Human Factors and Systems Interaction, Vol. 84, 2023, 189–209
https://doi.org/10.54941/ahfe1003605

Unified Vision-Series of Military Exercise - The Methodological Struggles in Conducting Evaluation of Human Factors and Systems Interaction

Rune Stensrud, Torgar Haugen, and Sigmund Valaker

Norwegian Defence Research Establishment, N-2007 Kjeller, Norway

ABSTRACT

The complexity and practical constraints of sociotechnical systems demand that researchers develop better understanding and methods for testing and providing input on critical Human Factor (HF) and Systems Interaction problems. This article takes a concrete approach to this problem describing the NATO HFM-276 Task Group methodological struggles in conducting evaluation of a military exercise. To analyze the duality; theory and practice, in the area an overall Socio-Technical Multiteam (STM) model was used to identify critical Human Factors (HF) and Systems Interaction in order to survey the gap between simulation and best practice. A dilemma that occurs in these events is the one between being a training arena for military units versus being an arena for operational testing of systems and new capabilities, may lead to some challenges for HF methods. This speaks to the aspect of multiplicity and uncertainty. The behavior of the distributed cooperative systems and operators during the preparation phase and execution are made up of multiple, highly interconnected individuals who influence one another formally and informally. A particular aspect of the Unified Vision series of exercises is the ongoing modifications of systems (e.g. Command and control systems and information systems) and their use during the trial itself. Operators and Supervisors are to analyze data seeking to identify targets across Intelligence, Surveillance, and Reconnaissance (ISR) operations. Evaluating these operations included Human Factors and Systems Interaction (pre- and post-exercise) surveys, observational, archival, and document study. We discuss the utility of these methods using recent work in the HF and ergonomics literature on how to match HF methods to core aspects of the complexity of sociotechnical systems. We also discuss how the methods used could capture complexity, and how to improve HF evaluations of training and test practices in the military and other field settings taking into account some aspects of the complexity of ISR operations. The example of mapping organizational structure using a problem space, as well as experimental designs could be avenues for accommodating complexity in future HF evaluations of military exercises.

Keywords: Military exercise, Human systems integration, Evaluation, Intelligence, Surveillance and reconnaissance (ISR)

INTRODUCTION

The digitization of command and control and Human Factors in Defense requires analyzes beyond just looking at the main processes in a military headquarters, including intelligence, surveillance and reconnaissance (ISR). Understanding of people and the interaction between process and needs for Teamwork Assessment, Interface Evaluation, Usability Testing and Human Error Identification is extensive. The combination of theory and practice of Human factors and Cognitive Systems Engineering (Von Bertanalfy, 1960; Shannon & Weaver, 1948; Fitts, Schipper, Kidd & Shelly, 1957) in order to be able to evaluate the balance between technology, system and process. Further, the utility of these methods using recent work in the HF and ergonomics literature on how to match HF methods to core aspects of the complexity of sociotechnical systems may be overwhelming. In its one right, our research have the goal of study ISR issues, as best as possible, with the methods most convenient, and at hand to provide a basis for human decision-making and answer for investments and expectations for improvement of policy, doctrines, in addition to evaluating a military operational concept named: *Joint ISR*.

JISR is a joint operational process, where the joint operational headquarters is responsible for the implementation and development of the process. The intelligence element in a joint operational headquarters participates in the process together with the operational and planning element in the headquarters. (The NATO JISR concept emphasizes that JISR is an activity primarily related to data acquisition.) At the joint operational level, many parts of the management process are expressed in the form of JISR. Elements from this process can be useful to use on a tactical level as well. The purpose of JISR is to provide a faster and better basis for decision-making at all command levels, by coordinating all acquisition needs and all available sensors, both intelligence sensors and other sensors. Tactical commanders with intelligence sensors under their command lead these to support the tactical mission. However, the overall sensor and analysis capacity must also be seen as a whole. A key element is to establish an understanding of intelligence and collection needs across the various departments in the Allied Armed Forces. In this way, the Defence's overall sensor capacity will be better utilized to respond to the highest-priority acquisition needs. This may help to facilitate the efficient utilization of allocation of sensors and the processing resources, but it is also of great value when it comes to the exchange of data, information and intelligence.

Technological developments are enormous and continuously open up new opportunities for using military platforms as sensors to cover acquisition needs, even if this is not the platforms' primary function. These possibilities require that such use is planned as part of the other operational planning. The process must thus ensure an efficient, synchronized and coordinated use of available sensors and production capacity that are not part of the intelligence organisation. However, ISR technologies do not work in isolation. Rather than working in isolation, ISR technologies need to connect with human operators. This is certainly the case in ISR operations where intelligence operations are fundamentally human-centric. Consequently, it seems important to include a Human Factors (HF) research methodology in the ISR CD&E process. A full evaluation of an ISR concept's ability to improve operator decision-making, as well as to improve policy and higher-level decision-making, cannot simply rely on the technology alone.

Holman et al. (2021, p. 1408) claim that the HF discipline is moving "beyond analyses of individual people interacting with individual artifacts." The HF discipline moves and involves, requiring attention to details related to the user's perspective and taking care of the system owner, who must be enabled to realize that they are building a socio-technical system, and that human factors must therefore be taken care of from the beginning. Failure to understand basic needs and complexity when many people and system components work together will seriously hinder both effective collaboration and the safety of the final system, illustrated by complex military systems. Generally, from an ergonomics perspective, the concept of complexity has a number of distinct features Woods (1988). Walker et al. (2010), and more recently Read et al. (2021), established several ways in which complexity is a key aspect of the social systems studied by human factors (HF) and ergonomics, with complexity defined according to an attribute view (multiplicity, dynamism and uncertainty), as a quantitative measure or as an emergent phenomena (Righi & Saurin, 2015). A key implication for HF research based on these insights is the requirement to avoid reductionism and oversimplification in HF analysis of complex phenomena (Walker et al., 2010; Holman et al., 2021). An example is the emergence effects where higher order phenomena are based on simpler interaction. Focusing only on the simple interactions could preclude an understanding of the emergent phenomena. Based on this insight to mitigate shortcomings of various methods, Walker et al. (2010) suggest that human factor (HF) and ergonomics research should undertake a multi-method approach to increase "predictive efficiency" concerning ergonomics problems with varying levels of complexity. Holman et al. (2021) followed up on this idea by offering an analytic framework for matching HF method to attributes of complexity. Their analysis indicated that for less complex HF problems there are many methods (Matthews, Gerald, and Lauren Reinerman-Jones., 2017. ISR example p. 97), whereas there is a need to develop HF methods for more complex problems (Holman et al. 2021, p. 1431-1432).

