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

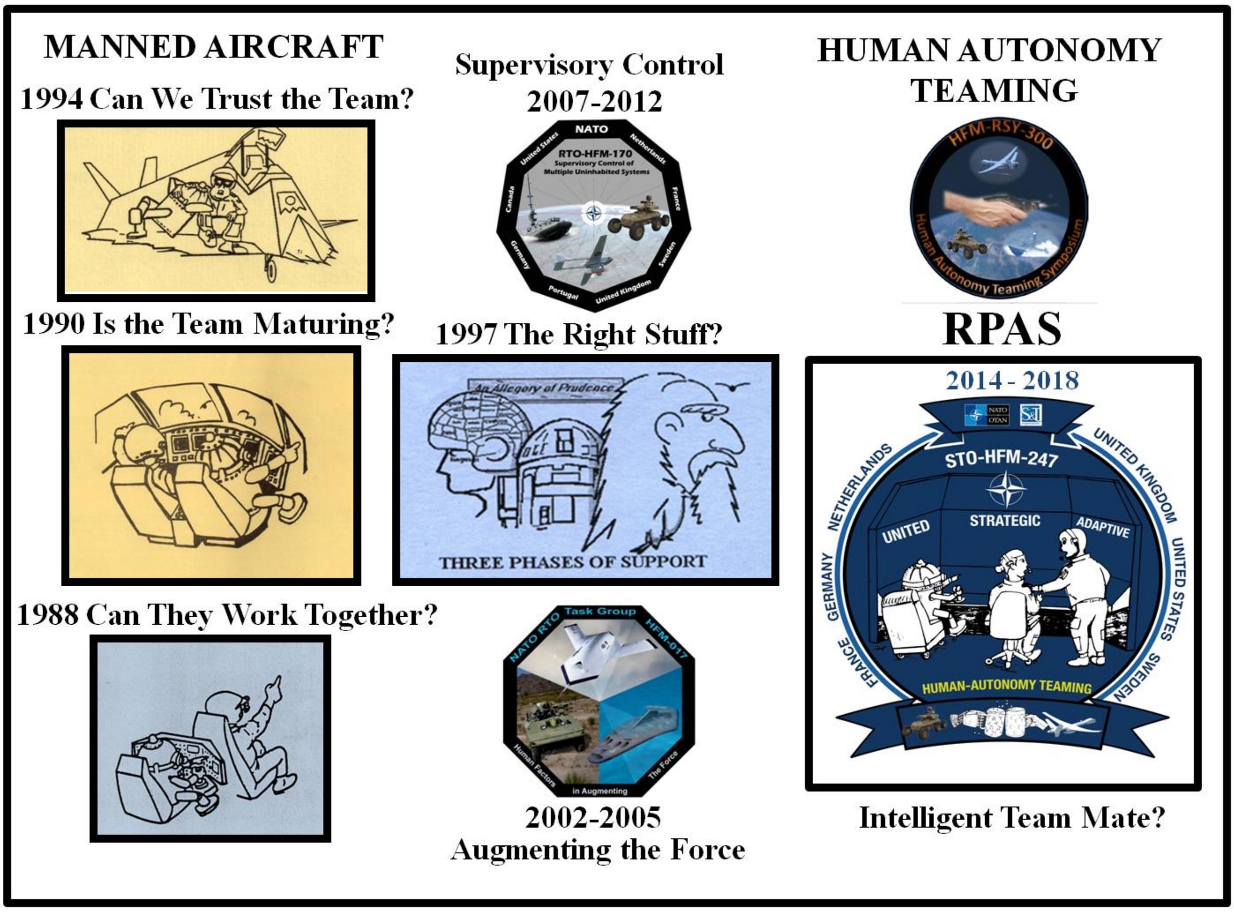

Figure 1: Precursor Human Autonomy Teaming Collaboration Activities.

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.

PAPER	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
TQ Average	5.00	3.76	5.12	4.80	5.84	4.45	4.53	4.97	5.62	5.67	5.97	3.63	6.00	5.48	5.80		3.68	4.67		4.27	4.60	3.33	5.70

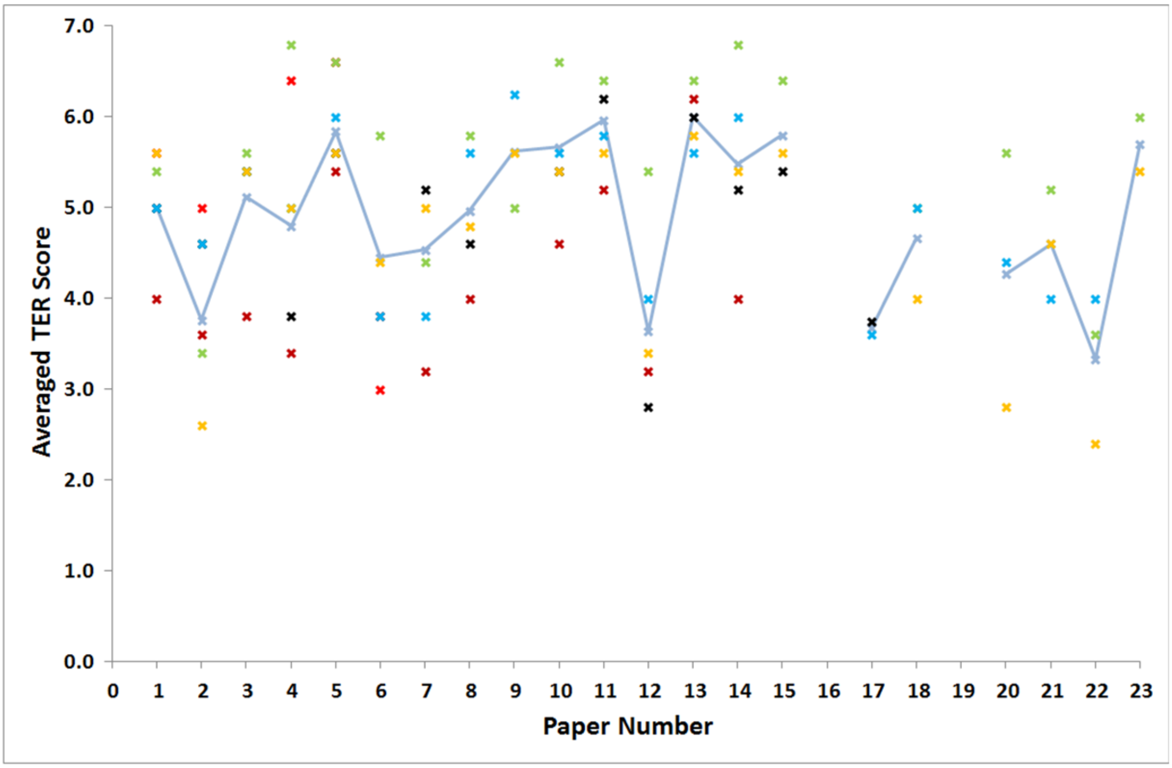

Figure 5: Averaged Ratings for Symposium Contributions

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.

