

A Reliability Game for Source Factors and Situational Awareness Experimentation

Francesca de Rosa¹, Anne-Laure Joussemme², Alessandro De Gloria³

^{1,2} NATO STO Centre for Maritime Research and Experimentation {Francesca.deRosa, Anne-Laure.Joussemme}@cmre.nato.int

^{1,3} University of Genoa, Italy

{S2822976, adg}@unige.it

Abstract

Systems to support Situational Awareness take increasingly advantage of data and information fusion techniques. Due to the rise in the variety of information sources (e.g., sensors, open-source, intelligence, historical databases) and their possible lack of veracity those techniques should account for source reliability. Although the appropriate mathematical instruments exist, it still needs to be fully understood what are the factors that contribute to source reliability and what is their relative and total impact on Situation Assessment, which is the process that builds up Situational Awareness. In order to characterise source factors impact on human belief assessment the Reliability Game has been developed. This is a data exchange game in which the players are requested to perform Situational Assessment tasks by mentally processing incoming information and meta-information, abstracted and provided through cards. This paper presents the method, the design choices and shows through a qualitative analysis that the proposed approach is indeed able to capture elements of source factors impact on players' belief changes. Data collected through the game will be further analysed to inform and improve the design of information correction methods in multi-source information fusion systems.

Keywords: Source Factors; Beliefs; Situational Awareness; Reliability; Game-Mechanics; Design.

1 Introduction

No unique definition for the term “serious game” exists, however it appears that there is a strong agreement on the fact that serious games are “designed for a primary purpose other than pure entertainment” [1]. Most of the research and applications in the domain of serious games have mainly focused on education, training and user learning objectives (e.g., [2], [3]). Different fields of engineering have started to look at serious games, not only from an educational perspective, but also as supporting design tools. A recent literature review of games used in engineering research [4] presents a classification of such games following the Gameplay/Purpose/Scope (G/P/S) taxonomy [5]. This model proposes game play, purpose and scope as relevant classification dimensions. The first dimension differentiates between play-based or game-based games. The former are games characterised by a lack of well-defined objectives and rules. The latter instead, have defined objectives and rules. The purpose dimension allows classifying games on the basis of their function (e.g., message broadcasting, training, data exchange). Message broadcasting games are the ones that have been developed with the aim of broadcasting a message (e.g., educative games, informative games, persuasive games and subjective games). Games for training are developed with the purpose of improving players (cognitive or physical) performances.



Finally, data exchange games have the specific purpose of supporting data exchange, such as “collecting information from [...] players” [5]. The scope dimension refers to the game market (e.g., state and government, military and defense, healthcare, education, corporate, religious, culture and art, ecology, politics, humanitarian, advertising, scientific research) and the target audience (e.g., professionals or general public).

From the above mentioned review it appears that in the realm of engineering research serious games focus has shifted from message broadcasting and training to the data exchange purpose. Two notable examples explore human problem-solving strategies to support computational algorithm optimisation in the context of protein structure design [6] and vehicle powertrain controller design respectively [7]. The findings derived from the use of the two games have shown that human-derived strategies can be a valuable resource when used in conjunction with computational algorithms.

Situational Awareness (SAW) is defined [8] as “the perception of the elements in the environment within a volume of time and space, comprehension of their meaning and the projection of their status in the near future” and the cognitive process that enables Situational Awareness as Situational Assessment (SA). Some serious games in the context of SAW have been developed (e.g., [9]). However, to the best of the authors’ knowledge, the focus of such games remains on training and message broadcasting, with the exception of [10] which presents a game for assessing team SAW.

With the final goal of informing the design of multi-source information fusion systems, we present in this paper a data exchange game, characterised by a game-based approach. The scope of this game, called the Reliability Game, is scientific research [5] as it aims at collecting data to be used in further research of source factors impact on human SA and consequent SAW. The term *source factor* is used in this paper with the specific meaning of element that characterises a source of information, such as its type (e.g., radar, human operator and historical databases), quality, reliability or attractiveness. It should be noted that although the target audience is professionals [5], namely subject matter experts (SMEs) in Maritime SAW, it could be extended to general public through the development of an appropriate scenario.

