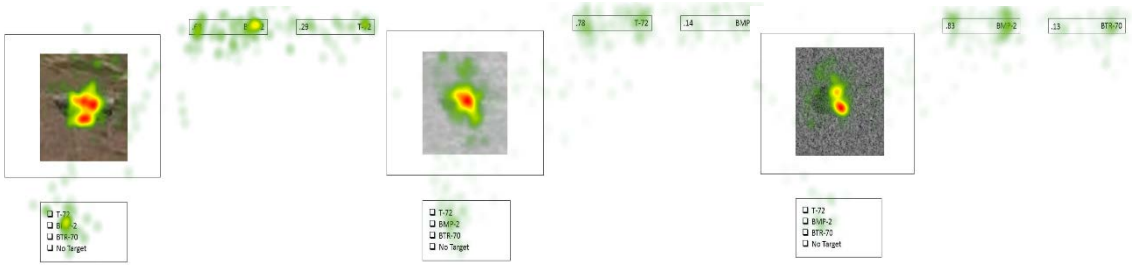


Figure 3: DSS 4 Heat Maps of Fixation Duration EO, IR, SAR

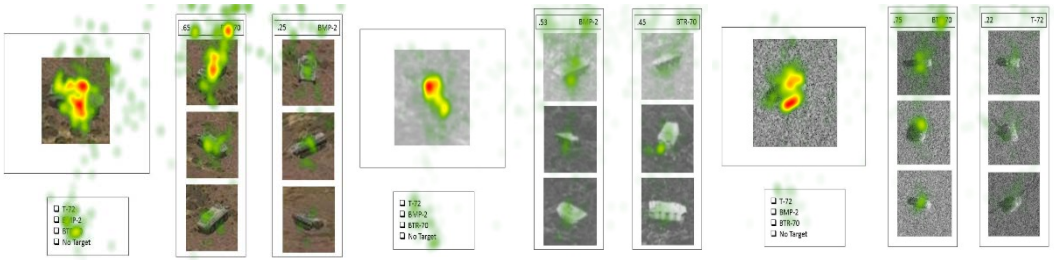


Figure 4: DSS 5 Heat Maps of Fixation Duration EO, IR, SAR

DSS level conclusion: The use of DSSs improved the confidence in the decision over no DSS provided, and these differences were statistically significant in all cases. In addition, the use of imagery in the DSS for all cases increased the trust (t=4.655146, p=<.0001) and confidence (t=4.96363, p=<.0001) over DSSs without imagery (trust with imagery M=3.95, trust with percentage M=3.78, confidence with imagery M=3.73, confidence with percentage M=3.57). It was interesting that DSS 3, the one with only one confidence value and one set of imagery examples was consistently the highest scoring DSS in both the trust and confidence category. The additional confidence and imagery information given in DSS 5 did not improve the confidence or trust overall but did slow the decision process. In fact, as the amount of information provided increased, the interpretation time increased.

3.3 Heuristic Data

We have established that cognitive heuristics are used to avoid mental workload and may have a negative impact on decision making. As discussed earlier, confirmation and anchoring were more clearly examined in

Experiment 2. For both confirmation and anchoring, there is a lack of adjustment in the decision making. For confirmation, there is no adjustment from the answer that the participant believes is true. The participant does this by only looking for confirming evidence of what they already believe is correct. When finding this confirming evidence, the affirmation can lead to higher confidence in the decision. For anchoring, there is lack of adjustment from the information given. The DSS provides a percentage and/or images that could be used by the participant. The participants could anchor on a piece or pieces of information and not adjust by using their own perception or secondary information. Overall, we are looking at the adjustment heuristic broken down into the subcategories of confirmation and anchoring. The question used to collect the heuristic and bias data with the explanation of the answers are shown in Table I.