In this article, we endeavor to characterize methods for gathering and analyzing data from a particular sociotechnical system (Bickman et al. (2009)), and reflect on how methods could be combined (Creswell et al., 2013) to accommodate potential complexity of the HF problems at hand. We reflect on the methods used to identify HF issues in the Unified Vision-series of military exercises. The specific task of these exercises where joint intelligence, surveillance and reconnaissance tasks (ISR) comprising both the command levels as well as the physical platforms collecting data and the teams processing data, exploiting and disseminating information (TCPED). Importantly new technical systems were tested and used in these exercises (i.e. command and control and information systems). The exercises exemplify many of the characteristics of a highly complexity described by Walker et al. (2010). An important question is how, in future evaluations, to ensure a better consideration for complexity, and that is the main topic of this article. This article discusses and proposes approaches to handle the challenge of evaluating trials and experiments within a complex military system.

Going forward, we can apply the findings from this study to future work in the JISR CD & E process as well as in other joint processes (Jassemi-Zargani et al., 2013). To be sure, a HF evaluation methodology involves a focus on the operators' perceptions of the effectiveness of an ISR concept. This HF evaluation can be divided into hard and soft elements (see Figure 1) where the hard elements refer to the operators' perception of the quality, quantity, completeness, and latency of the data they receive from the ISR system(s) and concept(s). In contrast, the soft elements are concerned with the meta-cognitive aspects of the decision-making process. Accordingly, the evaluation of these soft elements assesses the operators' trust in the system and others as well as their views on whether the system can meet the information requirements of the various groups involved in the operation. The evaluation would also examine the operators' assessments of the ISR concept to use different combinations of ISR platforms at any given time, organizational structures, the concept's ability to facilitate the development of a shared Common Operating Picture (COP) and SA, and how well the concept facilitates efficient dissemination of information and coordination among entities. Moreover, since much of JISR operations occur within multi-national contexts; this part of the HF evaluation can examine the operators' views on how differing cultural issues impact JISR operations. (Adapted from NATO-STO-HFM-276, 2023.)

We then provide some examples of how to include complexity in the study of ISR operations by applying a quantitative complexity metric using a dimension-reducing strategy projected on a "problem space" (Moffat, 2003; Alberts et al., 2006; Alberts and Hayes; 2006). The graphic presentation of changes within a problem space over time could provide a parsimonious way of illustrating both actual and potential complexity of a system, in particular if it is augmented with an experimental approach, which ensures a higher degree of control.

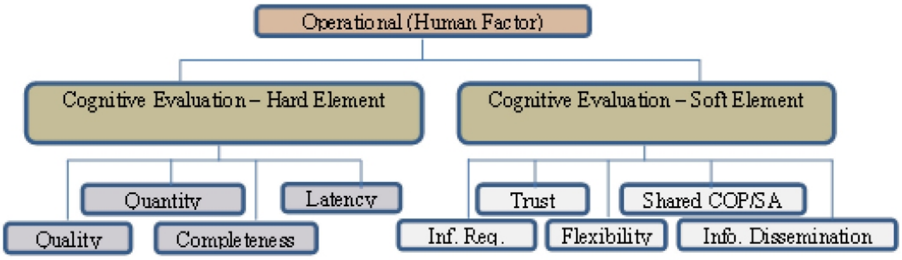

Figure 1: Operational (human factors) effectiveness evaluation criteria and sub-criteria hierarchy. (Adapted from NATO-STO-HFM-276, 2021; Jassemi-Zargani et al., 2013.)

Lastly, we discuss some opportunities for future research and practical implications.

THE TEST BED: UNIFIED VISION 2018

The modeling of the multinational organization and finding effective measurement variables for the socio-technical system that unfolded during the exercise Unified Vision 2018, was one of the goals of the evaluation that NATO HFM-276 RTG was in charge of during the NATO exercise that took place in NATO United Vision 2018 (UV18). The trial simulation was partly a traditional military exercise and a distributed simulated sensor-shooter trial, at the United States Air Force Europe (USAFE) Warrior Preparation Center (WPC), Einsiedlerhof, Germany June 11 to 26, 2018. The panel conducted similar investigations at the Bold Quest 2019 (BQ19) exercise conducted in Finland, May 2019 (Munday, 2018; U.S. European Command Public Affairs (May 2019)).

The Research Approach to Investigating HF-Issues in the UV-Series of Military Exercises

The goals of the NATO HFM-276 RTG were to identify critical HF issues for effective ISR operations, and use a theoretical model of behavior to develop our research methodology and understand our findings, and to make recommendations regarding the use and implementation of HF research in ISR CD&E operations.

Table 1 presents Summary of Research Techniques used by NATO-STO-HFM-276 group of experts. A range of techniques were used: in-depth interviews and observation on site, as well as surveys obtained at various time-points (pre post and daily surveys), as well as study of business documents and archival information.

Evaluating the Researcher's Toolkit – Finding the Fingerprints of Complexity When Investigating a Military Exercise

While the methods used provide rich empirical material, a military simulation trial exercise is sensitive to initial conditions e.g. resources available for the participants of the exercise and the level of preparations (such

Table 1. *Summary of research techniques and the techniques used in the UV-series used by HFM-276 in bold (green = methods in use, yellow = methods partly in use, red = not in use)* (NATO-STO-HFM-276, 2021).

Qualitative techniques	Quantitative techniques
Action research	Simple Experiments, Repeated-Measures Design Experiment, Matched-Pairs Designed Experiments and Live Inject Experiments
Interviews	Simulations (HITL)
Surveys	Surveys
Observational research	Archival studies
Business documents, TTPs, manuals	
External or archival data, data logs	

as the quality of exercise plans). It is also sensitive to fluctuations over time and the interpretation of the unfolding events. This indicate that the interpretation of the findings obtained through the chosen methods, should take into consideration initial conditions and fluctuations over time during the exercise in HF attributes (e.g. 'dynamism', 'uncertainty' and 'multiplicity') to have a better understanding of the complexities of the exercise.

Specifically, the Unified Vision-series of military simulation exercises are carried out every two years in a multi-lateral military setting (several nations taking part). There is no permanent exercise organization managing these events in-between the events, therefore over the years, these exercises are dependent on interconnected individuals who influence one another and ad hoc groups inside NATO and NATO member countries. In particular cross-sectional data gathering, e.g. surveys, could be sensitive to such aspects of dynamism because data obtained at one point in time likely do not reflect the situation at another time-point. This was partly mitigated by including surveys at different time points.

