Games to support disruptive technology adoption: the MUST Game use case

Francesca de Rosa[1] and Christopher Strode[1]
[1]*NATO STO Centre for Maritime Research and Experimentation,*
francesca.derosa@cmre.nato.int, christopher.strode@cmre.nato.int

Abstract

Serious games can be used as a means to explore complex systems and uncertainty related challenges, therefore they may have the potential of supporting the adoption of innovative and disruptive technologies. In this paper we present the use case of the Maritime Unmanned Systems Trust (MUST) Game, which goal is to capture beliefs, attitude and perspectives of the participants with respect to the employment of maritime unmanned systems (MUS) in the maritime domain. This novel game aims at better understanding the relation between trust factors and MUS. Moreover, it explores how players make decisions with respect to MUS deployments in an increasing threat scenario. This allows to capture important information on the trade-offs related to MUS use that have an impact on maritime missions planning activities (e.g., endurance, logistics, maintenance, cost, number of assets, security and type of assets). This paper describes the game and an analysis of the outcomes of its deployment. The results show how the MUST Game design has been effective in eliciting constructive discussion around the use of MUS in maritime missions, as well as in the collection of assessments and decisions, which are currently being used in algorithmic development.

Keywords: Analytical wargame, Maritime Unmanned System, Maritime operations, Trust, Deployment optimisation factor, Knowledge Acquisition Analytical Games

1 Introduction

Games not for entertainment, to which we will refer simply as games in the remainder of the paper, can serve different purposes. Regardless of the game sub-discipline (i.e., wargaming, serious gaming or simulation gaming) the main purpose of a game can be training, message broadcasting and data collection [1]. The data collection games aim at supporting data exchange, such as “collecting information from [...] players” [1], and the creation of knowledge about a topic of interest. They can be further distinguished in discovery games and analytical games, if they look at an unstructured problem or a structured problem respectively [2]. Analytical games are games designed to research a problem [2] and can be used as a means to explore complex systems and the inherent accompanying uncertainty. In this paper we present a use case on how to leverage this intrinsic characteristic of games to support the adoption of disruptive technologies and to design decision support systems (DSSs) associated to their use. The development of emerging and disruptive technologies (EDTs), such as artificial intelligence and autonomy, is growing in every domain and the issues of trust is an overarching problem that might hinder their adoption. While the topics and approaches discussed in this paper are applicable in most domains, we will contextualise the discussion into the maritime one. Specifically, we will present a game that explores the problem of employing maritime unmanned systems (MUS) into maritime operations. In fact, MUS are moving towards higher levels of technological maturity, however, it appears that within the maritime security and defence domain the doctrinal development that would allow their use is still lacking [3]. Although there is not a unique definition, MUS can be defined as an "unmanned system operating in the maritime environment (subsurface, surface, air) whose primary component is at least one unmanned vehicle. An unmanned vehicle is a [...] vehicle that does not carry a human operator and can: a. be operated autonomously or remotely; b. be expendable or recoverable; and c. can carry lethal or non-lethal payloads" [4] (e.g., sensors).

Building trust in MUS, through demonstration and progressive integration of MUS into training and operations, is highlighted as the main effort in the near future [3]. However, trust in automation and autonomy is an important and complex mental construct, that goes beyond mere familiarity and exposure to a certain technology. Other underpinning factors might play an important role (Section 2.2). Therefore, a comprehensive approach is needed to ensure an appropriate understanding of such a construct and its operationalisation.

With the goal of better understanding the relation between trust and MUS, contributing to the human-system integration efforts in this domain, the authors have developed the Maritime Unmanned Systems Trust (MUST) Game. This is a knowledge acquisition analytical game [5] that aims at capturing beliefs, attitude and perspectives of the participants with respect to the employment of MUS. The adoption of MUS in missions could prove to be effective. In fact, this could lead to an enhancement of situational awareness, reduction of human workload and enhancement of the operational performances. These systems would complement the manned platforms, positively impacting persistence, versatility, survivability, risk reduction and cost reductions [3]. However, many are the elements that decision makers will have to trade-off [6] at different levels (i.e., strategic, operational and tactical). Therefore, while focusing on trust in MUS, the MUST Game ensures the collection of relevant information also in relation to other mission planning factors that are key for the future DSSs, such as endurance, covertness, logistics, maintenance and cost. Moreover, the proposed operational analysis approach seeks to employ gaming methods to explore the real world complexities of large scale deployments of unmanned assets in the maritime. This work describes the game design features as well as the outcomes of the three Table Top Exercises (TTXs) that employed the MUST Game. Specifically, the remainder of this paper is organised as follows: Section 2 provides a brief overview of the use of games for technology investigation and introduces the concept of trust in autonomy; Section 3 details the design and the components of the MUST Game; Section 4 reports on an analysis of the collected data; and Section 5 summarises the lessons learned and future work.

2 Background

2.1 The game approach

The MUST Game is an analytical wargame based on the Knowledge Acquisition Analytical Game (K2AG) approach [5]. K2AGs are games used as knowledge acquisition (KA) tool. Knowledge acquisition refers to the extraction, structuring and organisation of expert knowledge to be encoded in algorithms and intelligent systems. In the last two decades the potential use of games for KA has started to gain attention. For example, MovIE WIzard, Book Wizad and MovIE Gurus [7] aim at using human computing to discover in text narratives relations between entities, which are hard-to-extract automatically. The SpotTheLink game [8], instead, was developed to motivate users in ontology alignment related tasks. Furthermore, the OntoGame [9] is a game framework which has the objective of deriving best practices and guidelines for semantic-content-authoring technologies [8]. Moreover, several methods have been designed to take advantage of games to collect useful information [10]. Examples include games looking at image annotation (e.g., [11], [12]), semantic web (e.g., [9], [13]) and commonsense KA (e.g., [14], [15]). A literature search shows that the work on games focusing on human decision-making to support the design of DSSs is limited. In fact, the concept of using games-for-modelling for the acquisition of highly structured domain-specific knowledge to be used in model-based methods for artificial intelligence (AI) appears in [16]. However, this work presents generic and preliminary results and it seems that little efforts have been devoted to formalise the approaches. Also the use of games to explore the trust construct appear to be scarce and not focused on innovative technologies and algorithmic design, but rather on behavioural aspects. In fact, they focus mainly on the collection of trust information datasets in the context of supply chains (e.g., [17–19]). Previously successfully deployed instances of K2AG include the Reliability Game [20, 21] and the MARISA Game [22], which data has been succesfully used to design algorithms to be employed in intelligent systems on the basis of cognitive mimetic principles [23]. K2AGs have proven to be very efficient and effective in terms of knowledge elicitation (i.e. time reduction, experiment simplicity and ability to extract the required qualitative and quantitative knowledge). Two main gaming aspects that characterise K2AGs are the use of knowledge cards (KCs) to render information and meta-information to the players and the use of innovative data gathering methods to easily collect players beliefs. These elements are at the core of the MUST Game mechanics as well. The MUST Game is a new K2AG which aims at exploring different mental constructs (i.e., trust) compared to previous K2AGs. Moreover, its goal is to inform the design of decision-support and planning tools, rather than the fusion algorithms computational strategies. This game combines the strengths of K2AG with elements typical of Disruptive Technology Assessment Games [24], such as cards containing information regarding new technologies. A Disruptive Technology Assessment Game (DTAG) is a table-top seminar game aiming at assessing potential disruptive technologies, through the assessment of their impact on the military operations and operational environments. DTAG makes use of Ideas of System (IoS) cards that include information regarding new systems (e.g., new technologies combined with equipment). New courses of actions are planned by the teams on the basis of the IoS provided and a technology led confrontation concludes the rounds. This method has proven to be an efficient tool when assessing early technology concepts and prototypes that have not yet been used in military operations [24]. However, they are a means to explore a concept, rather than collecting data that could be modelled in further knowledge engineering efforts. Differently than the previous K2AGs, the MUST Game is designed to collect data on a wider number of variables present in the conceptual model of the game, as well as capturing high level considerations on the use of the new family of systems under examination. Therefore, it enables to explore the depth and the breath of the challenges of employing disruptive technologies.

2.2 *Trust in autonomy*

It has been demonstrated how trust is an important mediating factor with respect to the employment of specific technologies, especially in high-risk situations [25]. In fact, low levels of trust might affect the willingness of humans to rely on information and suggestions provided by the technology, while unreasonable high levels of trust might lead to overreliance and misuse [26].