Table 1: Heuristic/Bias Question and Explanation

Heuristic/Bias Question Used in Experiment 2	Heuristic Being Used
1a. Confirmation of what I already thought was correct using the PERCENTAGE more	Participant has made their decision and uses the percentage on the DSS to confirm their decision- Confirmation
1b. Confirmation of what I already thought was correct using the IMAGES more	Participant has made their decision and uses the images on the DSS to confirm their decision- Confirmation
1c. Confirmation of what I already thought was correct using the BOTH the percentage and the images equally	Participant has made their decision and uses the both the percentage and the images on the DSS to confirm their decision equally - Confirmation
2a. Used the suggestion for decision using the PERCENTAGE more	Participant is unsure of the answer and uses the suggestion of the DSS based off of the percentage- Anchoring
2b. Used the suggestion for decision using the IMAGES more	Participant is unsure of the answer and uses the suggestion of the DSS based off of the image- Anchoring
2c. Used the suggestion for decision using BOTH the percentage and the images equally	Participant is unsure of the answer and uses the suggestion of the DSS based off of the percentage and images equally- Anchoring
3. A specific part of the image influenced your decision, solely using the test image and not using the decision aid	Did not use DSS, no heuristic used
4. None of the above	No heuristic used

The impact of the use of cognitive heuristics was analyzed by looking at the responses by sensor type, DSS type and incorrect responses, indicating a cognitive bias.

First, the correct answer heuristic data was evaluated as a function of sensor type. The initial question addressed

was whether or not the participants used various heuristics dependent upon the sensor type. Whereas, for the least literal sensor, the participants relied on the decision aid as the basis for their decision. For EO, confirmation was used 75 percent of the time when the correct answer was selected and 62 percent of the time for IR. Anchoring was used 62 percent of the time for SAR when the correct answer was selected. This indicates that, for the more literal sensors, the participants used the DSS to confirm their answer. All the differences among the three sensor types were statistically significant, as shown in Figure 5. This follows the pattern that was set for all heuristic data for all responses.

Categorical

Responses(Heuristics of Correct Responses) By Sensor

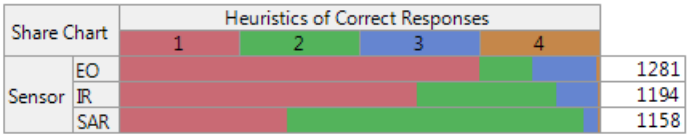
			Freq / Share	Heuristics of Correct Responses 1	2	3	4	Total Responses	Compare
Sensor	EO	A	958	140	171	12	1281	B,C
			74.8%	10.9%	13.3%	0.9%		
	IR	B	740	343	105	6	1194	C
			62.0%	28.7%	8.8%	0.5%		
	SAR	C	405	712	37	4	1158	
			35.0%	61.5%	3.2%	0.3%		

Default Comparison Groups: A/B/C

Shows letter of the category it is significantly different from at the higher share level

* Base count warning 100 Uppercase Alpha Level 0.05

** Base count minimum 30 Lowercase Alpha Level 0.1



Compare Each Sample

Letter comparisons use Pearson Chisq

Sensor, Heuristics of Correct Responses

LR PValues

LR Chi-square p-value on pairs

	A	B	C
A	1.0000	<.0001	<.0001
B	<.0001	1.0000	<.0001
C	<.0001	<.0001	1.0000

Pearson PValues

Pearson Chi-square p-value on pairs

	A	B	C
A	1.0000	<.0001	<.0001
B	<.0001	1.0000	<.0001
C	<.0001	<.0001	1.0000

Figure 5: Heuristic for Correct Responses by Sensor

The correct responses were next analyzed as a function of DSS type. The significant differences were seen between the following decision support levels: DSS 2 (confidence only) and DSS 5 (images and confidences), and DSS 4 (confidences only) and DSS 5 (images and confidences). This does follow a similar pattern with the statistical differences were between a DSS that used confidence(s) and a DSS that had both images and confidences. These differences are shown in Figure 6. Again, the differences as a function of DSS were subtle

as compared to the sensor differences. It would be hard to conclude that the DSS types had a significant effect on the heuristics that the analyst used.