A dilemma that occurs in these events is the one between being a training arena for military units versus being an arena for operational testing of systems and new capabilities. This may lead to some challenges choosing HF methods. This speaks to the aspect of multiplicity and uncertainty (Walker et al., 2010). The behavior of the distributed cooperative systems and operators during the preparation phase and execution are made up of multiple, highly interconnected individuals who influence one another formally and informally. A particular aspect of the UV trials is the ongoing modifications of systems (e.g. Command and control systems and information systems) and their use during the trial itself. This could lead to an increase in the multiplicity of events, because tasks are performed in multiple and to some extent unforeseen ways (for an example of a similar case of event complexity see: (Clewley & Nixon, 2019)). While the multiplicity of events by itself could be seen as an attribute of complexity, using the quantitative approach interconnections between events could inform about the potential for complexity, for example at different exercises there could be different number of goals and plans that produce the same overall goal (Walker et al., 2018).

Multiplicity could make it more difficult to interpret phenomena: e.g., are they examples of technical testing, of changes in sequences of actions or both. The use of direct observation as well as interviews could mitigate some of these challenges by allowing for the researcher access to a deeper understanding of activities and whether they are technical or of an operational character. In addition, the use of archival data could be a source of mapping the network between events, but was not used by the NATO-STO-HFM-276 evaluation team (Hærem, Pentland & Miller, 2015). Nevertheless, the researchers have done some groundwork to use the data for such analysis in the future.

Furthermore, the behavior of complex adaptive systems, which is said to be enhanced by informational influences as it spreads through the system would also cause challenges for data collection. In such systems, minor environmental events might generate minor changes in the behavior of a small subset of these individuals (adopters). Sometimes during the execution, slight behavioral change among these adopters (partners) might spread through the entire system and create larger changes. The entities (e.g. PED cells) where highly interdependent and thus changes in workflow in one cell, for example due to changes in C2 systems and the sequencing of actions may have had an important influence on the whole system of entities. As more entities and people join the adopters, the influence and strength of this force for change becomes magnified. The phenomenon known as "rippling" is discussed in the literature of complex adaptive systems (Moffat, 2003; Andersson, 1999 cited by Davison et al. 2012), and is a key way of conceptualizing complexity (Walker et al., 2010). Here the use of detailed archival study could discern some of the ways in which for example the exchange of information at one point in time could affect the operation at a later time. However, again the interpretation of events is challenging and could lead to misinterpretation of archival data. The archival data capture only part of the information flow, and hence effects of for example informal communication (face-to-face) may not be readily available.

For a series of military simulation exercises, trivial fluctuations in initial conditions (e.g. competence of those involved and participation) can result in dramatic differences in end states due to feedback loops that either amplify or suppress emergent patterns of behavior. Walker et al. (2010) indicate that such hard to predict effects should nevertheless inform method selection. As such, some outcomes cannot be anticipated or predicted; they are unknowable until after they occur (Moffat, 2003; Aiken and Hanges, 2012; Aiken et al. 2019). In the next sections, we are to explore ways to conduct evaluations more effectively and efficiently due to these challenges when investigating the content and context of a military simulation trial exercise.

As a backdrop to integrating the human aspects of system and equipment provision, to describe the way in which complex phenomena, empirical evidence, practical measures, and most importantly, a legacy of delivering on objectives (i.e. goal striving). We present part of a full information model of *ISR processing* that seeks to combine both feedback and feedforward. The reporting requirement and the need for information can be outlined as a feedback loop in the decision ladder, as shown in Figure A-1, and as an example of instantiation of the decision process and delegation of tasks.

The decision ladder, as shown in Figure A-1, has been used in an attempt to describe *a general psychological structure of Rasmussen's Ladder, as a schema for the decision making process and* the need for information during execution as well as control (reporting requirements) for PED operations. This is our attempt to create a projection of a simple decision loop (feedback) on Jens Rasmussen's decision ladder (adapted to the ISR process). A proposal that Moffat (2003) describes in words in his book: Complexity Science. Potential threat or target object is assumed to initiate data processing and change system states. The scheme is inspired by (Moffat, 2003; Vincente, 1999; and Jenkins, 2012 in the book (Stanton, 2013, 2nd ed.) [pages 76–79].

Why Cognitive Systems "Engineering"?

A common interpretation is C(SE), meaning cognitive (systems engineering) or systems engineering from a cognitive point of view. The other is (CS)E, meaning the engineering of (cognitive systems), or the design and building of joint (cognitive) systems. We will focus mainly on the latter – how to understand a joint human-machine system in a context. The former, C(SE) represents a more traditional (and common) point of view which focus on the dyadic relationship between a human and a machine – from a cognitive point of view (where cognition is ascribed to a specific part of the system) Neisser (1976).

However, the early history of the development of the computer (late 1940:ies) was followed by Cybernetics and early Human Factors – which in turn was followed by Cognitive science, based on the information Processing paradigm. Cybernetics was interested in describing control in "the animal and the machine" – i.e. modelling (feedback driven) behavior mathematically. Human Factors was a reaction to the fact that the growing complexity of technology made the human the limiting factor in a human-machine system. The very early studies were performed in the US (Fitt's and Jones) and in the UK (Bartlett and Craik). Intrinsically, the driving forces behind Human Factors and Cognitive Systems Engineering (CSE) and the foundation of the criticism of functionalism (Parsons) and then perhaps implicit criticism of goal hierarchy thinking in HF and CSE like that of Jens Rasmussen's decision ladder (both externally and of his own colleagues at Risø Research Center, Odense, Denmark: Hollnagel. A group of Risö-researchers founded their "own" branch of CSE, which largely rejected the idea of the Skills Rules Knowledge-model and the abstraction hierarchy, instead focusing on joint human-machine systems. The leading profiles in this branch are Erik Hollnagel and David D Woods (Hollnagel et al. 2006). This dispute can be compared to what has been prominent in sociology, including Habermas, who criticizes functionalism for unilaterally emphasizing an instrumental reason. In addition, Habermas believes that there is a communicative reason that goes into clarifying a situation and thereby coordinating action. Giddens is also possibly such a critic, but with a more processual approach where structure (which can indicate functions?) changes and is changed through human activity.

G. Bateson and the whole of anthropology probably also have this as their cornerstone in a way. Bateson with his (cybernetic) thinking about cognition as based on feedback in the dialectic between process (execution) and form (idea), but philosophically: can one criticize these critics for not recognizing that under some conditions it might be highly meaningful to talk about functions, hierarchy and structure as something that resists/is robust to change?