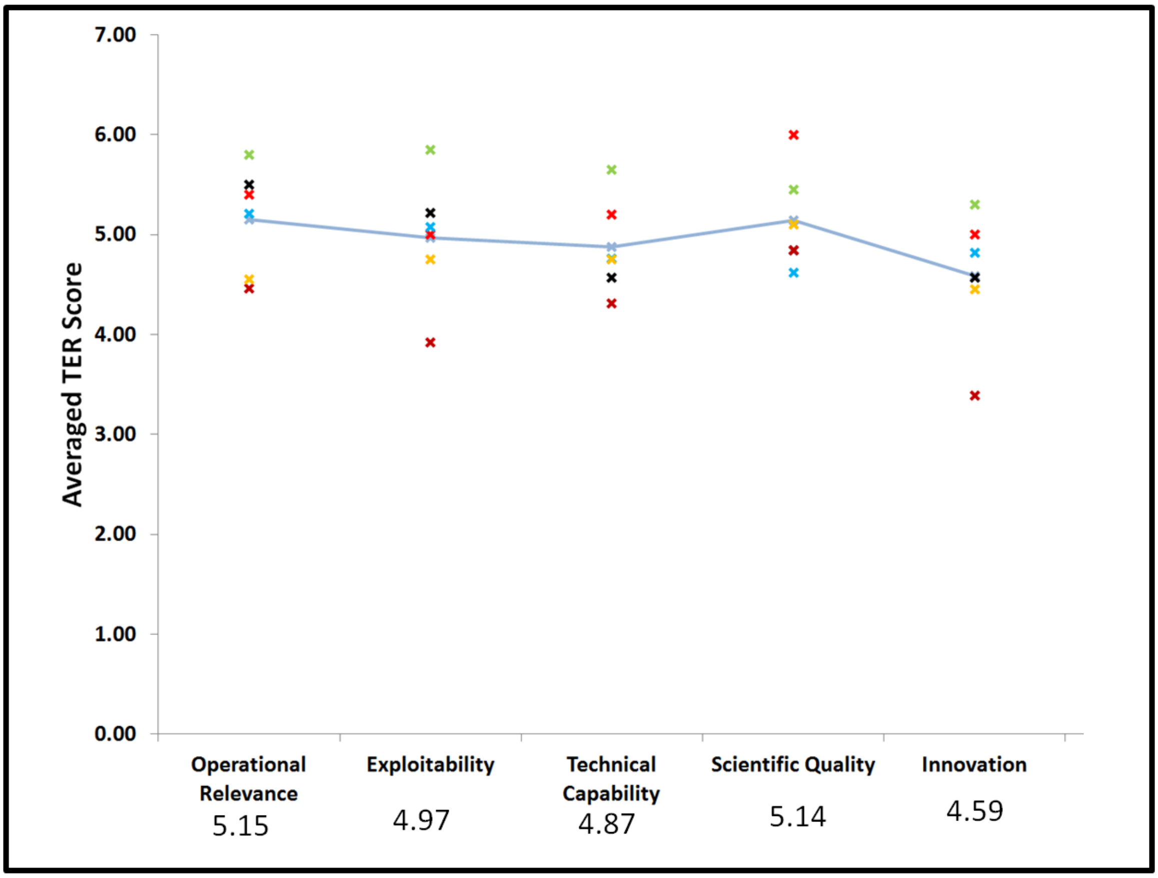

Figure 6: Averaged Ratings for Assessment Categories

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.

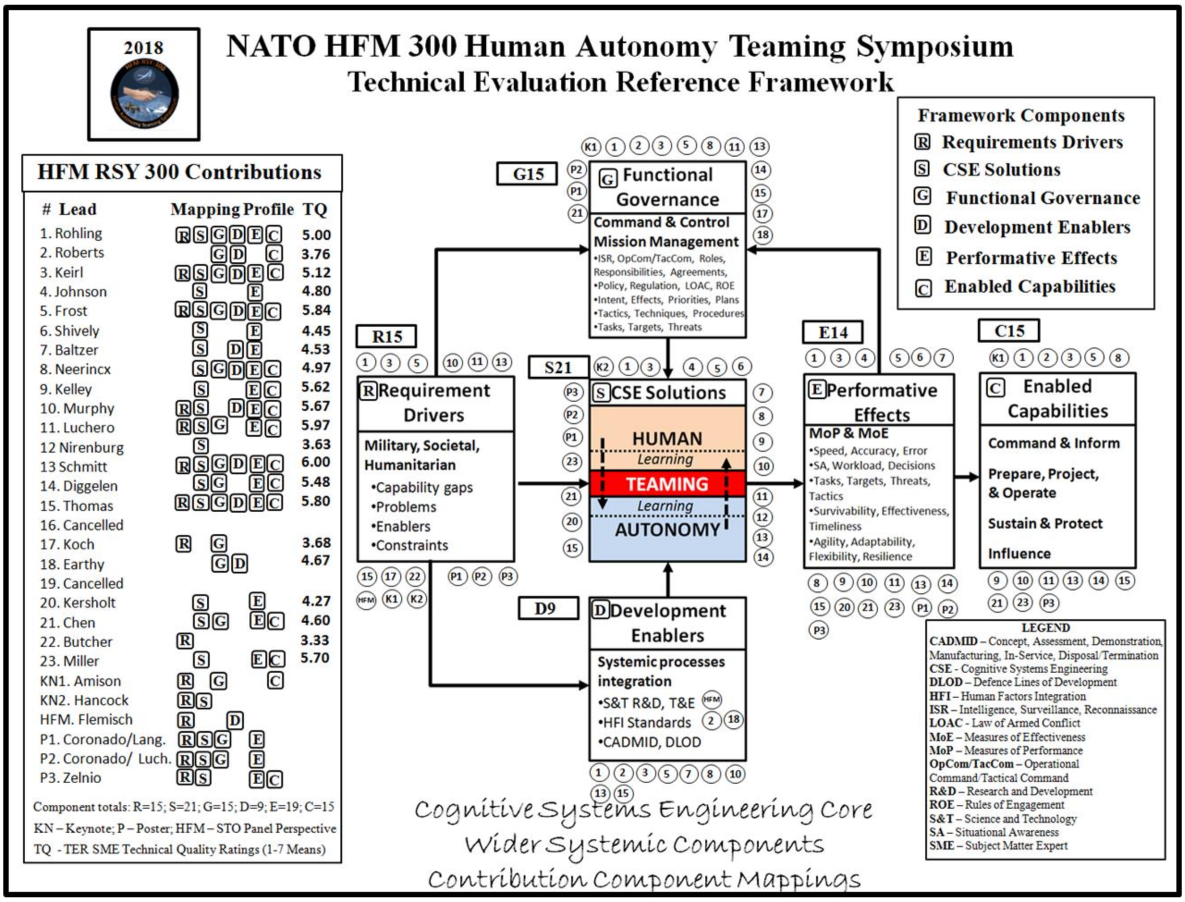

Figure 7: Results of Wider System Research Requirements Analysis

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.