Trust can be described as a tripartite relational property [27]. The first one identifies a component that is stable over time and represents the long-term propensity to trust of an individual. The second component is a state expression of trust, which is mainly a response to ambient conditions and is volatile. Finally, the third component represent the dynamic nature of trust, describing how trust evolves over time (i.e., trust development and loss). Recent meta-analysis in the domain of Human-Robot Interaction (HRI) [26] and Human-Automation

Interaction (HAI) [27] have investigated the factors influencing the development of trust in automation and autonomy. Automation and autonomy are two distinct concepts. Automation refers to “technique[s], methods[s] or system[s] of operating or controlling a process by [...] automatic means, as by electronic devices, reducing human intervention to a minimum” [28], while autonomy refer to the ability of machines (hardware and software) to perform independently under significant uncertain conditions for extended periods without external intervention [29]. Autonomous systems are able to operate with limited or non-existent communication and have the ability to compensate for system failures. Robots in these studies are treated as a subcategory of automated or autonomous systems that mainly differs from other systems in the fact that they generally are mobile, affect actions and might have an anthropomorphic appearance [26]. The meta-analysis on trust in HAI lead to a reorganisation of the moderating and antecedent factors to standardise the proposed model along different types of partner technology [27]. However, in our study we refer to the original model developed on HRI studies. This choice is not only correlated with the fact that HRI is the relevant HAI subset to contextualise the interaction between operators and MUS, but also with the specific factors that are included in the two proposed models. In fact, the way in which the factors are clustered appear to be more intuitive to be discussed with the MUST Game participants. For example, the system-related factors in the HRI framework explicitly refers to system failure rate, system false alarms and other factors that are of paramount importance in the use of MUS in missions. The meta-analysis on trust in HRI and HAI were driven by the assumption that introducing robotic and autonomous systems into human systems might not always result in improved team performance. This led to the need of an analysis of the role played by trust in this balance. The three-factor model of human-robot trust has identified several underpinning factors (i.e., antecedents of trust) that pertain to three main categories and analysed how they influence trust development. Specifically, such antecedents can be human-related, system-related or environmental. Human-related factors refer to either human characteristics (i.e., demographics, personality traits, attitude towards the systems, comfort with the system, self-confidence and propensity to trust) or ability-based aspects (i.e, attention capacity or engagement, expertise, competency, user workload, prior experience and user situational awareness). System related factors include performance-based and attribute-based elements. The performance-based ones are system behaviour, dependability, reliability, predictability, level of automation, failure rate, false alarms and transparency. Instead, attribute-based elements include proximity or co-location of system and user, system personality, adaptability, system type and anthropomorphism. Finally, environmental factors are divided into team collaboration factors (i.e., in-group membership, culture, communication, shared mental models) and tasking (i.e., task type, task complexity, multi-tasking requirements and the physical environment). Figure 1 illustrates the different factors and their categories. Some trust factors (i.e., in-group membership, culture, communication, shared mental models, system personality and anthropomorphism) were not included in the current study, as they were regarded as not applicable to the issue under investigation. In fact, the systems under investigation do act as tools rather than human-like teammates. The factors not included are depicted with a dotted contour. Many studies have focused on technology related aspects of MUS (e.g., [30]), however in this study we focus rather on the human assessment regarding MUS, to better understand the foreseen benefits, potential issues in technology uptake by operators and which factors are important in the decisions to deploy manned, mixed or fully autonomous forces. In order to provide the basis for a valuable and wide discussion on the use of MUS in maritime operations, within the MUST Game the set of factors currently employed in the optimisation algorithms (i.e., cost, number and variety) has been enriched with the set of trust factors (Section 2.2) applicable to the MUS. This information are expected to inform the next generation of decision support systems available to operators.

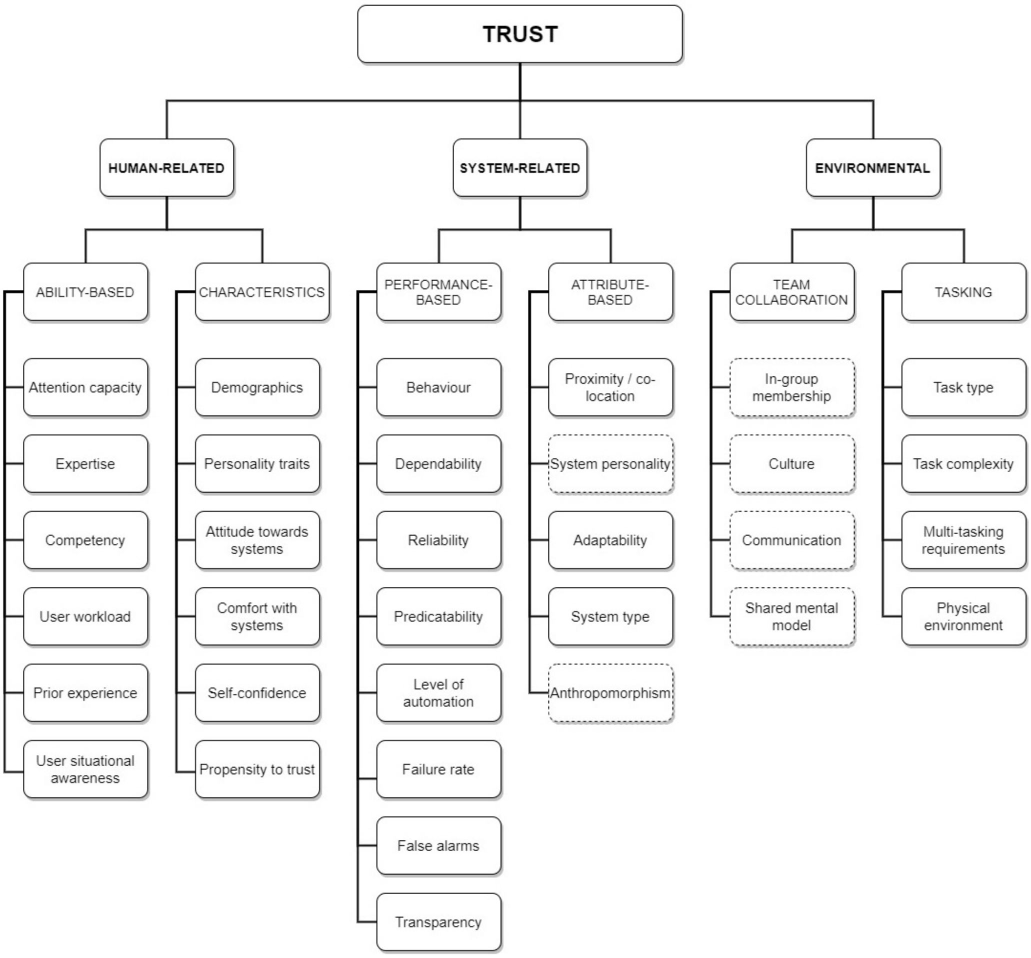

Figure 1: *Three factor model of trust for HRI [26]. The factors with a dotted contour line are not considered in the analysis conducted through the MUST Game.*

3 The MUST Game

3.1 The game purpose

The MUST Game is an analytical wargame, aiming at collecting useful information to support the development of future maritime Concept of Operations and the continuous development of decision support tools that will allow decision makers to fully take advantage of the use of MUS within maritime missions. It is not meant to measure subjective assessment of trust (e.g., through indirect measurements on constructs such as reliance), as that would require a different experimental setting than the one proposed by such kind of games. Instead, it focuses on providing an analysis regarding underpinning factors contributing to trust in a specific domain and mission type to drive future research and development. Moreover, it has a message broadcasting secondary purpose. In fact, it aims at discussing how useful advanced decision support tools might be with respect to these operations. The MUST Game has been deployed both as manual and as a distributed computer-assisted wargame.

3.2 Game design

3.2.1 Design approach

The MUST Game design followed the analytical game design framework proposed in [31]. Specifically, it followed an iterative process that led to the definition of a game conceptual model, a game design model and two game implementation models. One implementation model refers to the analog version of the game, while the second model refers to the computer assisted distributed game version. The game development started with a quick prototyping approach, followed by game verification activities that took place through a series of short play-testing sessions. The testing has been supported by experts in the field of maritime operations. The next sections describe the design constraints (Section 3.2.2), the world design (Section 3.2.3), the system design (Section 3.2.4) and the content design (Section 3.2.5), providing the relevant details of the aforementioned game models. Specifically, the world design refers to "the creation of the overall backstory, setting and theme" [32], while the system design corresponds to the "creation of rules and underlying mathematical patterns" [32]. Finally, the term content design refers to the "creation of characters, items, puzzles and missions" [32].