Categorical

Responses(Heuristics of Correct Responses) By DSS

	Freq / Share		Heuristics of Correct Responses 1	2	3	4	Total Responses	Compare
DSS	2	A	533	284	82	6	905	
			58.9%	31.4%	9.1%	0.7%		
	3	B	502	291	73	5	871	
			57.6%	33.4%	8.4%	0.6%		
	4	C	536	295	103	5	939	D
			57.1%	31.4%	11.0%	0.5%		
	5	D	532	325	55	6	918	A
			58.0%	35.4%	6.0%	0.7%		

Default Comparison Groups: A/B/C/D

Shows letter of the category it is significantly different from at the higher share level

* Base count warning 100 Uppercase Alpha Level 0.05

** Base count minimum 30 Lowercase Alpha Level 0.1

Share Chart		Heuristics of Correct Responses 1	2	3	4	
DSS	2					905
	3					871
	4					939
	5					918

Compare Each Sample

Letter comparisons use Pearson Chisq LR PValues Pearson PValues

DSS, Heuristics of Correct Responses

LR Chi-square p-value on pairs

	A	B	C	D
A	1.0000	0.8069	0.5579	0.0455
B	0.8069	1.0000	0.2928	0.2463
C	0.5579	0.2928	1.0000	0.0011
D	0.0455	0.2463	0.0011	1.0000

Pearson Chi-square p-value on pairs

	A	B	C	D
A	1.0000	0.8069	0.5588	0.0462
B	0.8069	1.0000	0.2951	0.2470
C	0.5588	0.2951	1.0000	0.0012
D	0.0462	0.2470	0.0012	1.0000

Figure 6: Heuristics for Correct Responses by DSS

Now that the use of the heuristics is established, the possible biases were analyzed by partitioning the data to contain only incorrect responses. As can be seen in Figure 7, the potential impact of the anchoring bias was demonstrated the most often when they answered incorrectly with all three sensors. The potential biases used were as follows: EO used anchoring 54 percent of the time while IR used anchoring 47 percent of the time. These differences were not found to be statistically significant; however, SAR used anchoring 63 percent of the time which was statistically different from both EO and IR.

Categorical

Responses(Bias) By Sensor

			Freq / Share	Bias 1	Bias 2	Bias 3	Bias 4	Total Responses	Compare
Sensor	EO	A	54	85	14	4	157	
			34.4%	54.1%	8.9%	2.5%		
	IR	B	89	112	32	8	241	
			36.9%	46.5%	13.3%	3.3%		
	SAR	C	56	176	40	7	279	A,B
			20.1%	63.1%	14.3%	2.5%		

Default Comparison Groups: A/B/C

Shows letter of the category it is significantly different from at the higher share level

* Base count warning 100 Uppercase Alpha Level 0.05

** Base count minimum 30 Lowercase Alpha Level 0.1

Share Chart		Bias 1	2	3	4	
Sensor	EO					157
	IR					241
	SAR					279

Compare Each Sample

Letter comparisons use Pearson Chisq LR PValues Pearson PValues

Sensor, Bias

LR Chi-square p-value on pairs

	A	B	C
A	1.0000	0.3768	0.0084
B	0.3768	1.0000	0.0002
C	0.0084	0.0002	1.0000

Pearson Chi-square p-value on pairs

	A	B	C
A	1.0000	0.3838	0.0078
B	0.3838	1.0000	0.0002
C	0.0078	0.0002	1.0000

Figure 7: Potential Bias by Sensor

When looking at the potential biases in terms of the DSSs, all DSS types had anchoring as the heuristic used most often. Secondarily, the potential biases were distributed differently among the four choices and these differences were statistically significant except for the difference between DSS 3 and DSS 5. These two DSS types both provided image examples as part of their decision systems which may explain their similar results. The statistical analysis is shown in Figure 8.