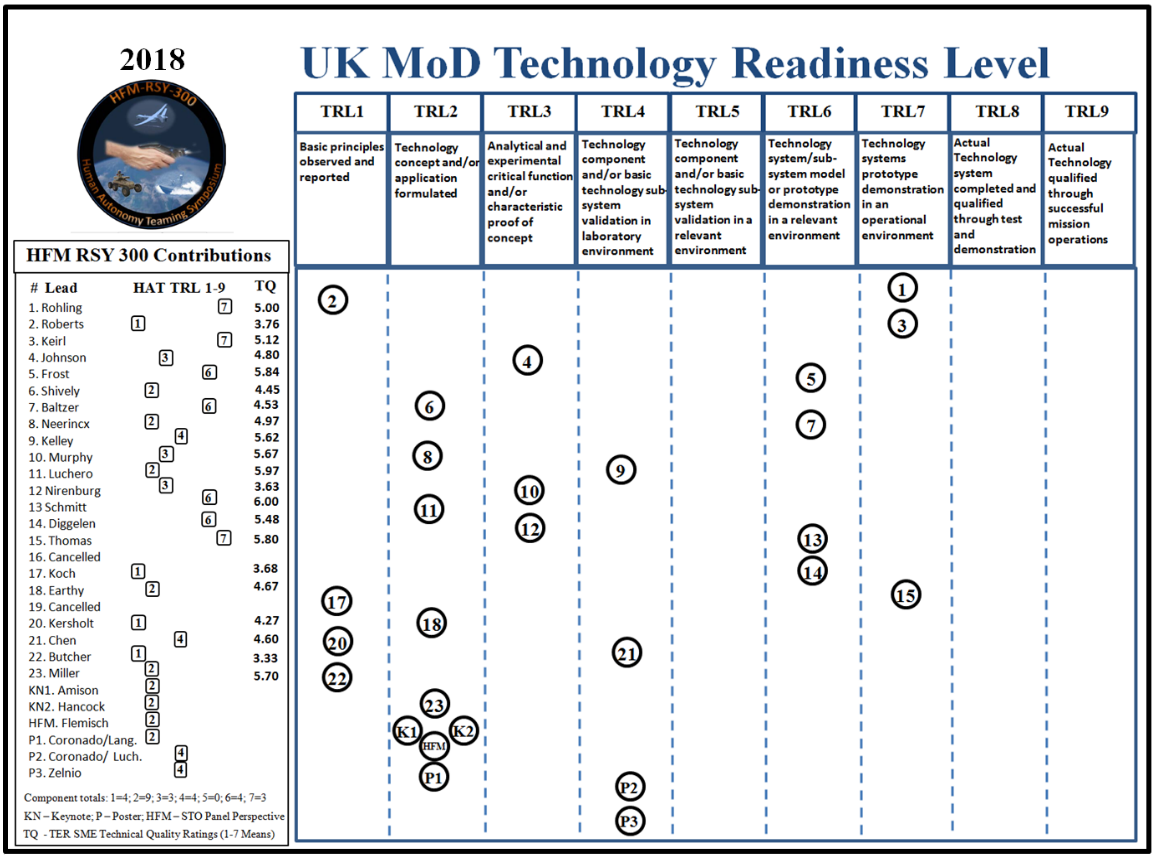

Figure 8: Results of Technology Readiness Levels Analysis

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].

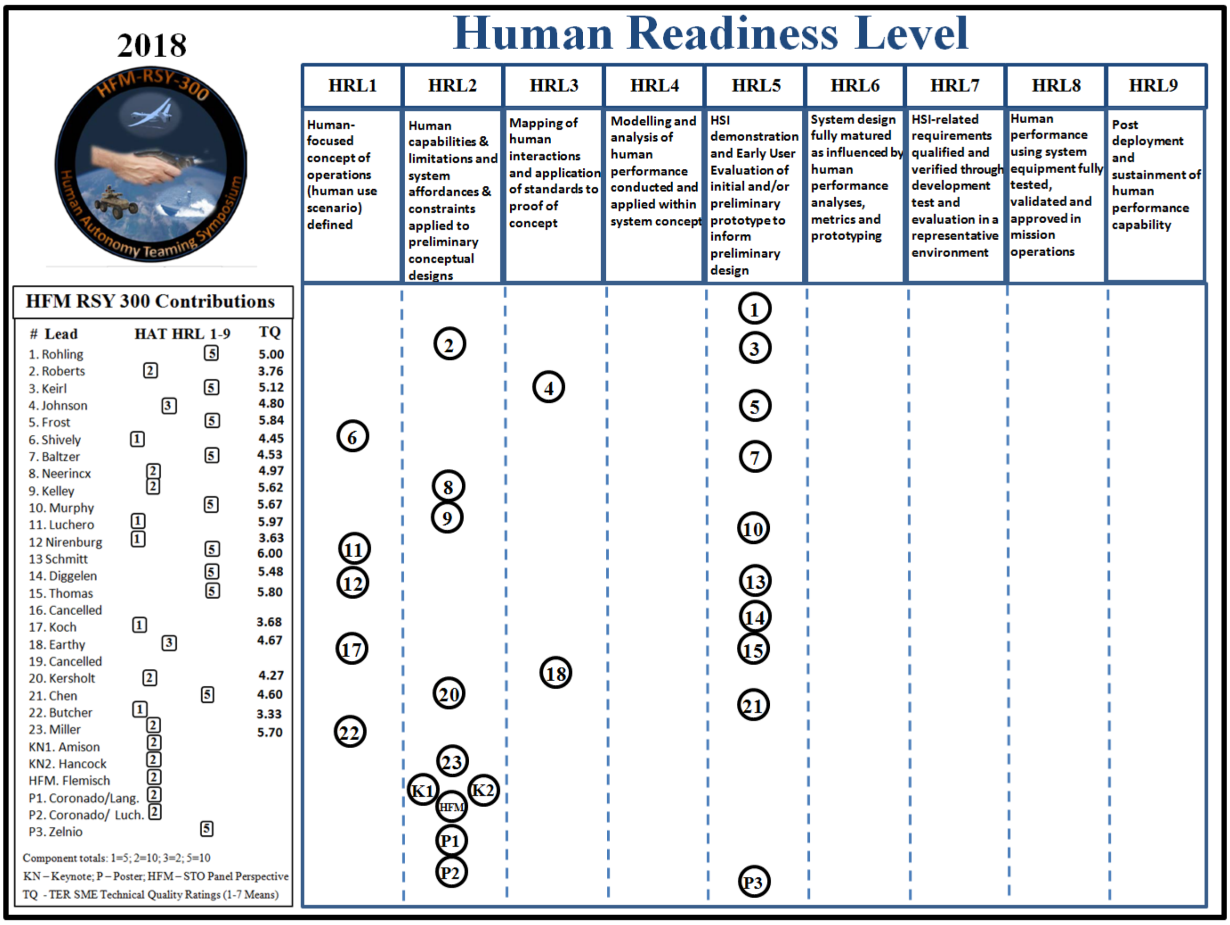

Figure 9: Results of Human Readiness Levels Analysis

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.