Anyway, we draw on cybernetic theory to develop an organizational efficiency model and analyze a collection of data from a multi-national task organization labelled the UV-series of military exercises. Findings from this setting could indicate that a balanced frequency of using various media and information systems, as well as using both feedforward (long-term request for information) as well as feedback (short-term request for information) could be influential in reaching desired levels of sustained performance despite exogenous challenges. For organizations, experiencing high dynamics of an environment, the use and designing of information systems that systematically incorporate feedback as well as feedforward could be essential. We explore some theoretical implications, directions for future research as well as practical implications.

Potential Ways of Improving on HF Evaluation of ISR to Account for Complexity

As described, there are many critical issues of complexity in the particular empirical case. We illustrate new methodological and analytical tools by focusing on a critical practical issue: that of determining the change of organizational structure (based on analysis of communication) which may indicate adaption and hence performance (Walker et al., 2009; Moffat & Manso, 2008; Aiken et al., 2019). Using archival data obtained at multiple time points we were able to reconstruct the evolution of the organizational structure over time.

One way to empirically examine coordination forms in a multiteam system (Marks et al.,2005; Zaccaro et al., 2012; Davison et al., 2012; DeChurch et al. 2010; 2020; Luciano et al., 2018) is to look at communication networks. A network is “a group or system of interconnected people or things that facilitates the exchange of data and/or information to achieve higher goals” (Mathieu et al., 2018). Teams communicating through different network structures show varying degrees of performance, as demonstrated by, for example, (Park et al., 2020).

We analyzed Teams tasking log, i.e. log data, from a military exercise consisting of PED teams collaborating in a multiteam environment (Davison et al., 2012). It made sense to only focus on the data log, as this was the main common means of tasking sensors and/or PED cells and tasks communication between teams at the higher echelons. The simulated exercise tasking log we accessed involved actors on the operational and tactical level, but we omitted the sensor operator level and their communication (Levitt, 1951), as it would typically be conveyed through voice chat/ or chat tools in simulation environment not logged. There was not imposed any restrictions on the PED teams with whom they could communicate with, meaning they had the opportunity to deviate from the formal chain of command (ORBAT). Based on the data log, we constructed an adjacency matrix, and from that, communication diagrams in order to visually identify informal network structures.

We present a visualization for PED operations based on a technical Coalition Shared Data (CSD) Architecture enabling all the participants (man and machine) to take part in the PED-process. Based on the same data we also calculated the communication patterns among the teams (sorted by type of messages; initial, and planned (RFIs), or dynamic (ISR-requests)). We broke down the communication data into three separate phenomena with each phenomena (by type of communication) being subject to a social network analysis (SNA) (Valaker et al., 2020). We broke down the communication data into three separate types of communication phenomena with each phenomena being subject to a social network analysis (SNA) (Borgatti et al., 2013). The data gathered covered approximately (50%-70%) of the accumulated tacit collaborative task work during the simulated military exercise, i.e., a data dump was prepared for the analysis. From viewing the patterns of interaction among all teams in the collective of teams we found that it proceeded from planning, request for information (RFI's) to more dynamic tasking i.e., ISRrequests (Figure 2, 3 and 4).

Visually we can see a less hierarchical organization. We assessed the patterns for task allocation, and workflow between teams. The ranked graph (in Figure 2) representing the initial entity set-up is based on relational data stored in a service cluster called ISR Workflow Services (IWS), and the service; Simple Persistence as a Service (SPS). Figure 2 shows the "*de jure*" hierarchy and command lines at time 1. De jure means it largely follows the pre-defined organization. Figure 3 shows the *pre-planned* communications between the teams based upon requirements for information, and this represents the collective of teams at time 2. The pattern follows to a degree the pre-defined organization (Figure 2). When comparing Figure 3 with Figure 4 we can see interaction between more teams emerging and this represents the collective of teams at time 3. Figure 4 is the *dynamic* communication, which consists of

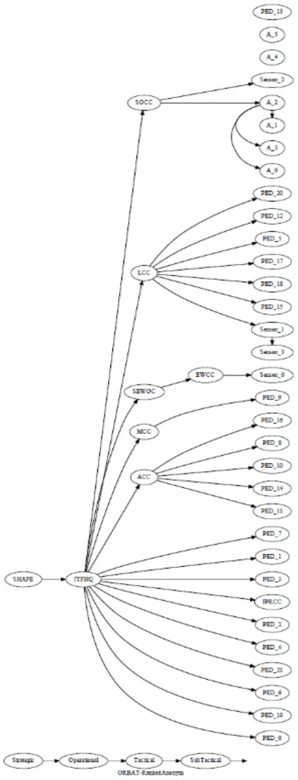

Figure 2: Ranked graph representing the initial relationships (e.g. de jure organization) between the participating teams.

Figure 3: Ranked graph representing the pre-planned relationships between the participating teams.

Figure 4: Ranked graph representing the emerging relationships between the participating teams.

teams communicating with a format prescribed for an ad-hoc and dynamic workflow.

Aiken et al., (2019) discuss pros and cons of SNA. One of our main findings so far on Practice Application with SNA is to recommend developing an "on-line" generation of SNA's daily. This could benefit reviewing a military exercise.

To be able to map the data onto a problem space which provides a 3D approximation of the attribute's view of complexity (Walker et al. (2009, p. 62). In our case, we used the NATO SAS-050 problem space dimensions named the collective "C2 Approach Space" (of the collective of federated PED cells and nodes). A matrix was constructed identifying who was requesting information from whom, producing network diagrams, and from them, computing various network statistics. The three separate social network analyses performed on the digital communication layer produced 3 separate diameter, density and sociometric status figures, i.e. three data points (x,y,z) in the collective C2 Approach Space (for an explication of these analysis see: Stanton et al., 2009): the three dimensions where: Patterns of interaction (PoI), distribution of information (DoI), and distribution of decision rights (DR).