In this paper after providing some details with respect to the motivation of the game (Section 2) and the notion of source reliability (Section 3), we will discuss the design approach and choices (Section 4). We will present the game outcomes, which demonstrate the effectiveness of the game mechanics (Section 5). Finally, the conclusions and the way ahead are discussed (Section 6).

2 Motivation

Decision support tools take greater and greater importance in daily life, may it be a common car navigator or more complex surveillance systems that support operators in different working environments such as safety, security, crises management, health and first aid. Those systems all aim at improving the user’s SAW by helping the user to better capture, understand and predict future states of the situation at hand, for more informed decisions. SAW and the corresponding information systems form an important building block of the dynamic decision making processes [11]. Although those systems provide support to humans in problem solving and decision making, they still remain an “enabler, facilitator, accelerator and magnifier of human capability, not its replacement” [12].

To obtain from data, information and finally insight, processing capabilities have to rely on human assets to work correctly. Indeed, the information value to a specific application cannot be decoupled from the human component, in terms of the ability to search, analyse and interpret the data or information provided. For example, it has been demonstrated that SA might be negatively impacted by system automation that drives operators out of the loop [13]. Moreover, within operational environments human can concurrently and interchangeably act in more than one of the following roles: “decision

maker, monitor, information processor, information encoder and storer, discriminator, pattern recognizer [,] . . . ingenious problem solver” [14] or disseminator. To improve human-machine synergy and user acceptance the system underlying reasoning and communication schemes should be intelligible [15] and possibly intuitive to the human.

In addition to the challenge of a desirable human-centered system design approach [16], the systems that support SAW have to deal with an ever increasing volume and velocity of the information, coupled with an increase of the variety of the information and corresponding sources with a potential lack of veracity. Data and information fusion technologies come into play to support operators’ SA and reduce the information overload. To this end information aggregation (e.g., data fusion) approaches have proven to be effective, provided that the outputs are presented in an intuitive and actionable format that engenders trust [17], [18].

To get full advantage of the variety of sources beyond the ones traditionally in use, we need not only to combine them but also to correctly account for source factors in fusion processes [19 - 21]. With respect to source reliability, most mathematical fusion operators assume that the sources are fully reliable or at least equally reliable and therefore assign an equal weight on the resulting combined belief assessment [22]. In reality this assumption is not always satisfied and sources can differ in reliability. Several strategies within different uncertainty frameworks (e.g., Bayesian, belief functions) have been proposed to account for partially reliable sources [23]. Generally, the consideration of source reliability in the fusion process relies on discounting, pruning or reinforcement operations [19 - 22], allowing for instance to completely discard a piece of information provided by an unreliable source or to strengthen the weight of information originating from a highly reliable source. However, further research is needed to clarify some concepts related to source reliability, to clarify the semantics of source quality dimensions and to ensure that the implementation of those reliability accounting strategies in current support systems meet some criteria of understandability or intuition.

A literary review on human factors methods [23] lists several methods for the assessment of SAW. However, none of those techniques is suitable for our work, as they do not specifically focus on the SA, which instead is at the core of the Reliability Game [24]. In fact, the purpose of this innovative approach in the context of SAW experimentation is to collect data regarding players’ belief changes as a function source factors, more specifically source type and quality. To gather this data each player is presented with a scenario and plays several rounds of the game, where the only variation consists in the knowledge regarding source type and quality. The corresponding belief changes are captured through game items’ position (cards) and final confidence assessment, as it will be explained in further details in the next sections.