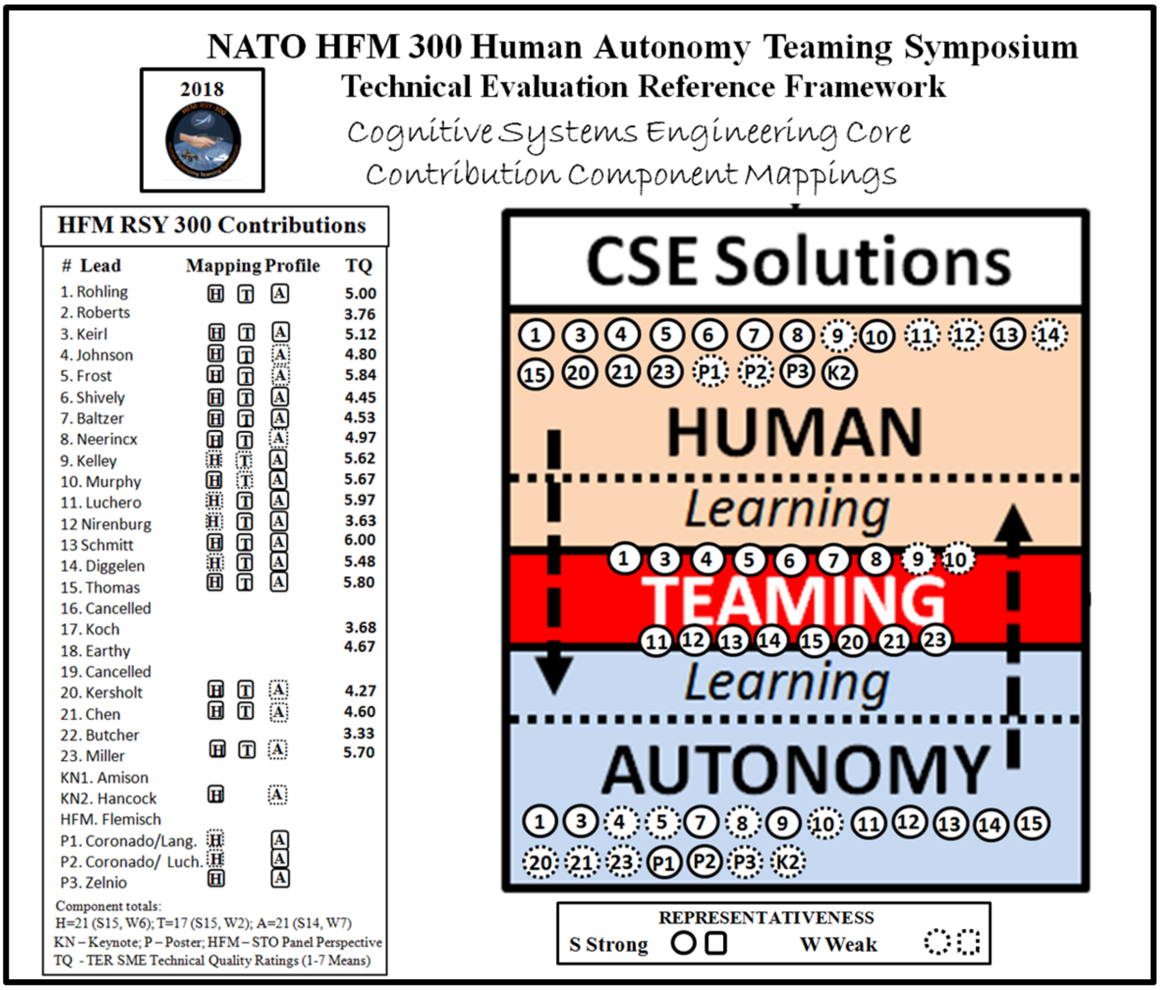

Figure 10: Human Autonomy Teaming Analysis

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.

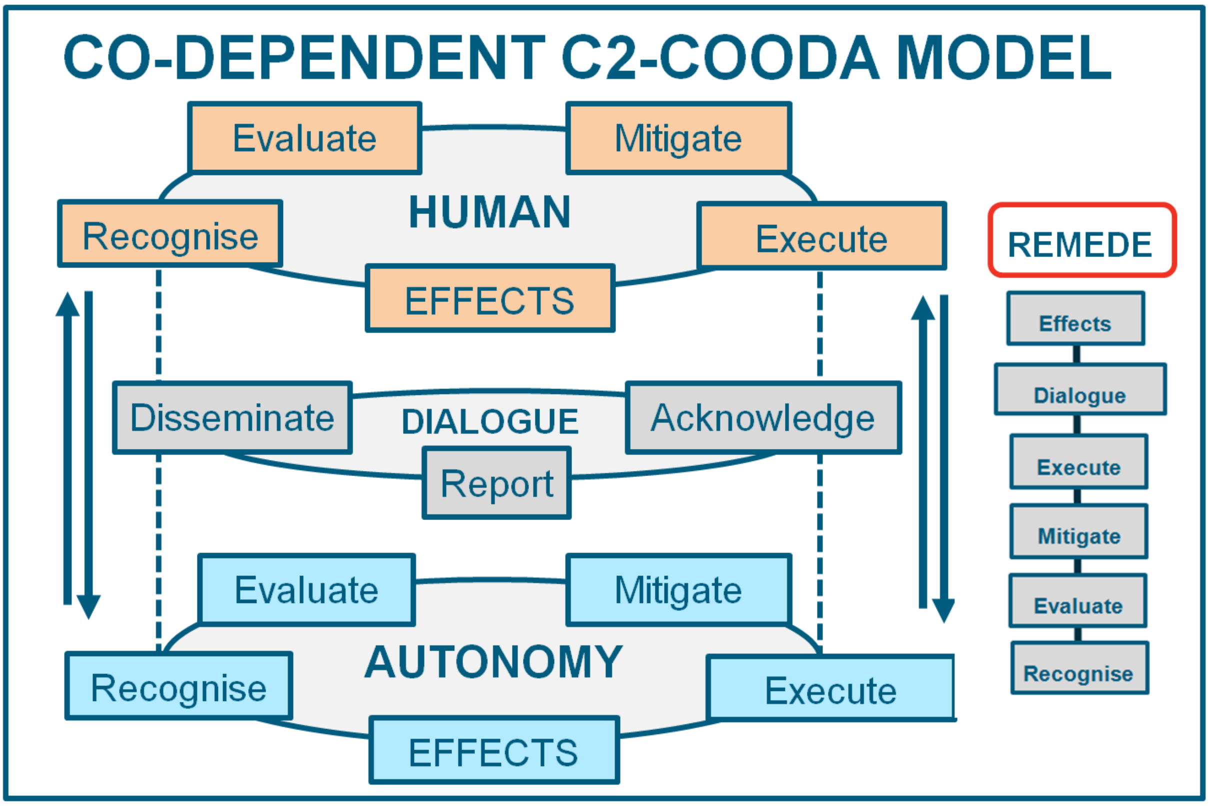

Figure 11: REMEDE Decision Comparison Model

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

Situation awareness-based Agent Transparency

Level 1: *Agent's current status/actions/plans*
- Purpose: Desire (Goal selection)
- Process: Intentions (Planning/Execution); Progress
- Performance