Categorical

Responses(Bias) By DSS

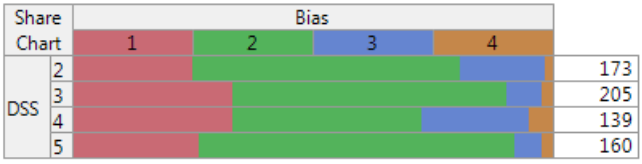
Freq / Share			Bias 1	Bias 2	Bias 3	Bias 4	Total Responses	Compare
DSS	2	A	43 24.9%	96 55.5%	31 17.9%	3 1.7%	173	C,D
	3	B	68 33.2%	117 57.1%	15 7.3%	5 2.4%	205	A,C
	4	C	46 33.1%	55 39.6%	31 22.3%	7 5.0%	139	
	5	D	42 26.3%	105 65.6%	9 5.6%	4 2.5%	160	C

Default Comparison Groups: A/B/C/D

Shows letter of the category it is significantly different from at the higher share level

* Base count warning 100 Uppercase Alpha Level 0.05

** Base count minimum 30 Lowercase Alpha Level 0.1



Compare Each Sample

Letter comparisons use Pearson Chisq LR PValues

DSS, Bias

LR Chi-square p-value on pairs

	A	B	C	D
A	1.0000	0.0105	0.0255	0.0050
B	0.0105	1.0000	0.0001	0.4044
C	0.0255	0.0001	1.0000	<.0001
D	0.0050	0.4044	<.0001	1.0000

Pearson PValues

Pearson Chi-square p-value on pairs

	A	B	C	D
A	1.0000	0.0110	0.0263	0.0068
B	0.0110	1.0000	0.0001	0.4071
C	0.0263	0.0001	1.0000	<.0001
D	0.0068	0.4071	<.0001	1.0000

Figure 8: Potential Bias by DSS

4.0 DISCUSSION

For classification accuracy, all the DSSs that were designed and tested outperformed the no DSS conditions. However, the interpretation time increased with additional information in the DSS. This finding suggests that for time critical tasks, careful attention must be paid to the amount of information in the decision aid as a more accurate decision that is not made within the necessary timeline would not meet mission objectives. From a sensor perspective, the classification accuracy of EO was greater than IR which, in turn, was greater than SAR which was in order of the literalness of the sensor. As may be anticipated, the timing numbers were reversed with SAR taking longer than IR which took longer than EO.

Both trust and confidence were enhanced with the DSSs; however, these subjective measures and eye tracking data showed that the reference imagery was the major contributor to the participant's trust and confidence, not the DSS algorithm probability. The study dug deeper in attempting to ascertain the dependency of the interpretation task on sensor type. To accomplish this goal, the heuristics that the participants used to make their

decisions were elicited. It was found that for EO and IR, confirmation (decision primarily driven by test imagery) was the heuristic that was used most often, and for SAR, anchoring (decision primarily driven by DSS) was used most often. Again, this is understandable based on the literalness of each sensor. Also, this finding was also supported by the eye tracking data as the fixation time on the reference imagery increased with the non-literalness of the sensor (i.e., SAR time > IR time > EO time). On the other hand, when considering possible biases (i.e., errors resulting from heuristic use), anchoring was the culprit for all sensor types.

5.0 FUTURE WORK

This experiment can be further analyzed by looking into the fixation count, visit count, and transitions of areas of interest within the DSS. The eye tracking data can be used to see if there are any other lessons learned, specifically, exploring the effect of the secondary suggestion. Did the secondary suggestion have a positive impact or was it too much information? Eye tracking will also be able to give a firm answer to whether the secondary suggestion was considered by determining if the participants looked at this area of the DSS.

REFERENCES

[1] Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5), 527-539.

[2] Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics*, 2, 1926-1943.

[3] Helldin, T., Falkman, G., Riveiro, M., Dahlbom, A., & Lebram, M. (2013). Transparency of military threat evaluation through visualizing uncertainty and system rationale. *Engineering Psychology and Cognitive Ergonomics*. Applications and Services, 8020, 263-272.