3 *The reliability concept*

For a proper consideration of reliability in the fusion process, it is helpful to understand what source reliability is and to define the underpinning elements that are central to its quantification. There is no universal definition of source reliability and even fields that have traditionally been working with multi-source information such as military intelligence neither have come to a definition, to a formalisation of the concept nor to an agreement on the rating of the source reliability [25]. Following [26], reliability is defined as “ability to rely on or depend on, as for accuracy, honesty and achievements”. It is important to underline that the term *ability* does not represent an ability of the source itself, rather our bet on the ability to rely on it. Therefore it is our own estimate, which is a function of many factors including the capacity and/or willingness of the source of providing good information. In the field of intelligence source reliability is evaluated on the basis of past meta-knowledge and experience with the specific source. However, in general it might depend on several other factors, such as similarity, perceived expertise, attractiveness [27]

of the source or experience with analogous sources (encapsulated in source type). The purpose of work described in the following sections is to understand how source factors underpinning source reliability, specifically source type and source quality, impact SA and SAW. The source reliability is treated as a latent variable, therefore is never specifically mentioned in the Reliability Game execution.

4 *The Reliability Game overall design*

The Reliability Game core is reasoning under uncertainty with information provided by sources of different type and quality, which are assumed as two underpinning factors of source reliability. The aim of the Reliability Game is to capture the impact of source factors on human SA. One of the final goals is to inform the design of automated reasoners to be included in multi-source information fusion systems.

This section summarises the design of the Reliability Game. More specifically, the following subsections provide details about the world design (Section 4.1), the system design (Section 4.2) and the content design (Section 4.3). The world design is defined [28] as “the creation of the overall backstory, setting and theme”, while the “creation of rules and underlying mathematical patterns” is identified under the definition of system design [28]. Finally with the term content design we refer to the “creation of characters, items, puzzles and missions” [28].

The Reliability Game design follows a mechanic-driven approach, which starts from the definition of the game core, followed by the selection of specific game mechanics (GM). The following GMs have been identified in an early stage of development:

- [GM1] Assessment of hypotheses relative to a missing vessel;
- [GM2] Use of cards to communicate messages to the player.

Those two mechanics were selected as they proved to be effective elements proposed in the Risk Game [29], which is a game that aims at eliciting experts’ knowledge regarding reasoning about concurrent events when dealing with information that differs in nature (e.g., from sensors or from humans) and in quality. The information quality dimensions considered in the Risk Game were accuracy, precision and trueness. Following [30], the term accuracy can be interpreted as the “closeness of agreement between a test result or measurement result and the true value”. The term precision refers to the “closeness of agreement between independent test/measurement results obtained under stipulated conditions” and the trueness refers to “closeness of agreement between the expectation of a test result or a measurement result and a true value”, where the measurement is the information.

Differently from the Risk Game, the Reliability Game assumes as fixed such dimensions in order to obtain a more rigid experiment control. This choice has been driven by the need to isolate the experiment variables (source type and source quality) impact on the final players’ belief assessment.

4.1 *World design*

The game is set in a maritime scenario and refers to a fictitious geographical area with the sovereign of three different countries:

- [L1] Right Land is a failed and poor state;
- [L2] Centre Land suffers of disorders due to the vicinity to Right Land;
- [L3] Left Land is a stable and rich country, thanks to the presence of oil extraction facilities within their Exclusive Economic Zone (EEZ) and to the Left Land Canal, which is a strategic waterway owned, managed and maintained by the Left Land government.

The player is part of Left Land Maritime Authority, which is the only authority with responsibilities within Left Land national waters. Therefore it is responsible for maritime safety, maritime security, environmental protection, customs and port state control. More specifically the player embodies the head of the monitoring department, who is informed by a subordinate that the Automatic Identification System (AIS) contact of the tanker ship MV Red Horizon (Figure 1) has been lost since six hours. The player is asked to assess what is currently happening to the ship in order to take further actions.



Figure 1. *MV Red Horizon information (vessel of interest) and its track before AIS contact loss as displayed in the scenario map by the red line*

The player is presented with a set of three collectively exhaustive and mutually exclusive hypotheses and is asked to perform a belief assessment about what is happening to the ship on the basis of the incoming information. The three candidate hypotheses for this scenario are:

- [H1] Nothing happened: the ship is continuing its voyage without safety or security issues;
- [H2] Safety issue: there is a safety issue with the ship;
- [H3] Security issue: the ship is engaged in oil smuggling activities.