Level 2: *Agent's reasoning process*
- Reasoning process (Belief/Purpose)
- Environmental & other constraints/affordances

Level 3: *Agent's projections/predictions; uncertainty*
- Projection of future outcomes
- Uncertainty and potential limitations; Likelihood of success/failure
- History of Performance

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.

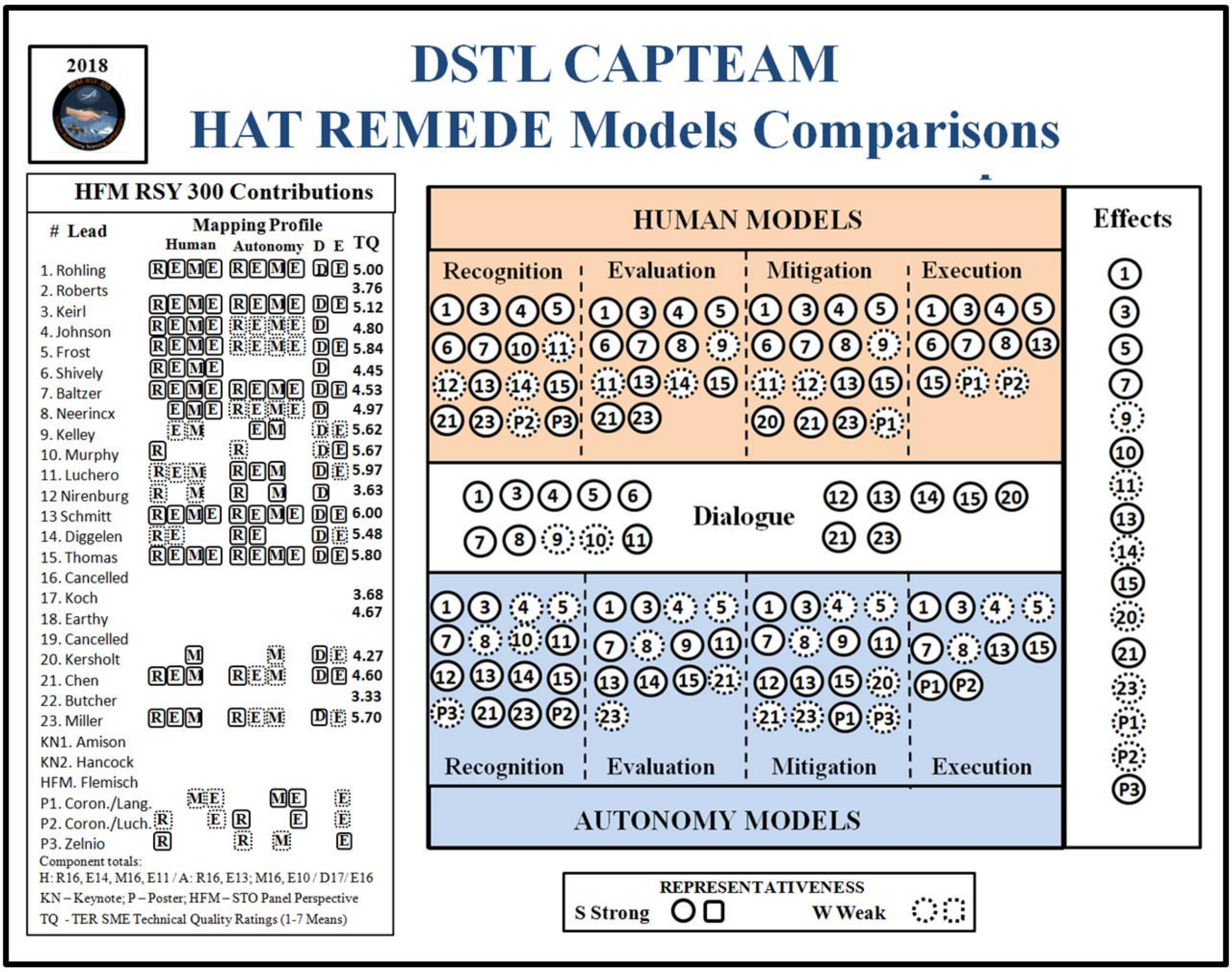

Figure 14: Results of Decision Models Analysis

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).