As points of reference in the SNA, we used the classic network archetypes "Chain," "Y," "Circle," and "Wheel" (Leavitt, 1951), along with the "Edge" and "Classic C2" networks added by Walker et al. (2017), as visualized in Figure 5. Stanton et al. (2015) are referring to social network research be Beveals (1948) and Leavitt (1951) when plotting archetypes into the C2 approach space. The chain structure is characterized by unitary decision rights, tightly controlled information flow and hierarchical patterns of interaction. The Y is similar in having unitary decision rights and hierarchical patterns of interaction, but has a broader flow of information (Stanton et al., 2015). The Circle tends to hierarchical patterns of interaction, along

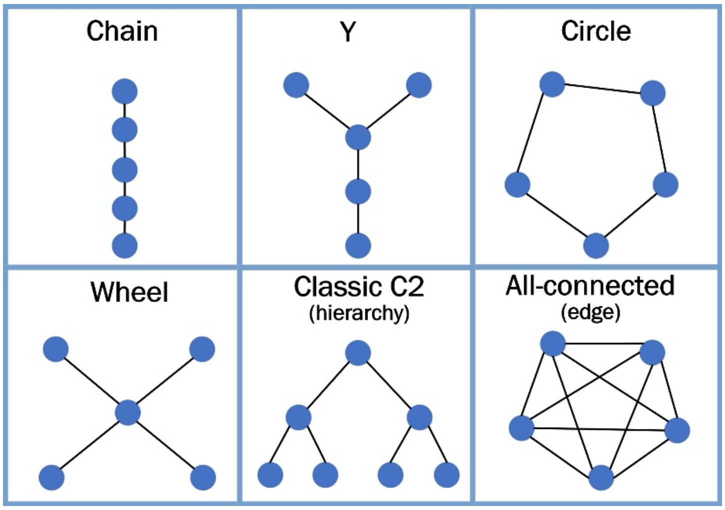

Figure 5: Six archetypical network structures (Leavitt, 1951; Walker et al., 2017).

with distributed decision allocation and broad distribution of information. The Wheel structure is similar to the chain in having unitary decision rights and tight control over the information flow, but with a more distributed interaction pattern (Stanton et al., 2015).

The All connected (edge) is essentially the opposite of the chain, characterized by distributed decision rights, broad information distribution, and distributed interaction patterns. Here all team members have a perception of equal independence and responsibility (Stanton et al., 2015). The Edge is considered a theoretical maximum rather than a practical structure (Walker et al., 2009). The Classic C2 (hierarchy) structure involves a hierarchical chain of command with layered echelons of units, with little horizontal coordination between these units (Walker et al., 2012; (Stanton et al., 2009, p. 101). We saw a slight change from hierarchical to networked structure over time from de jure, pre-plan to dynamic. The allocation of decision rights was mapped to the social network metric called 'socio-metric status' (Stanton et al., 2009, p. 101) illustrating a slight change of structure over time in Figure 6 (Valaker et al., 2020).

In the future, the positions in this problem space could be based on simulations rather than exercises. This could allow to take into account a more complete, and understandable, data set where we are able to exercise better control of initial conditions and influencing factors over time.

The mean sociometric status of the networks and standard deviation was estimated for the networks shown (Figure 2, 3, and 4) in Figure 7. (Figure 5 is re-labeled and moved to Appendix section).

Having described the method used, we now presents key findings related to the context. In context or situated: During the early history of Human factors (early 1980-1990:ies), it has been more and more acknowledged that

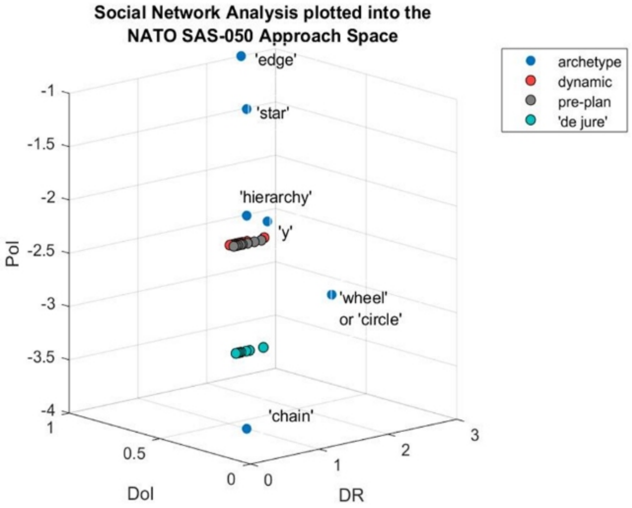

Figure 6: Allocation of decision rights was mapped to the social network metric called 'sociometric status' (Stanton et al., 2009, p. 101). Illustrating a slight change from hierarchical to networked structure over time from de jure, pre-plan to dynamic.

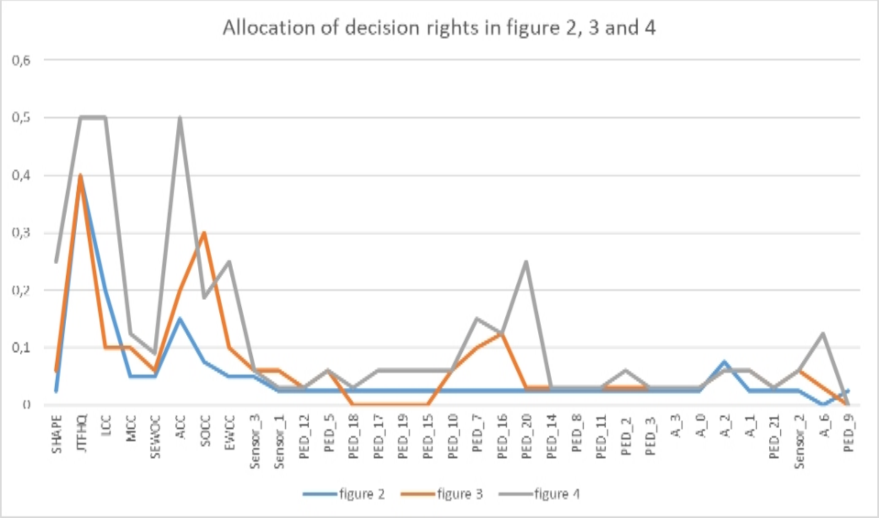

Figure 7: *Social network analysis plotted into the NATO SAS-050 approach space.* *(*Adapted from Kakimoto et al., 2006.)

cognition is "situated" and that the context in which cognition takes place is at least as important as the inner workings of the mind – be it from an information processing perspective or not.

Due to the dilemma of being a training arena for military units versus being an arena for operational testing of systems and new capabilities, we suggest an exploratory experimental design approach carefully integrating exploratory elements with controlled hypothesis testing as a method framework in future UV and other exercises.

Experimentation is an important enabler for identifying and assessing casual changes in emergent properties of a complex system resulting from technology insertion. Three popular experimental designs described in literature are Simple Experiments, Matched-Pairs Designed Experiments and Repeated-Measures Design Experiments (Frankfort-Nachmias et al., 2015).