3.2.2 Design constraints

The overall objective was to design a game considering a large area of operations. The basic assumption being that the area of interest has to be surveilled through sensing capabilities in order to detect the transit of one or more threats with a reasonable degree of certainty. The scenario had to be tailored in such a way that manned assets are not always readily available or not able to conduct the required mission for the necessary duration. Therefore, players have to deploy autonomous platforms in order to augment the limited manned coverage available. Moreover, the injects are designed in order to induce an initial threat escalation, followed by the requirement to reduce force escalation. This design choice was driven by the interest in exploring how changing threat levels could impact the player decisions with respect to MUS deployment.

Although environmental conditions are important factors in maritime operations and MUS use, the game only partially explored the topic through the use of a KC related to meteorological and oceanographic (METOC) conditions, mainly because of time constraints on the game session length. However, minor changes to the information contained in the KCs could be introduced to focus on METOC in future MUST Game deployments. In fact, this could include decreased detection range assumptions on the basis of worsening underwater conditions.

The COVID-19 pandemic crisis introduced a strong uncertainty with respect to the number, background and location of the MUST Game players. This called for a highly flexible and modular approach in the game design. Therefore, a modular approach was adopted (Section 3.2.4) that allows to adapt to the change in number of players, while the scenario is quite generic in order to adapt to the different expertise and experience of the participants. Specifically, the players are free to focus more on tactical/operational aspects or on strategic ones and the game will support this discussions allowing to capture interesting perspectives (i.e., on MUS trust) on the use of unmanned systems at different levels.

From a platform implementation perspective in a first instance it has been decided to develop the MUST Game in an analog form. This design choice was based on considerations that include, but are not limited to: (i) the kind of information to be collected; (ii) the kind of game experience provided to the players, (iii) the resources available and (iv) the limited pre-event access to the original TTX facility. COVID-19 break-out lead to the need for adjusting the game elements design to guarantee compliance with health and safety measures in place (e.g., personal separation and no exchange of game items between players). Moreover, the

Figure 2: *MUST Game map.*

persistence of the restrictions have led to the development of a computerised facilitation tool that allows the deployment of the MUST Game in a distributed fashion.

3.2.3 World design

The MUST Game scenario is strongly inspired from the High North Matrix Game [33] and the Bear Rising Matrix Game [34] scenarios, but it focuses on the maritime component. It presents a fictitious geographical area including the following countries: Grey Land, Yellow Land, White Land, Orange Land, Green Land and Red Land. Figure 2 presents the map of the area, with the geographical location of different states. As highlighted on the map the White Island - Middle Island gap, the Middle Island - South Island gap and the South Island - Green Land gap are 374 NM, 410 NM and 215 NM respectively.

Middle Island is under the sovereignty of Yellow Land. Grey Land, instead, is located outside the map on the left. The scenario starts with Grey Land policy-makers assessing collective defence arrangements with their historical allies (Yellow Land, White Land and Orange Land), in response to Red Land increasing aggressive military posture in the last years. Moreover, Red Land ships and submarines originating from the Red Land Naval Base and sailing into the Big Ocean are posing a serious threat to the allies' interests. In fact, they are conducting reconnaissance of allies naval bases and of transoceanic underwater cables.

While returning from an exercise, a White Land military vessel crashes into a Red Land fishing vessel and the incident results in the sinking of the fishing vessel, several deaths and casualties. Red Land declares that the allies are threatening its Sea Lines of Communication and there is evidence that they might deploy troops and coastal defense missiles to the Small Island, which is a demilitarised area, by virtue of a specific treaty.

The player has the role of a Grey Land military decision-maker that as part of the alliance is requested to assess on the basis of the available information the threat level and to decide whether to deploy or not a surveillance barrier. The barrier could be composed by manned assets, MUS or a mixture of manned and unmanned assets collaborating to support the maritime surveillance task. Moreover, the participant is requested to propose a possible composition of the deployed forces, choosing from several available assets, and to explain the reasons for the decisions made.

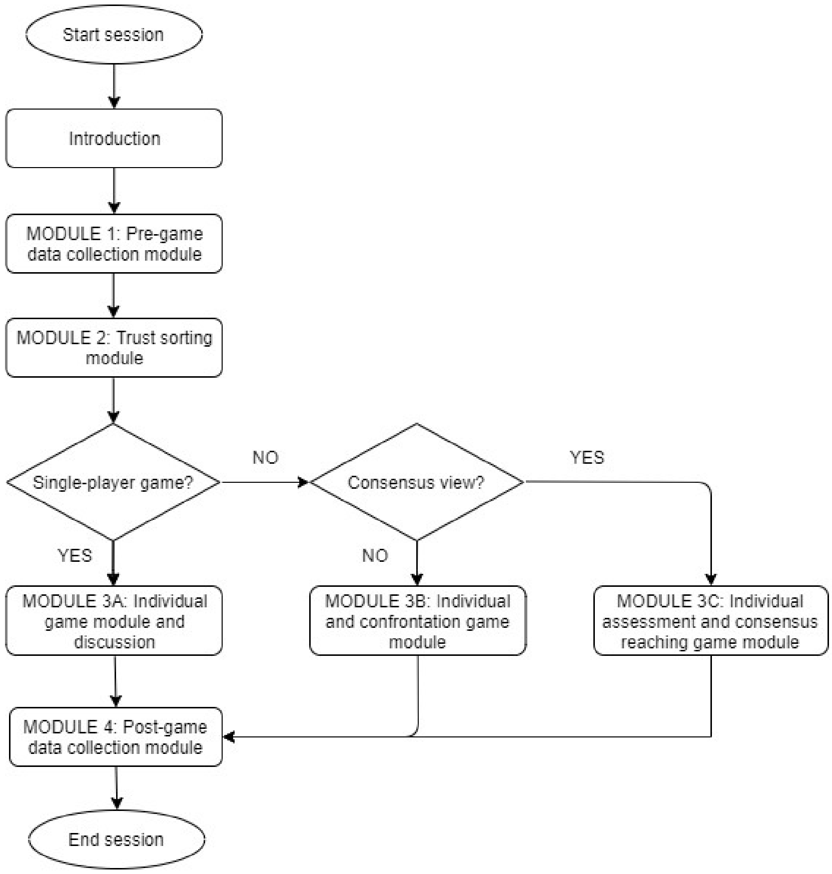

Figure 3: *MUST Game diagram*

3.2.4 System design

The MUST Game has been developed following a modular approach in order to assure the required flexibility to meet the challenges described in Section 3.2.2. A graphical representation of the different phases is depicted in Figure 3, where we can observe how the modular approach has been implemented to address the game design constraints (Section 3.2.2).

At the start of the game session the player is introduced by a facilitator to the game objectives and motivations. After this introduction the player is required to fill in a pre-game questionnaire (Module 1) that allows to collect not only demographic data, but also useful information with respect to the participant personality traits and characteristics. This information helps characterising the player, hence supports better data interpretation. Specifically, for the MUST Game we focused on the attitude towards new technology and the natural tendency to engage in effortful thinking rather than heuristic thinking [35]. To determine these aspects the Affinity to Technology (ATI) [36] and the short Need for Cognition (NFC) [37, 38] questionnaires have been used respectively. The short NFC questionnaire is an 18-item questionnaire designed to measure NFC, which is a stable individual difference (i.e., a personality trait) that indicates the "tendency to engage in and enjoy effortful cognitive endeavors" [38]. NFC refers not to the individual's cognitive ability of thinking, but rather to the motivation to think, with limited prompts. The motivation, in fact, varies along a continuum from individuals that find satisfaction in thinking (i.e., high NFC) and individuals who engage in thoughtful and effortful thinking only if incentivised [35]. Psychology research has widely investigated the link between NFC and individual's behaviors and attitude, therefore, the reader is referred to the relevant literature for additional details. The ATI 9-item questionnaire is used to assess an individual's tendency to actively engage in intensive technology interaction. ATI is a key facet of user characteristics used in research on technology interaction [36]. For example, it supports general research models of user-technology interaction, user experience and technology acceptance. In this game we propose to measure ATI as a means to characterise the players of this analytical game which is specifically designed to assess which and how new technologies might be employed in maritime operations. In fact, it is assumed that this personality trait might impact the data collected from the participants. Therefore, the measurement of ATI could support the analysis of such data. NFC and ATI are consider stable individual constructs to understand better the characteristic of the player, but do not measure trust in unmanned systems. In fact, the specific construct of trust is explored through the mini-game and main game.

Once this first data collection has been concluded participants perform a mini-game with a set of trust factors cards that report details on the trust factors considered into the MUST Game (Section 2.2). The mini-game is a gamification of a card sorting exercise (Module 2) where participants ordinal-rank these factors from the most important to the least important when it comes to trust in MUS. This exercise acts both as a data collection module, but also as a familiarisation exercise that allows the player to get exposed to the trust factors and start discussing about the topic. Then the players are introduced to the game core, game rules and game scenario (Section 3.2.3).