Hypothesis H1 would be explained by the fact that the AIS signal is not received due to a possible failure of the AIS, and no intervention would be required. On the contrary H2 or H3 would trigger respectively a Search and Rescue (SAR) operation or a security operation. The action phase *per se* is not part of the game as the game stops after the SA phase. However, it represents the driver and motivation of the player.

4.2 System design

A game session is divided in four rounds, in which a set of eleven cards is provided to the player. In each round the player is requested to assess what is happening to the ship on the basis of the available information and meta-information on source factors (source type and source quality) provided through cards. We will refer to the card as conveying a Message (*M*), which is composed by the information (*I*) and associated meta-information about source factors (*SF*), namely source quality (*Q*) and (*T*). The information provided might be true or false. Although it is not explicitly requested to assess information trueness, the player will implicitly assess this information dimension as a consequence of the game dynamics.

A summary of the game state, intended as the picture of all relevant variables that may change during the play [28] is reported in Table 1. Table 2 summarises the game view, which is the portion of the game state that is visible to the player in each round [28]. With respect to the game view it can be noticed that each round is exactly the same (e.g., scenario, triggering event, information presented in the cards, order of cards), with the only exception

of the meta-information about source factors (Q and T). The order of the cards is kept constant with the purpose of controlling the information presentation order effect [31].

Table 1. Reliability Game state

Variable	Description	Frame
H	Hypothesis	{H ₁ , H ₂ , H ₃ }
M	Message conveyed by a card	{M ₁ , M ₂ , M ₃ , M ₄ , M ₅ , M ₆ , M ₇ , M ₈ , M ₉ , M ₁₀ , M ₁₁ }
I	Information conveyed by a card	{I ₁ , I ₂ , I ₃ , I ₄ , I ₅ , I ₆ , I ₇ , I ₈ , I ₉ , I ₁₀ , I ₁₁ }
I _r	Information trueness	{True; False}
Q	Source quality	{1, 2, 3, 4, 5, Unknown}
T	Source type	{AIS, LRIT, Company Security Officer, National reporting procedure, Intelligence report, Maritime Safety Agency, Smart agent, Tool providing Patterns of Life on routes, Tool providing Patterns of Life on calls, Operator + Radio, Operator + VTS + SAR }
C	Confidence level	{1, 2, 3, 4, 5, Unknown}

Table 2. Reliability Game view*

Variable	Description	Round 1	Round 2	Round 3	Round 4
H	Hypothesis	Assessed	Assessed	Assessed	Assessed
M	Message conveyed by a card	Provided	Provided	Provided	Provided
I	Information conveyed by a card	Provided	Provided	Provided	Provided
I _r	Information trueness	Assessed Implicitly	Assessed Implicitly	Assessed Implicitly	Assessed Implicitly
Q	Source quality	Not provided	Provided	Assessed	Provided
T	Source type	Not provided	Not provided	Provided	Provided
C	Confidence	Assessed	Assessed	Assessed	Assessed

*Assessed = player has to assess the item and communicate it to the facilitator; Provided = item value provided to the player; Not Provided = item value not provided to the player; Assessed Implicitly = player has to assess the item but not to communicate it to the facilitator.

Each card needs to be positioned on a game board (Figure 2), which is specifically designed to capture the player's belief assessment toward the different hypotheses displayed in the corners of the triangle. The selected position reflects the weight of belief that the information contained in a card provides toward some subsets of hypotheses presented. For example, positioning a card in the lower corner of the triangle indicates that the specific piece of information provided by that card is pointing towards hypothesis H1 only, while positioning the card in H1 or H2 point would indicate that the card is pointing towards both hypotheses only (meaning excluding H3), but that the player could not discriminate between the two. The cards can be positioned not only on the points, but also on the axes connecting two points, to express some relative weight assigned to the belief towards a specific hypothesis or set of hypotheses. The shuffling of the cards during the round, after new information is discovered, is possible to allow the player's belief updating based on new evidence. Once all the eleven cards have been processed and positioned on the board, the player is asked to rate the global confidence in the three hypotheses. For the purpose of the game, confidence was defined as "the state of feeling certain about the truth of something" [32]. The winning condition corresponds to the assignment of the highest confidence rate to the correct hypothesis. Details on the confidence rating can be found in next section. Figure 3 illustrates a diagram of a game session, explaining the main actions that the participant has to perform.