[4] Liebhaber, M. J., & Feher, B. (2002). Air threat assessment: Research, model, and display guidelines. SPACE AND NAVAL WARFARE SYSTEMS COMMAND SAN DIEGO CA.

[5] Hoffman, R. R., Johnson, M., Bradshaw, J. M., & Underbrink, A. (2013). Trust in automation. *Intelligent Systems*, IEEE, 28(1), 84-88.

[6] Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing Decisions Improved Decision Making With a Single Training Intervention. *Policy Insights from the Behavioral and Brain Sciences*, 2372732215600886.

[7] Fischoff, B. (1981). *Debiasing* (No. PTR-1092-81-3). DECISION RESEARCH EUGENE OR.

[8] Keren, G.; Caverni, J.P.; Fabre, J.M.; et al. *Cognitive Biases.* Amsterdam: North-Holland 1990.

[9] Soll, J. B., Milkman, K. L., & Payne, J. W. (2014). A user's guide to debiasing.

[10] Larrick R. P. (2004). “Debiasing,” in *Blackwell Handbook of Judgement and Decision Making* eds Koehler D. J., Harvey N., editors. (Malden, MA: Blackwell Publishing Ltd.) 316–337.

[11] Arnott, D. (2006). Cognitive biases and decision support systems development: a design science approach.

Information Systems Journal, *16*(1), 55-78.

[12] Fendley, M., & Narayanan, S. (2012). Decision aiding to overcome biases in object identification. *Advances in Human-Computer Interaction*, *2012*, 7.

[13] Zelnio, H., & Fendley, M. (2018). Human performance modelling for image analyst decision support design. *International Journal of Human Factors Modelling and Simulation*, *6*(2-3), 184-202.

[14] Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, *66*(1), 102-110.

[15] Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory.* Psychology Press.

[16] Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

[17] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

[18] Wall, E., Blaha, L. M., Paul, C. L., Cook, K., & Endert, A. (2017). Four perspectives on human bias in visual analytics. In *DECISIVe: Workshop on Dealing with Cognitive Biases in Visualizations*.

[19] George, J. F., Duffy, K., & Ahuja, M. (2000). Countering the anchoring and adjustment bias with decision support systems. *Decision Support Systems*,*29*(2), 195-206

[20] Bhandari, G., Hassanein, K., & Deaves, R. (2008). Debiasing investors with decision support systems: An experimental investigation. *Decision Support Systems*, *46*(1), 399-410.

[21] Zenker, F., Dahlman, C., & Sarwar, F. (2015). Reliable Debiasing Techniques in Legal Contexts? Weak Signals from a Darker Corner of the Social Science Universe. *The psychology of argument: cognitive approaches to argumentation and persuasion*.

[22] Zachary, W. W. (1980). *Decision aids for naval air ASW* (No. 1366-A). ANALYTICS INC WILLOW GROVE PA.

[23] Woods, D. D., & Roth, E. M. (1988). Cognitive Systems Engineering. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 3-43): Elsevier Science.

[24] Hopson, J., Zachary, W., & Lane, N. (1981). The intelligent use of intelligent systems: Problems in engineering man/machine symbiosis. *AGARD The Impact of New Guidance and Control Systems on Mil. Aircraft Cockpit Design 16 p(SEE N 82-13048 04-01)*.

[25] Mitchell, C.M. & Miller, R.A. (1986) A discrete control model of operator function: A methodology for information display design, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 16 No.3, pp.343-357.

[26] Muller, Amanda C., & S. Narayanan. Cognitively-engineered multisensor image fusion for military

applications. *Information Fusion* 10.2 (2009): 137-149.

[27] Mitchell, Christine M. "GT-MSOCC: A domain for research on human-computer interaction and decision aiding in supervisory control systems." Systems, Man and Cybernetics, IEEE Transactions on 17.4 (1987): 553-572.