Each game session is composed of one round in which nine knowledge cards (KCs) are provided one at the time to the player. Each time the player receives a new knowledge card, the amount of information regarding the situation available to the players is updated. Therefore, they are requested to repeat the series of assessment regarding: (i) the threat level, (ii) the decision on the barrier deployment, (iii) the decision on its compositions and (iv) the factors that played a major role in this decision.

If the game is played in single-player mode, the participants are requested to fill in the data recording sheet (see Section 3.2.5) for the corresponding knowledge card and are encouraged by the facilitator to share all the relevant information related to their decisions (Module 3A). Instead, if the game is played in a multi-player mode, the participants can play either individually or as a team. In the first case they will be requested to compile the data recording sheet for each knowledge card and take notes regarding their decisions, to be discussed later in a seminar like setup after the last recording sheet has been compiled by all players (Module 3B). In the second case they will discuss each knowledge card together and will provide a consolidated answer for each knowledge card once consensus in the team has been achieved (Module 3C). The selection of the multi-player mode relates to the data collection and research objectives. In fact, a confrontation of the different perspectives of the players during (or soon after) each assessment phase might influence the decision-making cycle of the players in the next iteration. Therefore, if the focus is on the single player reasoning, rather than on the consensus assessment the individual mode should be adopted (Module 3B), alternatevely the consensus mode should be selected (Module 3C).

As for other K2AGs the KCs convey a message (M). However, differently than in other K2AGs the knowledge cards contained only an information (I), while no additional meta-information about uncertainty related factors was provided (i.e., source reliability, information credibility, information trueness, information precision). Sources of information are assumed reliable for the purpose of this game and the information provided through the knowledge cards is always true. This design decision was driven by the need of collecting data related to a wide space of variables without the additional layer of complexity brought by information uncertainty factors. Future work could explore how additional degrees of uncertainty could

Variable	Description	Frame
$Threat$	Threat level	{High, Medium, Low}
$DepProp$	Deployment propensity	{Deploy, Not Deploy}
$DepDec$	Deployment decision	{Deploy, Not Deploy}
$ForceType$	Force type	{Manned, MUS, Mixed}
$ForceComp$	Force Composition	{Frigate, Replenishment, Support platform, Destroyer, Large displacement UUV, Medium displacement UUV, Small displacement UUV, Large USV, Medium USV, Underwater glider, Wave glider}
POF	Primary optimisation factors	{Cost, Variety of assets, Number of assets }
DF	Additional decision factors	{Attention capacity, Expertise, Competency, User workload, Prior experience, User situational awareness, System behavious, Dependability, Reliability, Predictability, Level of Automation, Failure rate, False alarms, Transparency, Co-location of system and user, System personality, Adaptability, System type, Task type, Task complexity, Multi-tasking requirements, Physical environment, Safety, Security, Endurance, Deployment time}
M	Message conveyed by a card	$\{M_1, \ldots, M_k\}$
I	Information conveyed by a card	{Relocation of forces, Allies assets, Convoy plans, Threat detection, Intelligence report on threat, METOC forecast, Order of force escalation reduction, Plans of future convoy, Prediction of threat position}
IT	Information trueness	{True}
SR	Source reliability	{Reliable}

Table 1: *MUST Game state*

Variable	Description	Frame
$Threat$	Threat level	Assessed[1]
$DepProp$	Deployment options	Assessed
$DepDec$	Deployment decision	Assessed
$ForceType$	Force type	Assessed
$ForceComp$	Force Composition	Assessed
POF	Primary optimisation factors	Assessed
DF	Additional decision factors	Assessed
M	Message conveyed by a card	Provided[2] for M_1 to M_n Not Provided for M_{n+1} to M_k
I	Information conveyed by a card	Provided for M_1 to M_n Not Provided for M_{n+1} to M_k
IT	Information trueness	Provided
SR	Source reliability	Provided

[1] Assessed = player has to assess the item and communicate it to the facilitator;
[2] Provided = item value provided to the player; Not Provided = item value not provided to the player;

Table 2: *MUST Game view for game iteration n*

impact the player assessment. Table 1 summarises the relevant variables of the game model that might change during the gameplay, known as game state [32]. Let us define as n each iteration within the individual modules (Module 3A, Module 3B or Module 3C). An iteration

Figure 4: *Example of a trust factors card (green) and a knowledge card (blue).*

corresponds to the injection of a new knowledge card, followed by a new player assessment phase. Therefore, $n = 1, \ldots, k$ where k is the maximum number of knowledge cards. The portion of the game state that is visible to the player in each n, known as game view [32], remains unaltered with the exception of the one referring to the knowledge card (M and I). The game view is summarised in Table 2. The sequence in which the knowledge cards are presented is kept constant to allow a comparison across the different participants.

At each iteration n the players have to record the outcomes of their assessments on the data recording sheet and explain the reasons behind the assessments, either while performing the assessment or in the confrontation module.

Finally the players are requested to compile the post-game questionnaire (Module 4), which aims at collecting general feedback as well as data regarding the player experience, usability of the game as a tool and players workload. The post-game questionnaire makes use of a mixture of validated methods. Specifically, it employs the MEEGA+, which is an update to the Model for the Evaluation of Educational Games (MEEGA) [39], the Questionnaire for User Interaction Satisfaction (QUIS) [40] and the NASA Task Load Index (NASA-TLX) [41].

3.2.5 Content design

The participants all have their own game set, which is composed by: (i) a set of trust factors cards; (ii) a set of k maps; (iii) a set of k knowledge cards; (iv) a set of k data recording sheets; (v) a set of unmanned asset cards; (vi) a set of manned asset cards; (vii) a set of decision support cards.

The trust factor cards (Figure 4), which are used in the card sorting exercise, include a trust factor each and the additional information regarding the element to which they relate (i.e., human-related, robot-related or environmental) and the type factor (i.e., ability based, performance based, attribute based and tasking) [26]. The card sorting exercise is used instead of a simple questionnaire, as element sorting is easier than numerically ranking twenty-one elements in a list.

Differently than in many manual games, where a map is provided and counters and tokens are positioned on it during gameplay, due to COVID-19 restrictions and the need to minimise objects touched by different persons, each participant has a set of maps printed on paper. Every time a new knowledge card is received the player is allowed to see the following map, which has increasing information displayed on it, that refer to the new information received through the card. The maps can also be used by the players to annotate relevant information that pertains to the specific game iteration (i.e., asset operating areas). A set of manned assets cards and a set of unmanned assets cards (Figure 5) are provided to support players

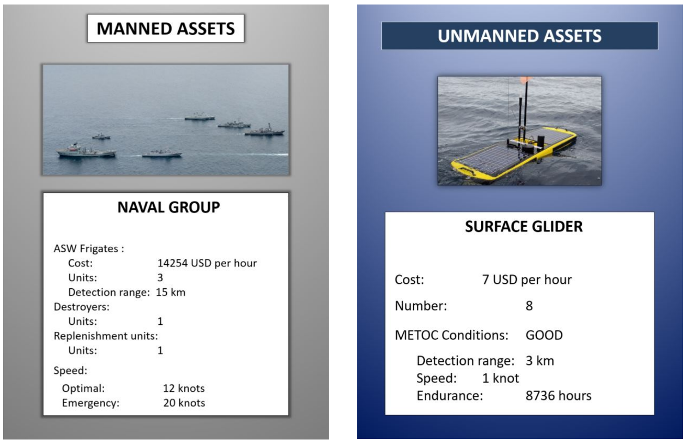

Figure 5: *Example of MUST Game manned and unmanned assets cards*

assessments. These cards report high-level information regarding the different assets, such as numbers available and performances (i.e., speed, endurance and detection range).

Once a knowledge card (Figure 4) is received and the corresponding map is accessed by the players, they need to start their individual set of assessments and decisions. These are self-recorded on data collection sheets, which are depicted in Figure 6. The data recording sheets are divided in different areas that somewhat correspond to the decision-making steps performed by the players along the game. Specifically, the areas are:

D1 - Threat level: the threat assessment area, where players record their belief regarding the threat level (i.e., high, medium and low);

D2 - Deployment: the deployment decision area, where players record their propensity towards one of the two decision options (i.e., deploying assets and not deploying assets) and the decision in that specific iteration;

D3- Asset type: the force type assessment area, where players record their belief regarding which force type could be more appropriate in the current situation (i.e., manned, MUS or mixed);

D4 - Asset selection: the force composition assessment area, where players record the number of assets per type they would deploy; their position can be marked on the corresponding map;

D5 - Optimisation: the optimisation assessment area, where the players record whether in their decision regarding the deployment they were trying to minimise one or more of the decision factors listed (i.e., cost, asset variety and asset numbers);

D6 - Factors: the deployment decision factors area, where the players state if there are additional factors (i.e., the trust factors, safety, security, endurance and covertness) that played a major role in the previous decisions;

D7 - Comments: the comment area, where players can take notes.