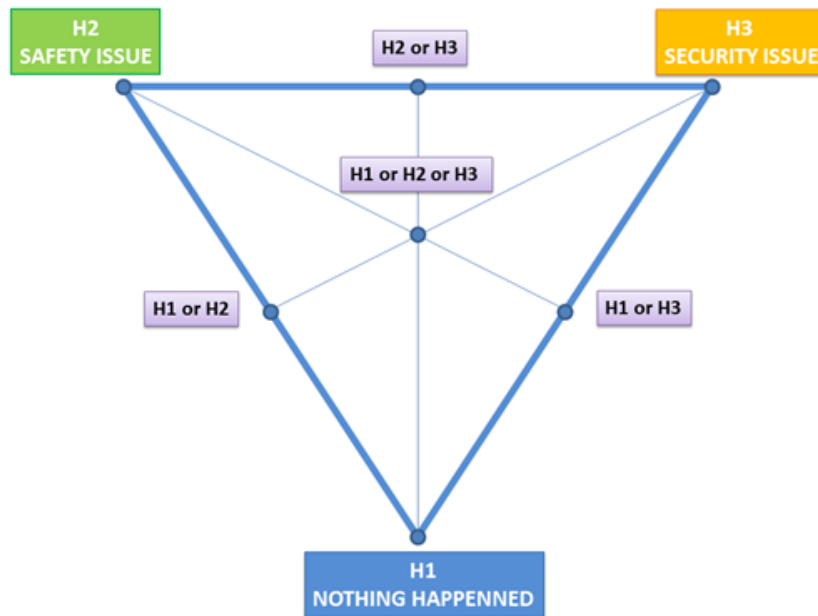


Figure 2. *Game board on which the cards need to be positioned*

To summarise the basic game mechanics are:

- [GM1] Assessment of hypotheses relative to a missing vessel;
- [GM2] Use of cards to communicate messages to the player;
- [GM3] The investigation component;
- [GM4] The card positioning on the board to rate the support of a message towards hypotheses;
- [GM5] The shuffling of cards as a consequence of new evidence acquisition (optional);
- [GM6] The global confidence rating of the hypotheses at the end of each round.

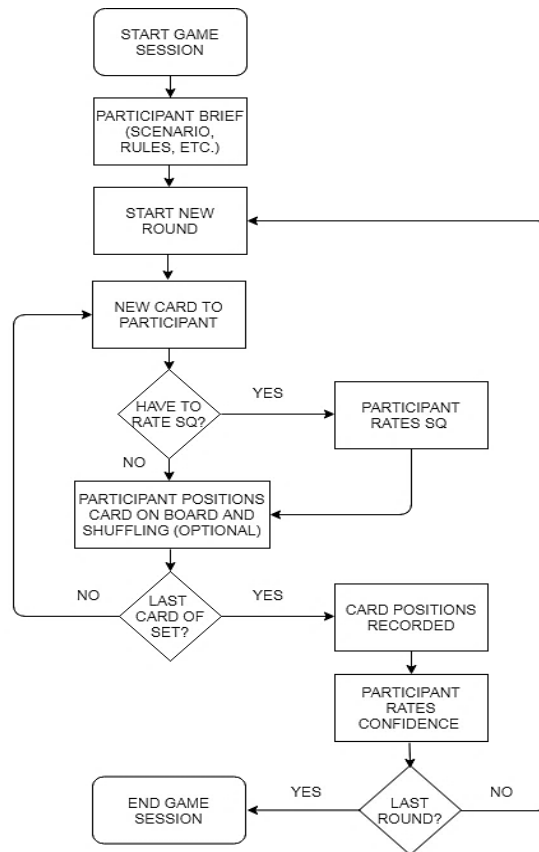


Figure 3. Diagram of the a session of Reliability Game

4.3 Content design

At the start of the session the player is introduced by a facilitator to the game core and to the game mechanics. During this introduction session the scenario, rules and different game elements (e.g., game board, scenario map, cards and flashcards) are presented to the player.