As an example on a Simple Experiment, Unified Vision 18 was introducing an Information Quality Manager (IQM) with a prototype IQM tool system, monitoring data quality. In addition, as a mitigating Repeated-Measures Design Experiment (e.g. CIWIX measuring interoperability), and, as an example of a live inject exercise element brought into the exercise, the Baltic CESMO is a live Electronic Warfare exercise held in North Germany, 2018. Due to the design of an experiment, the logic and requirements for experiments should be discussed, as well. Although the logic of an experimental approach is simple, there are some conditions that must be met in order for us to talk about a true experiment: (1) required randomization, (2) the need of equal treatment, (3) control of environment and conditions.

According to (1): The groups must be identical before the experiment starts. In experiments like this, this will typically be achieved by randomly distributing the participants to the two conditions of the experiment. It is called randomization. Random distribution of terms may be the most important requirement for an experiment. If randomization is not done, but the study still manipulates variables, this is commonly referred to as a "quasi-experimental" survey, or a quasi experiment.

Due to (2): Participants within the same condition must receive the same treatment, and regarding (3): The investigator must have good control over other conditions that may affect the participants while the experiment is in progress. The condition or behavior of interest in the experiment must be measured in a proper way. This can happen in different ways, such as interviews or questionnaires. Typically, the researcher will use statistical procedures to draw conclusions about difference (Cacciattolo, 2015; Szabo & Strang, 1997). There was examples where (1)-(3) were not satisfied during UV18. UV18 Assessment Team reported that participants were being forced to use (a) ISR support system(s) that were not/was not fit-for-purpose and spending considerable unnecessary time to employ alternative work practices to fit the system. Other Variables Not Kept Constant: (1) Removing operators, participants, (2) Introducing unplanned systems, (3) Changing plans.

Logic-scientific testing focuses on uncovering the "truth" and providing and testing good arguments grounded on solid theory, quantified data and hypothesis testing (Aiken et al., 2019, p. 127). A hypothesis gives a proposed explanation to explain a phenomenon under investigation. Hypothesis (significance) testing involves experimenting on a sample to determine whether sufficient evidence exists to support the hypothesis. The construction of specific metrics for guiding the development of a particular information support system, staff training program, personnel management policy, organizational design, or other aspect (e.g., requirement management Hull et al., 2005) would be dependent upon the nature and focus of the individual program. Hence, it is beyond the scope of this article to present such metrics. However, it is believed that an on-going literature study provides the theoretical foundation for just such an undertaking. Additionally, according to complexity literature (Hærem et al., 2015), researchers need to consider time not only as a variable useful for predicting nonlinear changes and (time stamp) fluctuations but as part as the context surrounding and influencing the observations of behavior (Aiken et al., 2019, p. 126). In addition, may wish to use simulation to test theories, then time and logging time series; this is of interest when using modeling and simulations investigating (macro level) behavior. In this way reductionism both in terms of time, entities and their interrelations (e.g., only focusing on some entities at a certain time-point) could be mitigated. As suggested by Holman et al. (2021) this line of inquiry could be further developed by using novel computational approaches, simulations and agent based modeling.

DISCUSSION

This article discusses the methodological struggles faced conducting trials and experiments with-in a complex military system of systems. The uncontrolled nature of the activities suggests that researchers and practitioners could augment their analysis of "live" data with exploratory experiments that concentrate on specific attributes of a problem space. An exploratory experimental design approach carefully integrates exploratory elements with controlled hypothesis testing.

Using an HF methodology combined with experiments Walker et al. (2010), will make it possible to study how the PED team operators will perform critical tasks on future systems (prototypes) without ties to existing procedures (Bakdash & Pizzocaro, 2013).

FUTURE RESEARCH

Future research should more rigorously test and develop effective and practical means of analyzing and evaluating current and proposed future ISR capabilities, including distributed simulations. In particular, the identification of key attributes to be examined should be given more consideration. Modeling efforts have been done in a NATO context for real-time analysis of exercises and could be further developed (Manso, & Manso, 2010). This topic covers the design, conduct, and analysis of experiments related to any aspect of ISR, including management or governance of sensors, interoperability and sensor integration, information sharing, trust, shared awareness, shared understanding, task allocation, decision-making, collaboration, planning, coordination execution, and assessment of operations. Examples of the theoretical concepts to be further investigated are situation awareness, coordination, decision-making, trust, quality of data and stress due to the Human Factors Influencing JISR and Its Output Factors – Theoretical Framework, presented in Figure A-2. Key factors identified in the Theoretical Framework and related ISR and organizational research from military contexts. We propose to delineate how personal and interpersonal factors, organizational factors, cultural factors, task factors, system factors and team factors influence JISR. This proposed connection between input and output factors is portrayed in (Figure A-2); the figure provides an overview of possible factors to be studied in future work.

A future research challenge we need to take into consideration concerning approaches to complexity is identifying and utilizing the attributes to consider quantitative and adaptive systems (Bertino et al., 2015; Matei et al., 2015). For example, the techniques for using agent-based modeling could be further refined to take into account emergent phenomena (Stensrud et al., 2023).

PRACTICAL IMPLICATIONS

Our research identifies some key practical implications. In particular, this could be important in fast-paced contexts where there are challenges of developing an accurate understanding of the environment. This should probably be an ongoing process where the information system is updated to the needs of the context. Through such updating, an information system may embed both relevant feedforward, such as orders and as requests for information, as well as feedback, such as reports and short-term requests. This may require a rethinking of how technology is designed and employed.

Summing up the practical implications of improving and using tools and analysis of complex adaptive systems the experimental design qualifying both the operators and the (new) systems for use in experiments, should be qualified according to an agreed metric (Shepard et al., 2017). Ideally, the systems in use should be controlled and introduced in an interoperability trail (E.g. CWIX) before entering the environment of an UV exercise.

CONCLUSION

Components of complexity conceptions appear, for example using the *attribute view* describes certain aspects of complexity of a sociotechnical system, and could inform the view of such systems. Exploring for example through experimentation research designs the potential issues related to complexity could augment "field" collection of data, and provide means to investigate emergence for example in a more rigorous way. This HF methodology would work as a quality control component for the technical and procedural ISR concept development.

APPENDIX

Figure A-1 visualize a *generic robust controller architecture* following the general psychological structure of Rasmussen's Ladder, as a schema for the decision making process. The scheme is inspired by (Brehmer, 1992; Moffat, 2003;Vicente, 1999; and Jenkins, 2012 in the book (Stanton, 2013, 2nd ed.) [pages 76–79].