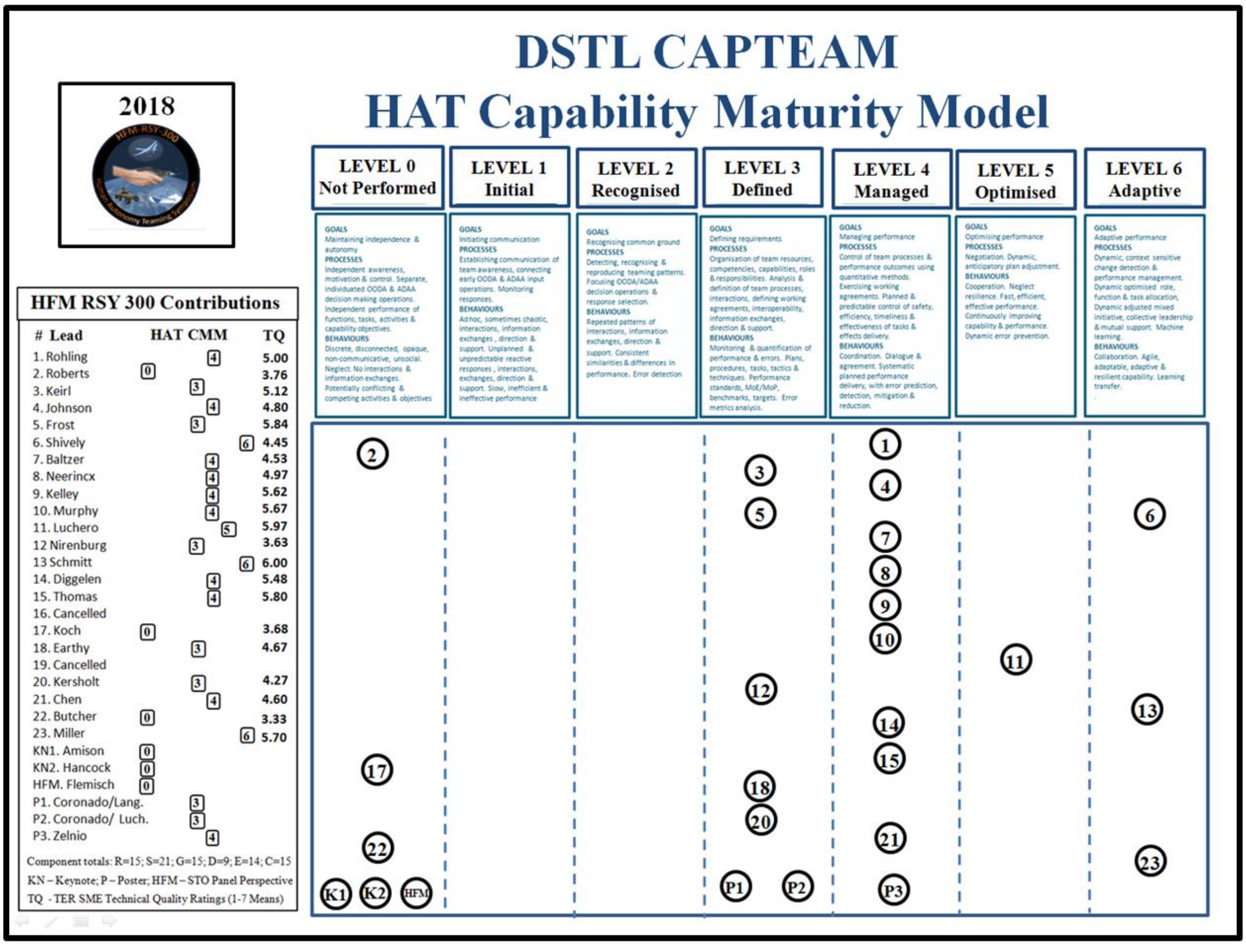

Figure 15: Results of Capability Maturity Level Analysis

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	

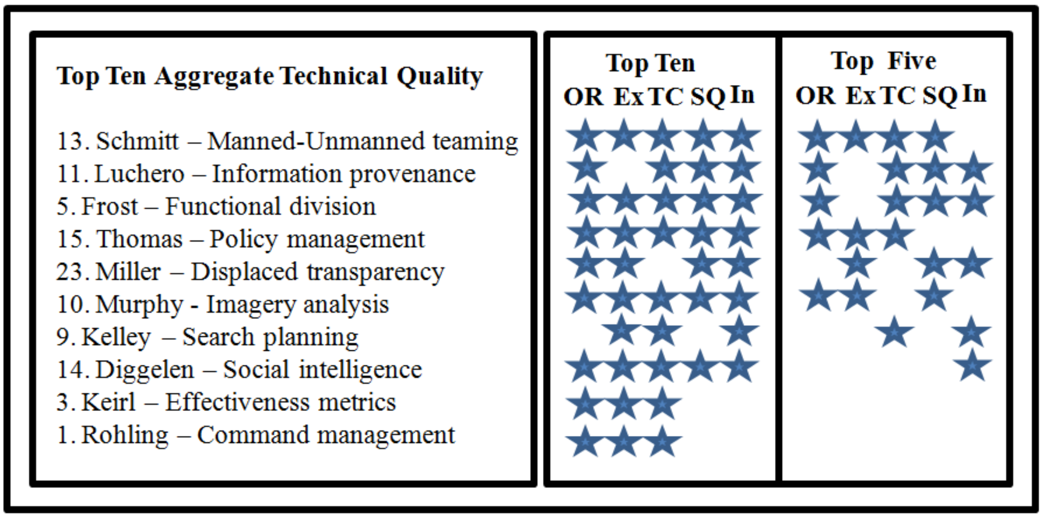

Figure 16: Top Ten TQ Obtaining Top Ten and Top Five Technical Component Ratings

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].

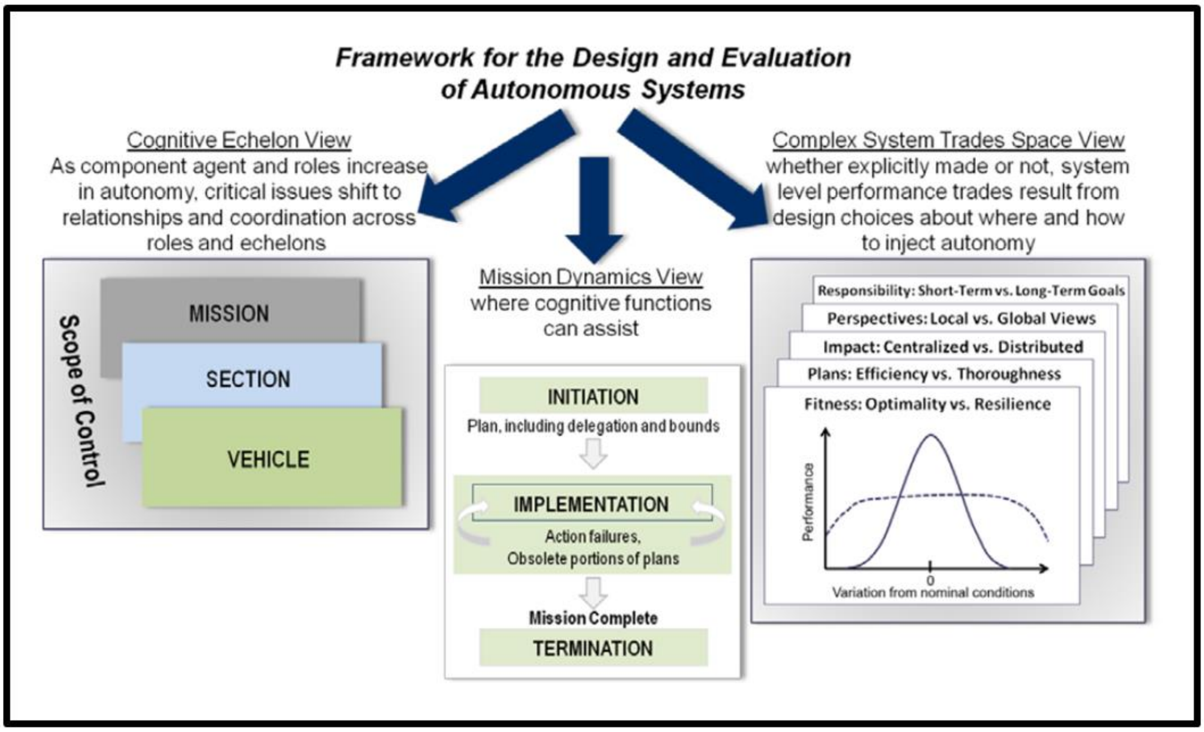

Figure 17: US Defense Science Board DSB 2012 Task Force Framework for the Design and Evaluation of Autonomous Systems [4]

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.