The scenario map (Figure 1) depicts the geographical area and other relevant geographical contextual information, such as the location of borders, the location of oil installations and the presence of a primary shipping lane that crosses the Exclusive Economic Zone (EEZ) of Left Land, leading to the trans-oceanic channel. Moreover it visualises the AIS track of the ship of interest before the contact was lost.

The Messages displayed on the cards are divided in three areas, namely the source type area, the source quality area and the information area as can be seen in Figure 4.

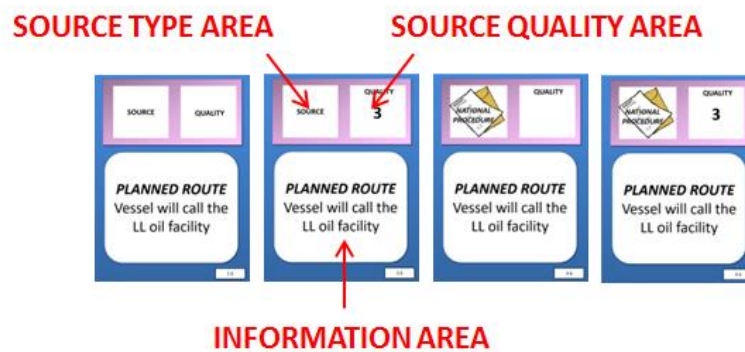


Figure 4. Example of the presentation of the same message in the four different rounds

As previously mentioned the only variation between the cards in the different rounds is in the meta-information on source type and source quality. As can be noted the information area is kept constant, while the source type area and source quality area are changing. Examples of message content in terms of the conveyed information, source type, source quality assigned (in Round 2 and Round 4) and information trueness is provided in Table 3. For instance, Message 7 reads that the Company Security Officer, which quality is unknown, reports false information regarding the fact that nothing happened to the ship.

Table 3. *Example of Reliability Game messages*

M	Conveyed information	Source type	Source quality assigned	Information trueness
1	Current ship position in X	AIS	4	False
2	Ship not answering to radio calls	Operator + Radio	5	True
7	Report that nothing is happening to the ship	Company Security Officer	U	False
8	Comparison of position X with usual ship routes	Tool providing Patterns of Life on routes	3	True

In addition to the message cards the player is also presented with flashcards supporting the player's rating and providing additional contextual information. An example is provided in Figure 5, which shows the flashcards regarding the vessel of interest, the one on the source quality rating scale and finally the one on the confidence rating scale. While the first contains the relevant information regarding the ship (e.g., ship type, dimensions, flag state and ownership), the one on the source quality scale represents visually the suggested rating scale for the specific variable, which ranges from 1 to 5 or *Unknown*, with source quality 1 meaning low quality. A source quality scale of six levels has been selected to align with most of the existing standards of source reliability rating in the intelligence domain [25]. We explicitly avoided using percentage ranges and the reliability verbal expressions as studies demonstrated the subjective interpretation of the word reliability and of the matching between the verbal and numerical expression [25]. To provide an intuitive visual support to the understanding of the ranking a graphical representation of the scale has been included, which is inspired to the home energy efficiency rating chart [33].