To play the MUST Game in a distributed fashion a facilitation tool has been designed that allowed to visualise all the game elements (e.g., maps and knowledge cards) at the right step of the game, which was facilitated online (Figure 7) via commercial video conferencing platforms.

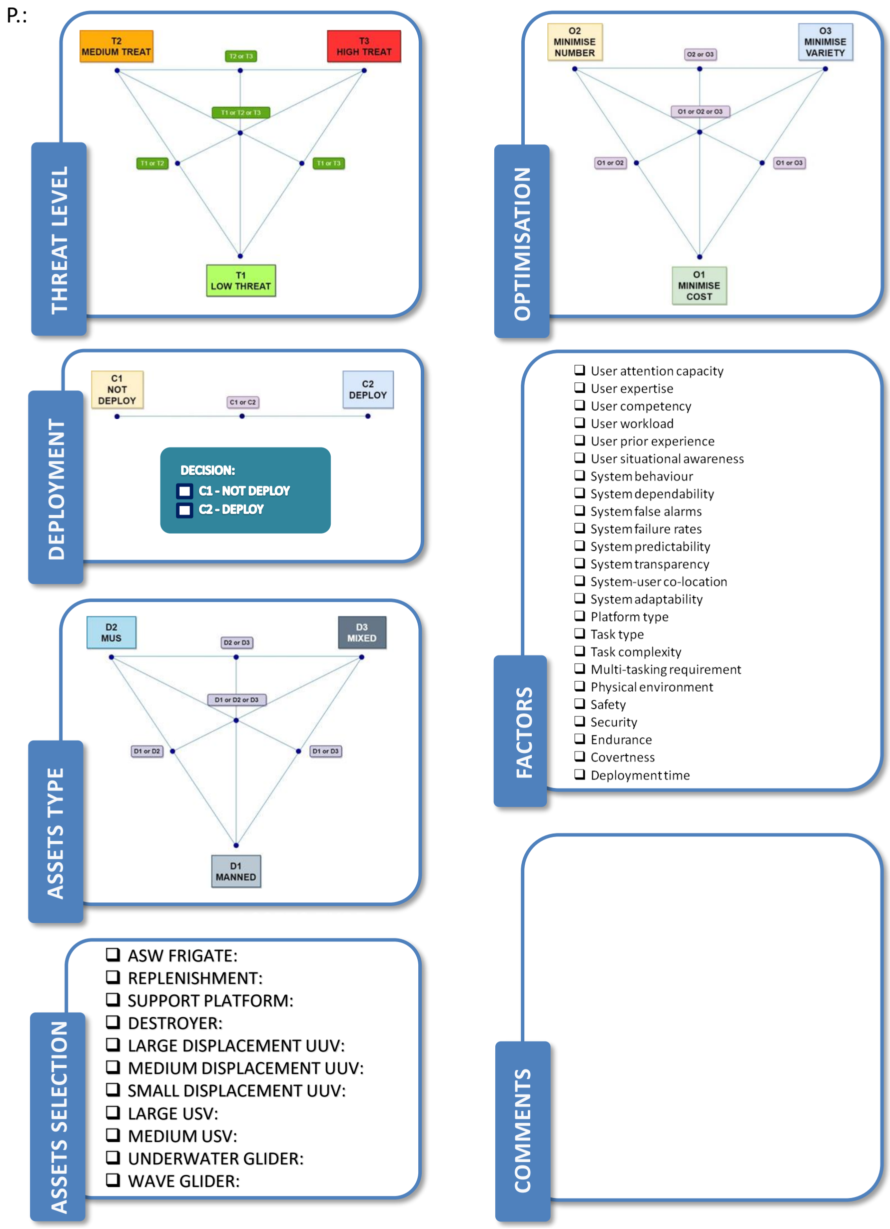

Figure 6: *MUST Game data recording sheets*

Figure 7: *The MUST Game Distributed Facilitation Tool main screen and gaming screen*

Parameter	Manual game sessions	I distributed TTX	II distributed TTX
Participants number	7	5	18
Gender			
Male	57%	80%	89%
Female	43%	20%	11%
Age			
Average	47.6 years	-	45.6
Standard dev.	11.4 years	-	10.4
Status			
Military	50%	100%	72%
Civilians	50%	0%	28% [1]
Nationality			
Australia	0%	0%	6%
Belgium	0%	0%	6%
France	0%	0%	6%
Germany	0%	20%	6%
Italy	47%	0%	6%
Portugal	0%	0%	11%
Spain	0%	80%	0%
The Netherlands	0%	0%	6%
United Kingdom	0%	0%	33%
United States	43%	0%	22%
Year of relevant working experience			
Average	24.7 years	-	21.7
Standard dev.	11.2 years	-	10.9

[1] Most civilian participants had a military background

Table 3: *MUST Game participants demographic information (for the distributed session some demographic information has not been provided)*

4 Playing the MUST Game

4.1 The Table Top Exercises

The pandemic crisis impacted the game design as well as the game deployment. Therefore, the TTX was performed as a distributed activity, where the term distributed does refer both to time and location. Specifically, several game sessions have been played over a time frame of eight months, when circumstances allowed for it. In the first TTX, the manual version of the MUST Game has been played in a series of game sessions with the participation of a small sample of players with relevant expertise in the field. The players are either military or scientists invited to participate because of their background in maritime operations and autonomy. One game session involved two players, while the remaining ones were single-player sessions. In most cases the single-player mode was selected, as a consequence of the reduced physical attendance of personnel at the game premises in response to COVID-19 limitations.

The computer assisted MUST Game, instead, was used to run two distributed TTXs. The first TTX took place with participants from NATO standing maritime forces as part of a wider military exercise. A second distributed TTX has been organised in which participants were appointed by several national and NATO delegations to the NATO MUS Initiative in response to a NATO calling notice. Table 3 summarises the demographic information of the players. Although there are no specific constraints on the number of players that can participate in a multi-player session, in the second distributed TTX players were divided into four groups. In fact, we wanted to ensure the engagement of each player without periods of inactivity and to limit the session length at maximum four hours.

It should be noted that during the first computer assisted TTX the participants were requested to play individually, however, they strongly insisted to play as a team and because of the flexibility of the MUST Game design this was allowed. Therefore, although we have been able to capture distinct aspects of the individual assessments while they were talking and

Frequency of play	Digital games	Manual games
Never	33%	17%
Rarely: from time to time	44%	50%
Monthly: at least once a month	6%	17%
Weekly: at least once a week	17%	17%
Daily: every day	0%	0%

Table 4: *MUST Game participant gaming expertise characterisation*

reaching a consensus regarding the answer to be provided, the data collected in this session is treated as the one provided by a single participant in the next sections.

In addition to the demographic information, pre-game data has been collected that allowed to characterise the player with respect to their propensity towards playing games, their affinity to technology and their need for cognition. The data on the frequency at which participants play games showed that they are not often playing games (Table 4). This information is important as it might give useful insight on the interpretation of the usability and PX data collected. Moreover, it might help the facilitator in characterising the player gaming expertise and to adapt the facilitation style.

With respect to the other data collected, it has been possible to observe that all the participant (except one) have shown a high level of affinity to technology (average $M_{ATI} = 4.15$ out of 6, standard deviation $SD_{ATI} = 0.24$). This indicates that the participants have a natural tendency towards the interaction with technology (i.e., explore new technical systems).

Finally, all the participants exhibit a high level of NFC. A high level of NFC has been demonstrated as being positively correlated with engagement in thinking with little prompting, higher likelihood of careful and systematic processing of information (e.g., [37, 38]) and openness to experiences (e.g., [42]). These aspects play an important role on the attitude of the participant to the game, on the game facilitation and the in-game data collected.

4.2 Post-game data and game validation

Post-game data has been collected in order to support the validation activity of the MUST Game, as well as collecting general feedback from the participants. Figure 8 shows the results of the answers of the players on several questions related to the usefulness of decision support tools (DDSs) for the use of MUS and the relevance of the MUST Game. The answers are provided on a five-point Likert scale ranging from *not at all* to *extremely*. We can easily see how the players regarded the availability of DSSs as highly useful (i.e., 86% extremely useful and 14% very much useful) as the tasks to be performed are difficult to plan (i.e., 32% extremely difficult, 37% very much difficult and 31% moderately difficult). The MUST Game is considered sufficiently realistic (i.e., 5% extremely realistic, 47% very much realistic and 42% moderately realistic) and its purpose was easily understood by the players (i.e., 11% extremely well understood, 74% very much well understood and 11% moderately well understood). Moreover, the players considered the topics explored with the MUST Game as operationally relevant (i.e., 37% extremely relevant, 53% very much relevant and 10% moderately relevant). The percentages reported account only for the compiled questionnaires received. In fact, some participants did not provide the feedback through the questionnaire, but rather in free text format. Nevertheless, their qualitative assessments aligned with the one provided in the feedback questionnaire.