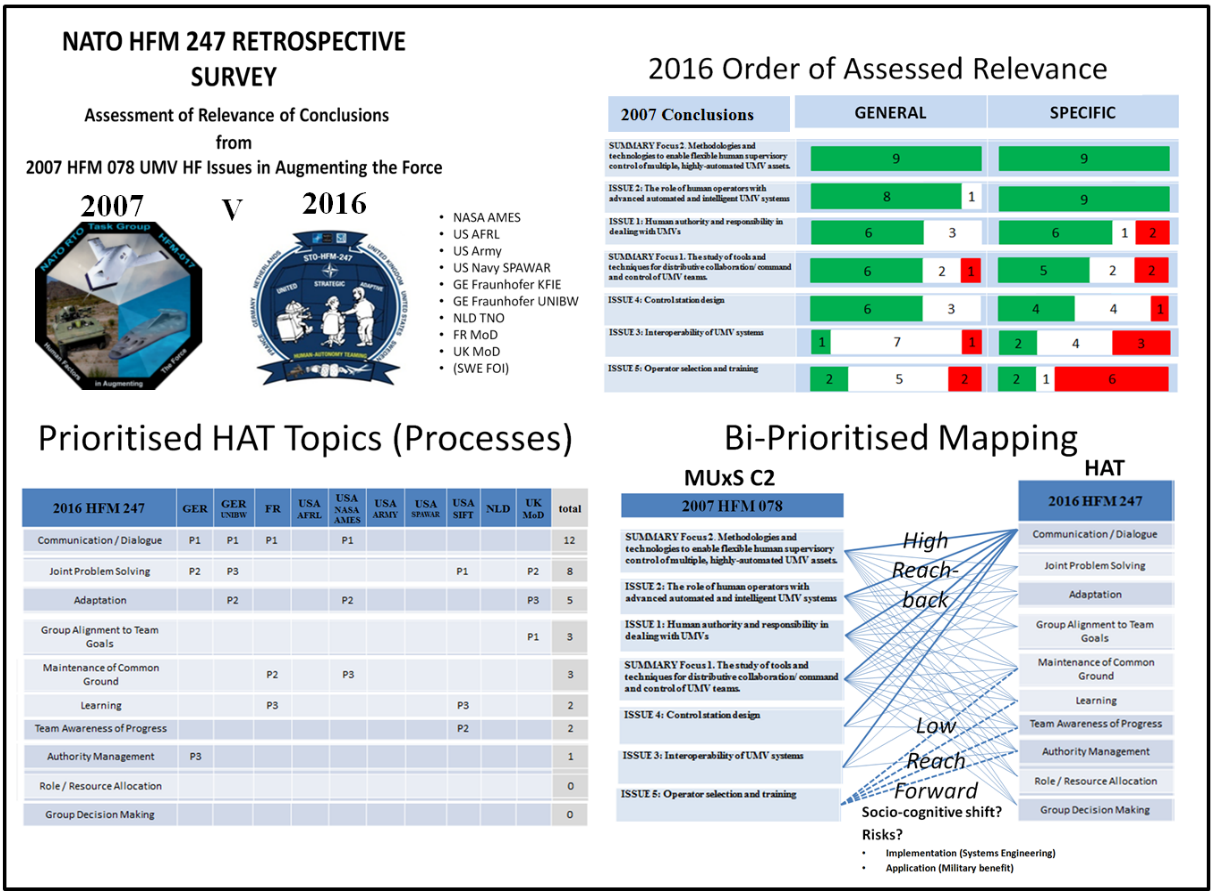

Figure 18: NATO HFM Task Group 2016 Retrospective Survey

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

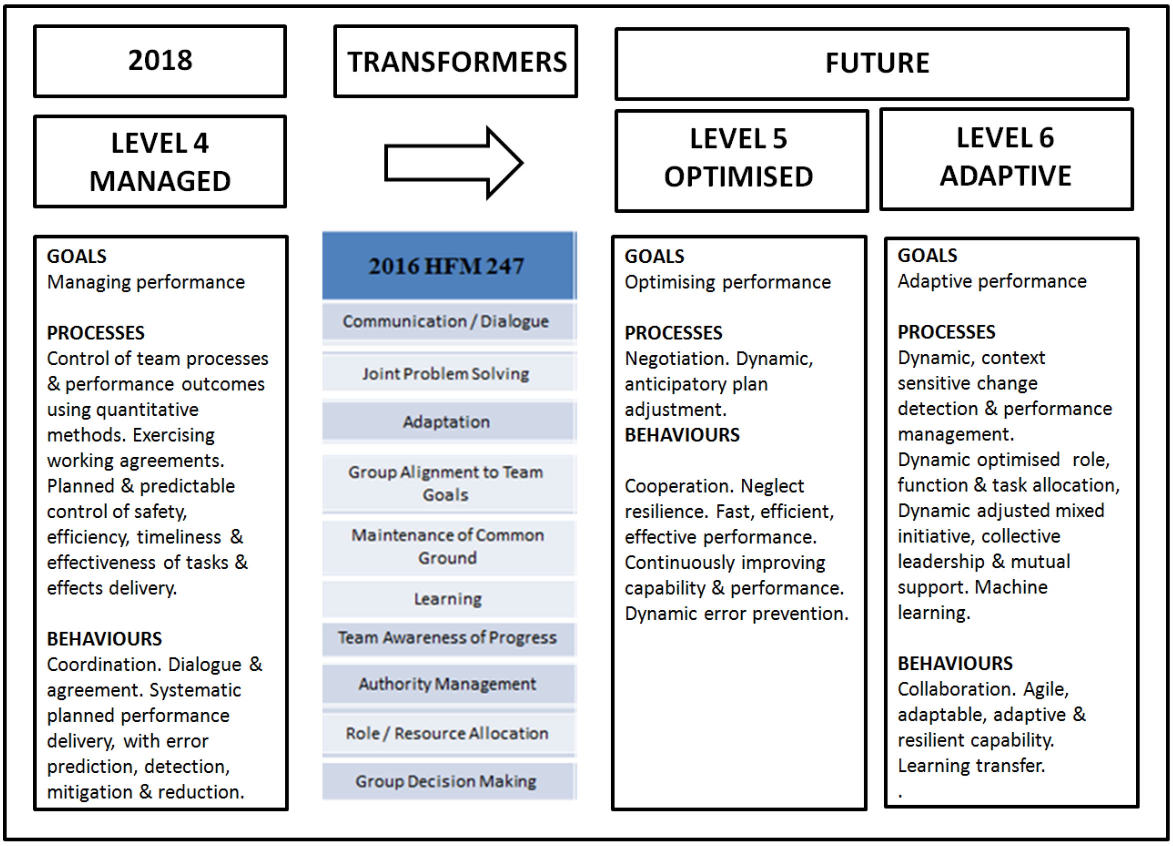

Figure 19: Means for Progression to Teaming Goal of Adaptive Performance

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.