Figure A-2 are a proposed Theoretical Framework of Human Factors Influencing JISR and Its Output Factors (HFM-276, 2021).

ACKNOWLEDGMENT

This paper draws heavily upon the work and the products of the (NATO) (STO) TR-HFM-276 *Human Factors and ISR Concept Development and Evaluation* (2021). The authors would like to acknowledge TR-HFM-276 Chair: Dr. Fred LICHACZ, Defence Research and Development Canada (DRDC Ottawa).

REFERENCES

Aiken, J. R., & Hanges, P. J. (2012). Research methodology for studying dynamic multiteam systems: Application of complexity science. In *Multiteam systems: An organization form for dynamic and complex environments* (pp. 431–458). Routledge/Taylor & Francis Group.

Aiken, J. R., Hanges, P. J., & Chen, T. (2019). The means are the end: Complexity science in organizational research. In S. E. Humphrey & J. M. LeBreton (Eds.), *The handbook of multilevel theory, measurement, and analysis*. (pp. 115–140). American Psychological Association. https://doi.org/10.1037/0000115-006.

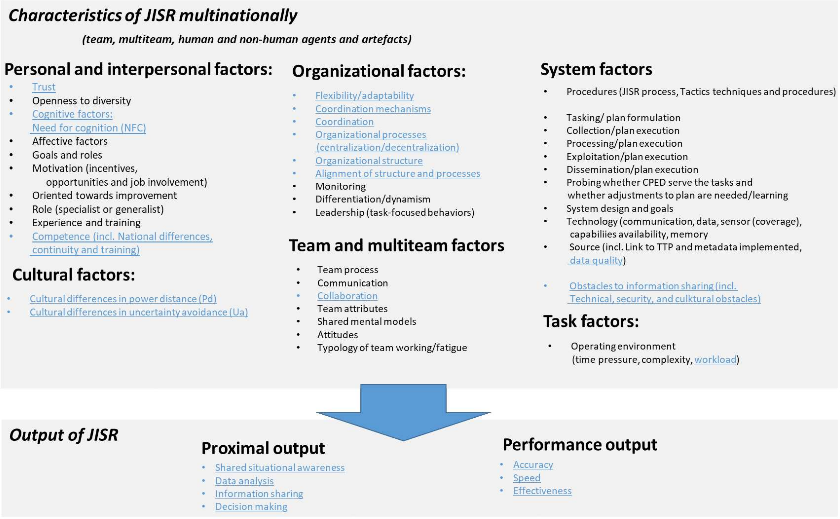

Figure A-1: A schema for the decision making process. (Adapted from Moffat, 2003; Vicente, 1999; and Jenkins, 2012 in the book (Stanton, 2013, 2nd ed.) [pages 76–79].)

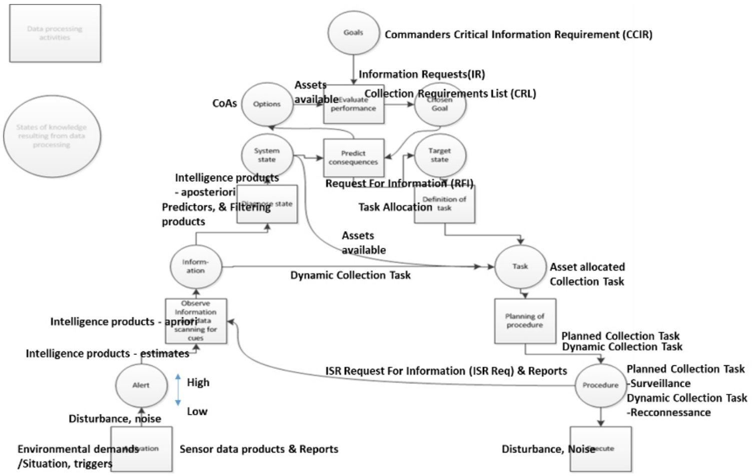

Figure A-2: Human factors influencing JISR and its output factors – theoretical framework (HFM-276, 2021).

Bakdash, J. Z., Pizzocaro, D., & Precee, A. (2013). Human Factors in Intelligence, Surveillance, and Reconnaissance: Gaps for Soldiers and Technology Recommendations. *MILCOM 2013 - 2013 IEEE Military Communications Conference*, 1900–1905. https://doi.org/10.1109/MILCOM.2013.320.

Bertino, E., Matei, S. A., & Russell, M. G. (Eds.). (2015). *Transparency in Social Media: Tools, Methods and Algorithms for Mediating Online Interactions* (1st ed. 2015). Springer International Publishing : Imprint: Springer.

Bickman, L., & Rog, D. (2009). *The SAGE Handbook of Applied Social Research Methods*. SAGE Publications, Inc. https://doi.org/10.4135/9781483348858

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. SAGE.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, *81*(3), 211–241. https://doi.org/10.1016/0001-6918(92)90019-A.

Cacciattolo, K. (2015). *Analysis of the Effectiveness of the Secondary Analysis of Existing Data in Quantitative Techniques*. https://doi.org/10.13140/RG.2.1.3368.1123.

Clewley, R., & Nixon, J. (2019). Understanding pilot response to flight safety events using categorisation theory. *Theoretical Issues in Ergonomics Science*, *20*(5), 572–589. https://doi.org/10.1080/1463922X.2019.1574929.

Clewley, R., & Stupple, E. J. N. (2015). The vulnerability of rules in complex work environments: dynamism and uncertainty pose problems for cognition. *Ergonomics*, *58*(6), 935–941. https://doi.org/10.1080/00140139.2014.997804.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (Third Edition). SAGE.

Davison, R. B., Hollenbeck, J. R., Barnes, C. M., Sleesman, D. J., & Ilgen, D. R. (2012). Coordinated action in multiteam systems. *Journal of Applied Psychology*, *97*(4), 808–824. https://doi.org/10.1037/a0026682.

DeChurch, L. A., & Zaccaro, S. J. (2010). Perspectives: Teams Won't Solve This Problem. *Human Factors*, *52*(2), 329–334. https://doi.org/10.1177/0018720810374736.

DeChurch, L. A., Bufton, G. M., Kay, S. A., Velez, C. V., & Contractor, N. (2020). Organizational Learning and Multiteam Systems. In L. Argote & J. M. Levine (Eds.), *The Oxford Handbook of Group and Organizational Learning* (pp. 602–620). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263362.013.7.

Frankfort-Nachmias, C., Nachmias, D., & DeWaard, J. (2015). *Research methods in the social sciences* (Eighth edition). Worth Publishers, a Macmillan Education Company.