The results of the questionnaire regarding player experience (PX) are summarised in Table 5. They show an overall good level of PX. In fact, we can see how the game as an artifact was considered adequate (e.g., game element design, colors, fonts, content), especially in the manual version. The game was perceived as easy to learn, engaging, appropriately challenging, not monotonous and governed by clear and easy rules. As expected, the game did not

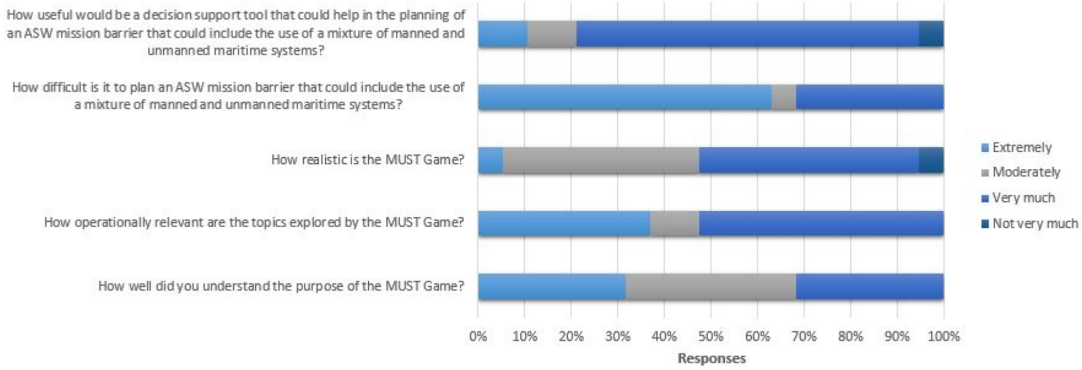

Figure 8: *Overall feedback provided by the players*

provide a strong immersive experience, as that was not the aim of this analytical wargame. In some cases the game scored low with respect to the social interactions components. This last point is due to the fact that in most sessions of the first TTX the game was administered as a single-player game. However, a very positive feedback on the social interaction aspect was provided by the players of the distributed computer-assisted game version (second and third TTX). It was highlighted that this experience provided them an important and quick training opportunity, which allowed to confront themselves with colleagues with different nationality and backgrounds. In fact, the need to reach a consensus starting from different perspectives was considered the first time in which they had the opportunity to interact in this manner with the other participants in a safe-to-fail environment. This feedback supports the perspective expressed in [43] which consider every wargame, including analytical wargames, as having an intrinsic training and educational component.

The method proposed was very well perceived by the participants, which judged it as valuable, thought provoking and encouraging “outside-the-box” thinking. It was stated that although wargaming was not new to them, the proposed approach was more effective and efficient as it was “simple, quick and strait to the point”. The players had fun, were satisfied by the experience and would recommend it to colleagues. Moreover, they perceived the game as an adequate knowledge acquisition method. Overall they appear to prefer gaming over other means as an elicitation mean (i.e., questionnaires). However, it appears that not all the players clearly understood how the content of the game and the stated objective are related. This actually supports the fact that the game is a useful and easy way to interact and elicit players knowledge on complex aspects. In fact, the game is purposely designed to render the elicitation task as natural and simple as possible, without exposing the participant to the complexity of the game model, which is designed to address the research objectives. However, additional communication strategies could be explored in order to ensure that all the players might gain a broader awareness on how the game is addressing these objectives, potentially also enhancing their learning performance.

From a workload perspective the NASA TLX questionnaire provided positive results. In fact, it appears that the overall workload score is 33, were the major contributions as expected were related to mental demand (score of 48), performance demand (score of 65) and the effort demand (score of 40). Low scores, instead, were obtained for physical demand (score 4), temporal demand (score 21) and frustration (score 21).

To ensure that the research design produces findings that contribute to the intended knowledge [44] we assessed the knowledge utility, interpreted as the extent to which the game “generates findings that actually measure what the researcher intended to measure” [44]. The analysis of the collected data shows where and how the collected expert knowledge contributes

Question	Strongly agree	Agree	Neither disagree nor agree	Disagree	Strongly disagree
Completing the game tasks gave me a satisfying feeling of accomplishment.	11.11%	55.56%	22.22%	11.11%	0.00%
I feel satisfied with my contribution to the stated game objective.	16.67%	50.00%	33.33%	0.00%	0.00%
I felt good interacting with other players during the game.	22.22%	33.33%	33.33%	5.56%	0.00%
I forgot about my immediate surroundings while playing this game.	5.56%	11.11%	27.78%	44.44%	11.11%
I had fun with the game.	44.44%	38.89%	5.56%	5.56%	0.00%
I needed to learn a few things before I could play the game.	27.78%	44.44%	11.11%	16.67%	0.00%
I prefer to perform knowledge acquisition through other ways (e.g. questionnaires).	0.00%	22.22%	38.89%	27.78%	11.11%
I think that most people would learn to play this game very quickly.	22.22%	66.67%	0.00%	11.11%	0.00%
I think that the game is easy to play.	33.33%	55.56%	5.56%	5.56%	0.00%
I was able to interact with other players during the game.	38.89%	22.22%	22.22%	5.56%	11.11%
I was so involved in my gaming task that I lost track of time.	11.11%	22.22%	27.78%	27.78%	11.11%
I would recommend this game to my colleagues.	27.78%	44.44%	16.67%	11.11%	0.00%
It is clear to me how the contents of the game are related to the objective.	22.22%	44.44%	22.22%	11.11%	0.00%
It is due to my personal effort that I managed to advance in the game.	0.00%	22.22%	55.56%	22.22%	0.00%
Learning to play this game was easy for me.	38.89%	55.56%	0.00%	5.56%	0.00%
Something happened during the game (game elements, competition, etc.) which made me smile.	33.33%	38.89%	16.67%	5.56%	0.00%
The colors used in the game are meaningful.	22.22%	27.78%	33.33%	16.67%	0.00%
The contents and structure helped me to become confident that I would suppor the stated objective with this game.	11.11%	50.00%	22.22%	5.56%	0.00%
The fonts (size and style) used in the game are easy to read.	33.33%	50.00%	0.00%	11.11%	5.56%
The game contents are relevant to my interests.	38.89%	55.56%	0.00%	5.56%	0.00%
The game design is attractive (boards, cards, etc.).	38.89%	44.44%	16.67%	0.00%	0.00%
The game does not become monotonous as it progresses (repetitive or boring tasks).	22.22%	66.67%	11.11%	0.00%	0.00%
The game promotes cooperation and/or competition among the players.	5.56%	27.78%	44.44%	16.67%	0.00%
The game provides new challenges (offers new obstacles, situations or variations) at an appropriate pace.	33.33%	38.89%	22.22%	5.56%	0.00%
The game rules are clear and easy to understand.	27.78%	44.44%	22.22%	5.56%	0.00%
The text font and colors are well blended and consistent.	27.78%	55.56%	11.11%	5.56%	0.00%
There was something interesting at the beginning of the game that captured my attention.	22.22%	55.56%	22.22%	0.00%	0.00%
This game is an adequate knowledge acquisition method for the task.	16.67%	33.33%	44.44%	5.56%	0.00%
This game is appropriately challenging for me	16.67%	50.00%	11.11%	22.22%	0.00%
When I first looked at the game, I had the impression that it would be easy for me.	16.67%	50.00%	22.22%	11.11%	0.00%

Table 5: *Overall player experience (PX) of the participants of the three TTXs*

to the stated objective of the MUST Game. Specifically, we have been able to collect data on the relevant problem variables as identified by the game conceptual model (see Section 3). Moreover, after the conversion of the collected data into subjective probabilities following [5] we observed how the assessments changed in the different steps of the game. For example, Figure 9 illustrates how the deployment propensity, asset type selection and importance of optimisation decision factors changed with the assessed threat level along the turns played by two different players (Player A and Player B). The results of the analysis align with the validity concept of defence experiments as proposed in [45]. In fact, we have shown the ability to employ the new capability (i.e., the MUST Game), to detect the change during the experiment, to isolate the reason for change (e.g., evolving threat level) and to relate the results to actual operations. Specifically, the data collected has been analysed to extract the features relevant