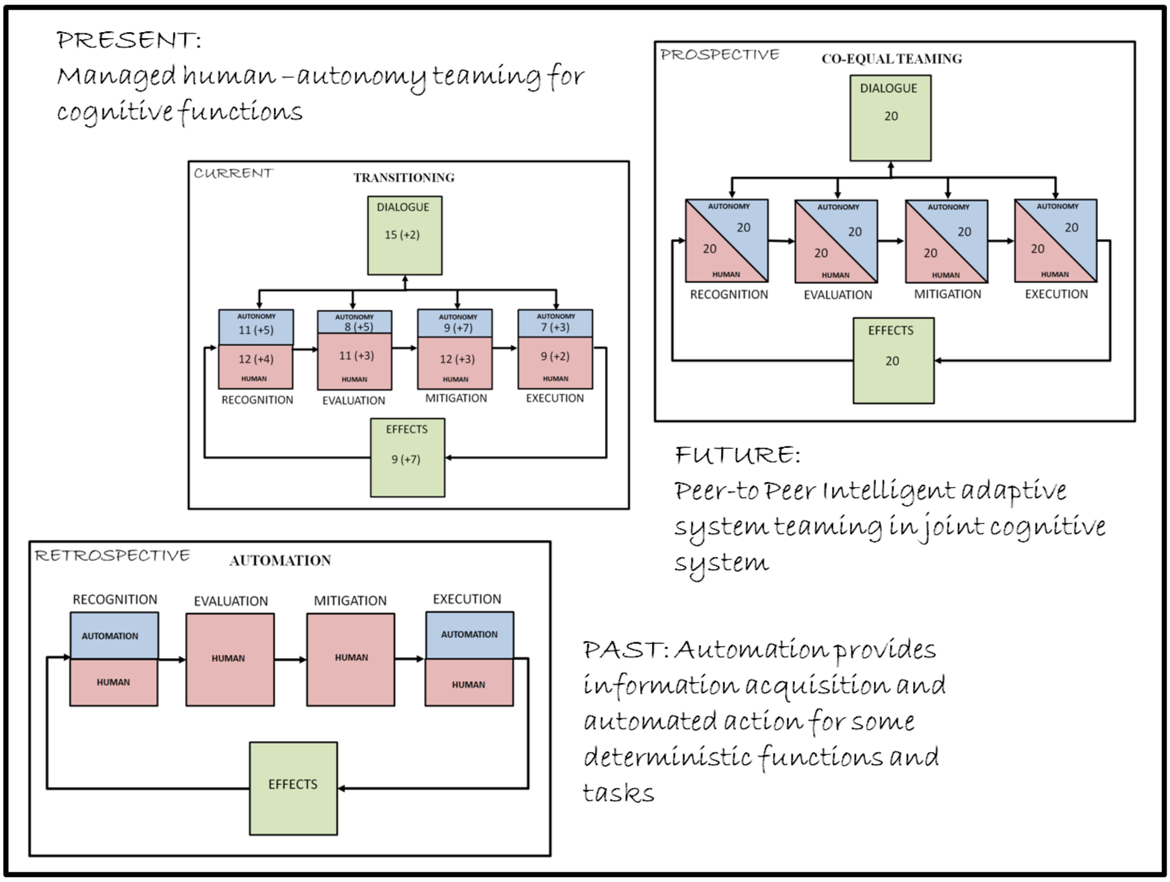

Figure 20: Past, Present and Future Perspectives

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

<table>
<tr><th colspan="3">SESSION 1: OPERATIONAL REQUIREMENTS AND MEASURING HAT EFFECTIVENESS</th></tr>
<tr><td>No.</td><td>AUTHORS</td><td>TITLE</td></tr>
<tr><td>1</td><td>Rohling.</td><td>Supporting Infantry Units with UGV and UAS</td></tr>
<tr><td colspan="3">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.</td></tr>
<tr><td>2</td><td>Roberts, Soper.</td><td>Identifying HF considerations associated with command, control, communication of autonomous vehicles in the maritime domain</td></tr>
<tr><td colspan="3">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</td></tr>
<tr><td>3</td><td>Keirl, Thorpe, Taylor, Grabham.</td><td>Measuring effectiveness of HAT</td></tr>
<tr><td colspan="3">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.</td></tr>
<tr><td>4</td><td>Johnson, Vignati, Duran.</td><td>Understanding human autonomy teaming through interdependence analysis</td></tr>
<tr><td colspan="3">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.</td></tr>
</table>

<table>
<tr><th colspan="3">SESSION 2: HUMAN AUTONOMY TEAM STRUCTURE AND COMMUNICATIONS (1)</th></tr>
<tr><td>No.</td><td>AUTHORS</td><td>TITLE</td></tr>
<tr><td>5</td><td>Frost, Bartik, Calhoun, Spriggs, Ruff, Behymer.</td><td>Creating a well situated human autonomy team: The effects of team structure</td></tr>
<tr><td colspan="3">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.</td></tr>
<tr><td>6</td><td>Shively, Coppin, Lachter.</td><td>Bi-directional communication in human-autonomy teaming</td></tr>
<tr><td colspan="3">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</td></tr>
</table>

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		

<table>
<tr><td colspan="3">SESSION 5: HAT INTERACTION AND DESIGN (PART1)</td></tr>
<tr><td>NO.</td><td>AUTHORS</td><td>TITLE</td></tr>
<tr><td>13</td><td>Schmitt, Brand, Rudnick, Schulte.</td><td>Experimental evaluation of a cooperative automation approach for manned-unmanned teaming in future military helicopter missions</td></tr>
<tr><td colspan="3">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.</td></tr>
<tr><td>14</td><td>Diggelen, Barnhoorn, Peeters, Staal, Stolk, Vecht, Waa, Schragen</td><td>Pluggable social artificial intelligence for enabling human-agent teaming</td></tr>
<tr><td colspan="3">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.</td></tr>
<tr><td>15</td><td>Thomas, Cottrell, Grabham.</td><td>Policy management and negotiation: Enabling effective human autonomy teaming</td></tr>
<tr><td colspan="3">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.</td></tr>
</table>

<table>
<tr><td colspan="3">PANEL: HAT INSTITUTIONAL INTEGRATION</td></tr>
<tr><td>No.</td><td>AUTHORS</td><td>TITLE</td></tr>
<tr><td>17</td><td>Koch</td><td>Ethical challenges of lethal autonomous weapons systems and manned-unmanned teaming.</td></tr>
<tr><td colspan="3">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.</td></tr>
<tr><td>18</td><td>Earthy, Downs</td><td>Human-system interaction standards for robotic, intelligent, autonomous systems</td></tr>
<tr><td colspan="3">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.</td></tr>
<tr><td>20</td><td>Kersholt, Barnhoorn, Hueting, Schuilenborg</td><td>Automation as an intelligent teammate: Social psychological implications</td></tr>
<tr><td colspan="3">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.</td></tr>
</table>

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	