Hærem, T., Pentland, B. T., & Miller, K. D. (2015). Task Complexity: Extending a Core Concept. *Academy of Management Review*, *40*(3), 446–460. https://doi.org/10.5465/amr.2013.0350.

Hollnagel, E., Woods, D. D., & Leveson, N. (Eds.). (2006). *Resilience engineering: concepts and precepts*. Ashgate.

Holman, M., Walker, G., Lansdown, T., Salmon, P., Read, G., & Stanton, N. (2021). The Binary-Based Model (BBM) for Improved Human Factors Method Selection. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *63*(8), 1408–1436. https://doi.org/10.1177/0018720820926875.

Hull, E., Jackson, K., & Dick, J. (2005). *Requirements engineering* (2nd ed). Springer.

Kakimoto, T. (2006). *Social Network Analysis on Communications for Knowledge Collaboration in OSS Communities*. http://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.112.8536&rep=rep1&type=pdf#page=43.

Levitt. (1951). Some Effects of Certain Communication Patterns on Group Performance. *The Journal of Abnormal and Social Psychology, 46*, 38–50. https://doi.apa.org/doiLanding?doi=10.1037/h0057189.

Luciano, M. M., DeChurch, L. A., & Mathieu, J. E. (2018). Multiteam Systems: A Structural Framework and Meso-Theory of System Functioning. *Journal of Management*, *44*(3), 1065–1096. https://doi.org/10.1177/0149206315601184.

Manso, B., & Manso, M. (2010). Know the Network, Knit the Network: Applying SNA to N2C2 Maturity Model Experiments. *Presented at the 15th International Command and Control Research and Technology Symposium (ICCRTS 2010),* *Santa Monica, CA, June 22-24, 2010. U. S. Government or Federal Rights License.* ICCRTS 2010, USA, CA.

Marks, M. A., DeChurch, L. A., Mathieu, J. E., Panzer, F. J., & Alonso, A. (2005). Teamwork in Multiteam Systems. *Journal of Applied Psychology*, *90*(5), 964–971. https://doi.org/10.1037/0021-9010.90.5.964.

Matei, S. A., Bertino, E., Zhu, M., Liu, C., Si, L., & Britt, B. (2015). A Research Agenda for the Study of Entropic Social Structural Evolution, Functional Roles, Adhocratic Leadership Styles, and Credibility in Online Organizations and Knowledge Markets. In E. Bertino & S. A. Matei (Eds.), *Roles, Trust, and Reputation in Social Media Knowledge Markets* (pp. 3–33). Springer International Publishing. https://doi.org/10.1007/978-3-319-05467-4_1.

Mathieu, J. E., Luciano, M. M., & DeChurch, L. A. (2018). Multiteam Systems: The Next Chapter. In *The SAGE Handbook of Industrial, Work & Organizational Psychology* (pp. 333–353). SAGE Publications Ltd. https://doi.org/10.4135/9781473914957.

Matthews, G., & Reinerman-Jones, L. (2017). *Workload assessment: how to diagnose workload issues and enhance performance*. Human Factors and Ergonomics Society.

Moffat, J. (2003). *Complexity theory and network centric warfare*. CCRP Publication Series.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* https://psycnet.apa.org/record/2003-00066-000.

Rasmussen, J. (1982). Human errors. A taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents*, *4*(2–4), 311–333. https://doi.org/10.1016/0376-6349(82)90041-4.

Read, G. J. M., Shorrock, S., Walker, G. H., & Salmon, P. M. (2021). State of science: evolving perspectives on 'human error.' *Ergonomics*, *64*(9), 1091–1114. https://doi.org/10.1080/00140139.2021.1953615.

Righi, A. W., & Saurin, T. A. (2015). Complex socio-technical systems: Characterization and management guidelines. *Applied Ergonomics*, *50*, 19–30. https://doi.org/10.1016/j.apergo.2015.02.003.

Shepard, L. A., Penuel, W. R., & Davidson, K. L. (2017). Design principles for new systems of assessment. *Phi Delta Kappan*, *98*(6), 47–52. https://doi.org/10.1177/0031721717696478.

Stanton, N. (2017). *Digitising command and control: a human factors and ergonomics analysis of mission planning and battlespace management*.

Stanton, N. A. (2013). *Human factors methods: a practical guide for engineering and design*. Ashgate Publishing Company.

Stanton, N. A., Baber, C., & Harris, D. (2009). *Modelling command and control: event analysis of systemic teamwork*. Ashgate.

Stensrud, R., Valaker, S., & Mikkelsen, B. (2023). *Orchestrating Humans and Teammates to counter security threats: Human-autonomy teaming in high and low environmental complexity and dynamism*. Intelligent Human Systems Integration (IHSI 2023) Integrating People and Intelligent Systems. https://doi.org/10.54941/ahfe1002830.

STO-TR-HFM-276, N. (2021). *Human Factors and ISR Concept Development and Evaluation*. NATO.

STO-TR-SAS-050. (2007). *Exploring new command and control concepts and capabilities = Exploration de nouveaux concepts de commandement et controle et de leurs capacits*. NATO Research & Technology Organisation.

Szabo, V., & Strang, V. R. (1997). Secondary Analysis of Qualitative Data: *Advances in Nursing Science*, *20*(2), 66–74. https://doi.org/10.1097/00012272-199712000-00008.

Valaker, S., & Stensrud, R. (2020). The influence of harmonization levels on multi-domain operations effectiveness: The moderating roles of organizational environment and command and control variety. *International Command Control and Technology Symposium*. https://www.academia.edu/45184131/The_influence_of_harmonization_levels_on_multi_domain_operations_effectiveness_The_moderating_roles_of_organizational_environment_and_command_and_control_variety.

Vicente, K. J. (1999). *Cognitive work analysis: toward safe, productive, and healthy computer-based work*. Lawrence Erlbaum Associates.

Walker, G. (Ed.). (2009). *Command and control: the sociotechnical perspective*. Ashgate.

Walker, G. H., Stanton, N. A., Salmon, P. M., Jenkins, D. P., & Rafferty, L. (2010). Translating concepts of complexity to the field of ergonomics. *Ergonomics*, *53*(10), 1175–1186. https://doi.org/10.1080/00140139.2010.513453.

Walker, G. H., Stanton, N., Salmon, Paul M., & Jenkins, D. P. (2018). *Command and control: the sociotechnical perspective*.

Zaccaro, S. J., Marks, M. A., & DeChurch, L. (Eds.). (2012). *Multiteam Systems* (ed.). Routledge. https://doi.org/10.4324/9780203814772.