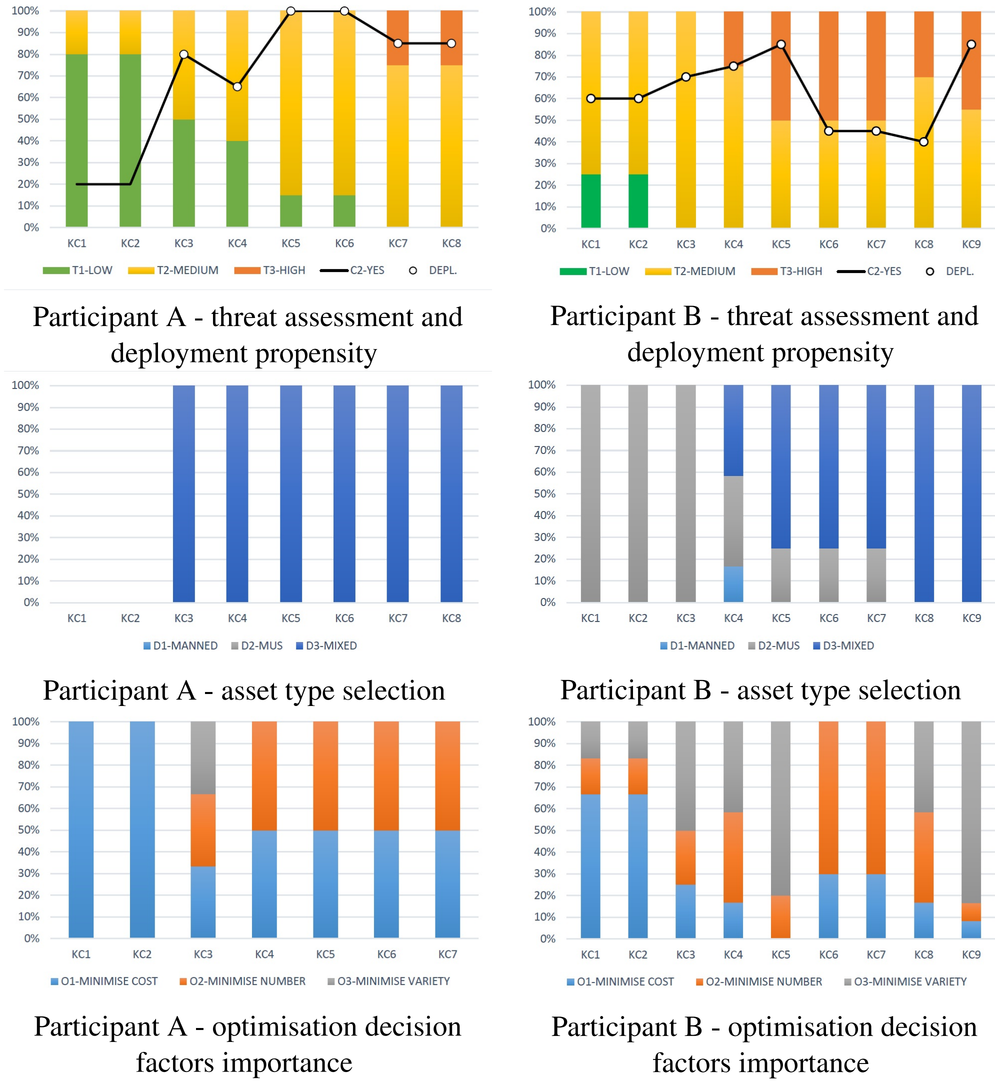

Figure 9: *Examples of players' individual threat assessment (green = low, yellow = medium and red = high), propensity towards deployment (black line) and the decision to deploy (white dot).*

to the current algorithmic development efforts and is currently further investigated to build the corresponding computational model.

4.3 Additional observations on the MUST Game design

The game appears to be sufficiently intuitive to start the conversation soon after the introduction to the scenario. The pre-game data collection phase, in which players are ranking the trust factors, appears to meet both objectives of familiarising with these factors and collecting useful data with respect to the players subjective assessment of their importance. Moreover, it acts as a useful ice-breaker at the start of the game.

In most sessions the facilitator had to perform also the role of the data collector. Although the data collection sheet was created with the intent to maximise the data recorded by the player itself, the amount of unstructured data collected appears to be considerable, which on one side made it complex for the facilitator to perform both tasks, but on the other side shows the effectiveness of the MUST Game as a means to stimulate the discussion and elicit the required information. The level of abstraction selected for the scenario, knowledge cards and technology cards appears to be adequate to stimulate relevant discussions, while making the game flexible to adapt to the level of analysis where the player feels more comfortable. In fact, depending on the kind of player the discussion converged more towards tactical considerations or higher level operational aspects, with some incursions into the strategic level.

Depending on the number of players the time required to play the full game, including the pre-game and post-game data collection, appears to be bounded between three (3) and four (4) hours. It is interesting to notice that while it was expected that multi-player distributed sessions would require more time, those sessions require less time than the sessions run in presence. This might be related to the more strict game control that needs to take place in the distributed environment as well as an increased tendency towards open discussions when physical co-location occurs.

During the first computer-assisted TTX the participant insisted in providing the results in an aggregated consensus format, reached after consultation and confrontation between the players (Module 3C in Figure 3). This had the undoubted advantage of generating discussion between them, which was perceived as a useful training session from the player perspective, as they were all militaries of a multi-national group. However, there should be a preliminary assessment to verify if this result is in line with the data collection and overall research objectives. In fact, the problem at hand might require to collect the subjective assessments before consensus reaching. Moreover, this data might be a useful support in the analysis phase, for example if the analyst did not follow the full interaction (e.g., due to communication issues). Therefore, if the individual assessments are needed the designers should carefully assess strategies that might incentivise participants to submit individual results. The second distributed TTX was run following the approach outlined in Module 3B of Figure 3, where individual assessments on the single KC were followed by a group confrontation, which generated valuable insights on the problem.

5 Conclusion

This paper shows how analytical games could be used to support the development of concept of use and decision support tools related to innovative and disruptive technologies. Specifically, it presents the use case of the Maritime Unmanned System Trust (MUST) Game. The game has been able to capture how participants with different background and expertise perceive the use of MUS in maritime operation scenarios with an evolving threat level. From the decision made by the players and their explanations we have been able to observe and record players' perspectives on aspects related to the peculiarity of the use of MUS in the maritime compared to other warfighting domains, potential barriers to uptake and which factors (i.e., cost, variety and number) have a stronger impact on the decision-making process. Furthermore, the results appear to shine an interesting light on the perceived importance of the trust factors, which could have relevant implications on directing future research and development efforts of these technologies. The outcomes of the deployment of the game in three exercises overall demonstrate that the MUST Game is effective in eliciting constructive insights around the use of MUS and very positive feedback has been received on the potential future employment of this approach to explore more in-depth the aspects of interest related to MUS employment in maritime missions. Future work will explore the exploitation of the in game data to inform the design of the models to be included in dedicated decision support tools. The concepts and proposed approaches have been contextualised into a maritime scenario making use of MUS. However, they are generic and could be applied to many other domains, as well as other EDTs. Therefore, future work could consider to adapt the game to other contexts.

References

[1] D. Djaouti, J. Alvarez, and J.-P. Jessel, "Classifying serious games: the G/P/S model," *Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches*, vol. 1, pp. 118–136, 2011. doi: 10.4018/978-1-60960-495-0.ch006

[2] E. Bartels, "Innovative education: Gaming - learning at play," *ORMS Today*, vol. 41, no. 4, 2014.

[3] G. Le Bourhis, L. Bertonati, and M. Murphy, *Maritime Unmanned Systems in ASW. Collaborative ASW study*. North Atlantic Treaty Organisation, Combined Joint Operations from the Sea Centre of Excellence: North Atlantic Treaty Organisation, 2018.

[4] A. Evangelio, O. Nyaas, G. Yuzichuck, S. Sweeney, M. Zaragoz, M. Coffman, and J. Fox, *NATO Guidance for developing Maritiem Unmanned Systems (MUS) capability*. North Atlantic Treaty Organisation, Combined Joint Operations from the Sea Centre of Excellence: North Atlantic Treaty Organisation, 2012.

[5] F. de Rosa, "Knowledge Acquisition Analytical Games: games for cognitive system design," Ph.D. dissertation, University of Genoa, Italy, 2020.

[6] J. Baron and E. Weber, "Introduction," in *Conflict and Trade offs in Decision making*, E. Weber, J. Baron, and L. G., Eds. Cambridge University Press, 2011, pp. 1–23.

[7] S. K. Kondreddi, P. Triantafillou, and G. Weikum, "Human computing games for knowledge acquisition," in *Proceedings of the 22nd ACM International on Information & Knowledge Management*, ser. CIKM '13. New York, NY, USA: ACM, 2013. doi: 10.1145/2505515.2508213. ISBN 978-1-4503-2263-8 pp. 2513–2516.

[8] S. E. Thaler, S. and K. Siorpaes, "Spotthelink: a game for ontology alignment," in *6th Conference for Professional Knowledge Management WM2011*, 2011.

[9] K. Siorpaes and M. Hepp, "Games with a purpose for the semantic web," *IEEE Intelligent Systems*, vol. 23, no. 3, pp. 50–60, May 2008. doi: 10.1109/MIS.2008.45

[10] L. von Ahn, "Games with a purpose," *Computer*, vol. 39, no. 6, pp. 92–94, Jun. 2006. doi: 10.1109/MC.2006.196

[11] L. von Ahn and L. Dabbish, "Labeling images with a computer game," in *CHI Conference on Human Factors in Computing Systems*, 2004. doi: 10.1145/985692.985733 pp. 319–326.

[12] D. Morrison, S. Marchand-Maillet, and E. Bruno, "Tagcaptcha: Annotating images with captchas," in *Proceedings of the 18th ACM International Conference on Multimedia MM '10*. New York, NY, USA: ACM, 2010. doi: 10.1145/1873951.1874284 pp. 1557–1558.

[13] T. Markotschi and J. Völker, "Guess What?! Human intelligence for mining linked data," in *Proceedings of the Workshop on Knowledge Injection into and Extraction from Linked Data (KIELD) at the International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, 2010.

[14] T. Chklovski, "Learner: A system for acquiring commonsense knowledge by analogy," in *Proceedings of the 2nd International Conference on Knowledge Capture K-CAP '03*. New York, NY, USA: ACM, 2003. doi: 10.1145/945645.945650 pp. 4–12.

[15] A. Herdagdelen and M. Baroni, "The concept game: better commonsense knowledge extraction by combining text mining and game with a purpose," in *Proceedings of the 2010 Commonsense Knowledge Symposium (AAAI CKS)*, 2010.

[16] S. Hoppenbrouwers, B. Schotten, and P. Lucas, "Towards games for knowledge acquisition and modeling," *International Journal of Gaming and Computer-Mediated Simulations*, vol. 2, no. 4, pp. 48–66, 2010. doi: 10.4018/jgcms.2010100104

[17] S. Meijer and G. Hofstede, "The trust and tracing game," in *Proceedings of 7th International Workshop of the IFIP, Aalborg, Denmark*, 2003.

[18] S. Meijer, G. Zuniga-Arias, and S. Sterrenburg, "Experiences with the mango chain game," in *Proceedings of the 9th International workshop of the IFIP, Espoo, Finland*, R. Smeds, J. Riis, P. Haho, and M. Jaatinen, Eds. 123–132, 2005.

[19] D. de Siqueira Braga, M. Niemann, B. Hellingrath, and F. B. de Lima Neto, "The game of trust: Using behavioral experiment as a tool to assess and collect trust-related data," in *Trust Management XI*, J.-P. Steghöfer and B. Esfandiari, Eds. Cham: Springer International Publishing, 2017, pp. 41–48.

[20] F. de Rosa, A.-L. Jousselme, and A. De Gloria, "A Reliability Game for Source Factors and Situational Awareness Experimentation," *International Journal of Serious Games*, vol. 5, no. 2, pp. 45–64, 2018. doi: 10.17083/ijsg.v5i2.243

[21] F. de Rosa, A. De Gloria, and A.-L. Jousselme, "Analytical games for knowledge engineering of expert systems in support to situational awareness: the reliability game case study," *Expert Systems with Applications*, vol. 138, pp. 112 800–112 811, 2019. doi: 10.1016/j.eswa.2019.07.017

[22] F. de Rosa and A. De Gloria, "An analytical game for knowledge acquisition for maritime behavioral analysis systems," *Applied Science*, vol. 10, no. 2, p. 591, 2020. doi: 10.3390/app10020591

[23] T. Kujala, P. Saariluoma, and M. Porta, "Cognitive mimetics for designing intelligent technologies," *Advances in Human-Computer Interaction*, vol. 2018, 2018. doi: 10.1155/2018/9215863

[24] S. Collins, *Disruptive Technology Assessment Game Handbook*. North Atlantic Treaty Organisation, Allied Command Transformation: North Atlantic Treaty Organisation, 2014.

[25] V. Groom and C. Nass, "Can robots be teammates?: Benchmarks in human-robot teams," *Interaction Studies*, vol. 8, pp. 483–500, 2007. doi: 10.1075/is.8.3.10gro

[26] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, vol. 53, no. 5, pp. 517–527, 2011. doi: 10.1177/0018720811417254

[27] K. E. Schaefer, J. Y. C. Chen, J. L. Szalma, and P. A. Hancock, "A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems," *Human Factors*, vol. 58, no. 3, pp. 377–400, 2016. doi: 10.1177/0018720816634228

[28] *Automation*. https://www.dictionary.com/browse/automation.

[29] U. Krogmann, "From automation to autonomy: Trends towards autonomous combat systems," in *Advances in Vehicle Systems Concepts and Integration (RTO MP-44)*. Neuilly Sur-Seine Cedex, France: NATO Research and Technology Organization, 1999.

[30] M. Bradley, D. C. Tarraf, T. C. Whitmore, J. DeWeese, C. Kenney, J. Schmid, and P. DeLuca, *Advancing Autonomous Systems - An Analysis of Current and Future Technology for Unmanned Maritime Vehicles*. Santa Monica, California: RAND Corporation, 2019.

[31] F. de Rosa and A. De Gloria, "Design methodology of analytical games for knowledge acquisition," *International Journal of Serious Games*, vol. 8, no. 4, pp. 3–23, 2021. doi: 10.17083/ijsg.v8i4.456

[32] B. Brathwaite and I. Schreiber, *Challenges for Game Designers*. Charles River Media, 2008.

[33] T. Price, *The High North - A Matrix Game of Artic Crises*, 2018.

[34] NATO Allied Command Transformation, *Bear Rising Matrix Game*. North Atlantic Treaty Organisation, Allied Command Transformation, 2019.

[35] J. T. Cacioppo, R. E. Petty, J. A. Feinstein, and W. B. G. Jarvis, "Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cog-

nition," *Psychological Bulletin*, vol. 119, no. 2, pp. 197–253, 1996. doi: 10.1037/0033-2909.119.2.197

[36] T. Franke, C. Attig, and D. Wessel, "A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale," *International Journal of Human-Computer Interaction*, vol. 35, no. 6, pp. 456–467, 2019. doi: 10.1080/10447318.2018.1456150

[37] J. T. Cacioppo and R. E. Petty, "The need for cognition," *Journal of Personality and Social Psychology*, vol. 42, no. 1, pp. 116–131, 1982. doi: 10.1037/0022-3514.42.1.116

[38] J. T. Cacioppo, R. E. Petty, and C. F. Kao, "The efficient assessment of need for cognition," *Journal of Personality Assessment*, vol. 48, pp. 306–307, 1984.

[39] G. Petri, C. G. von Wangenheim, and A. F. Borgatto, "A large-scale evaluation of a model for the evaluation of games for teaching software engineering," in *Proceedings of the 39th International Conference on Software Engineering: Software Engineering and Education Track*, ser. ICSE-SEET '17. Piscataway, NJ, USA: IEEE Press, 2017. doi: 10.1109/ICSE-SEET.2017.11 pp. 180–189.

[40] B. D. Harper and K. L. Norman, "Improving user satisfaction: The Questionnaire for User Interaction Satisfaction version 5.5," in *Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference*, Virginia Beach, VA, 1993, pp. 224–228.

[41] Human Performance Research Group, *NASA Task Load Index*. NASA Ames Research Center, 1986.

[42] C. J. Sadowski and H. E. Cogburn, "Need for cognition in the big-five factor structure," *Journal of Psychology*, vol. 131, pp. 307–312, 1997. doi: 10.1080/00223989709603517

[43] R. C. Rubel, "The epistemology of war gaming," *Naval War College Review*, vol. 59, no. 2, 2006.

[44] M. Polski, "Back to basics-research design for the operational level of war," *Naval War College Review*, vol. 72, no. 3, 2019.

[45] Joint Systems Analysis (JSA) Group, Methods and Approaches for Warfighting Experimentation Action Group 12 (AG-12), "Guide for understanding and implementing defence experimentation (GUIDEx)," Bedford, 2006.