Figure 5. *Flashcards – Vessel of interest, source quality levels and confidence levels in the analysis*

At the end of each round players are asked to rate the global confidence in their analysis of the current situation. The levels relative to the confidence rating are analogous to the ones for the source quality, but provide an additional definition for confidence. The rating

scale was selected with reference to the analysis performed in [25]. This review shows that with respect to confidence the available rating scales differ in the proposed number of rating levels. In fact, most scales vary from a five level scale to a three level scale. To minimise confusion and errors it was decided to adopt a six level scale in agreement with the source quality scale. Those levels correspond to the five levels present in the intelligence scales plus the *Unknown* value. This value is deliberately not included in intelligence scales as it is expected that intelligence analysts are able to state their confidence in an analysis [34], however it was deemed interesting its inclusion in order to verify if and how this value would be used if available. It is important to highlight that several standards defining confidence levels map the confidence terms with specific probability intervals. However as there is no agreement on the correspondence, such a mapping was not considered in the Reliability Game.

4.4 Game design constraints

The main constraints that had to be accounted for during the design phase can be categorised as physical constraints and cognitive constraints. The first ones are those acting on the physical elements of game or related to logistical aspects, while the later are the ones dealing with cognitive tasks to be performed by the player.

The main physical constraints are the dimensions of the game elements such as the cards that had to be manageable, readable and had to be moved easily. In addition to this, another important limitation is that in a non-digital game not all item moves can be easily recorded unless an external observer constantly records the moves (e.g., through notes or pictures).

The main cognitive constraints are the number of cards that have to be provided to the player, the game session length and the need for supporting elements to compensate for the fact that in real world activities operators can rely on background knowledge and on the support of real systems (e.g., a display showing the AIS track of the lost ship).

The size of the set of cards has been selected as a trade-off between the ability of the player to manage the set of cards and the attempt to minimise some effects that might impact the experiment results. Two notable effects are the random responding by the players [35] and the carryover effect [36]. The carryover effect takes place within subject experiments when one test might impact the one of the following tests. In order to minimise the carryover effect due to memorising the information from one round to the following one (also referred to as practice effect) it was decided to have a card set size major than seven. In fact, it has been suggested that the storage capacity of the short-term memory of an average person is approximately seven items, plus or minus two [37].

The game session length is a relevant cognitive constraint, as the game had to be short enough to keep the players attention, avoiding mind-wandering effects. Mind-wandering refers to the effect of the mind not focusing on a specific topic for a long period of time, which might occur especially when engaged in attention-demanding tasks [38].

5 Game evaluation

5.1 Game evaluation

Game evaluation is an important and critical part of design processes. In this context the term evaluation is used as proposed in [39], to avoid confusion with the concept of validation that in some contexts refers to measurements validity (e.g., measurements accuracy). Therefore, we will refer to the term evaluation as the “confirmation through the provision of objective evidence that the requirements for a specific intended use or application of a system have been fulfilled” [39], where the term “system” refers to the game to be evaluated.

As previously mentioned the purpose of the Reliability Game is to collect data regarding source quality and source type impact on SA and SAW. Therefore, in order to evaluate the effectiveness of the game with respect to the above mentioned scope the main criteria are the observation of *variations of card positions* and *confidence rating* between rounds. Because the only input variation between rounds consists in the meta-information about source type and source quality, it is assumed that the two above mentioned criteria are able to capture the corresponding impact on SA as belief change.

The following sections report the outcomes of a qualitative analysis of the data collected during an experiment run with the Reliability Game.

5.2 Experiment set-up

The game underwent a quick prototyping and play testing phase that allowed verifying the board design, the scenario, the information items proposed and the facilitation approach. After minor changes to some information items, a revised version has been issued. The collection of data is still ongoing, but we present herein data collected on a small but relevant sample of SMEs that allowed verifying the effectiveness of the proposed GMs. At the time of this paper the game has been played with twenty-one (21) players, which demographics and characteristics are reported in Table 4. Participants' selection was performed on a voluntary base from maritime SMEs, with either civil or military status. The experimental set-up followed a within subject design, in which the participants have been exposed to four different conditions, namely the game rounds. The conditions variation corresponds to the game view summarised in Table 2.

For each player the following in-game data has been collected:

- [D1] a picture of the final cards position at the end of each of the four rounds;
- [D2] the source quality rating during the third round;
- [D3] the confidence rating in the hypothesis at the end of each round.

Table 4. *Participants demographics and characteristics*

Feature		
Gender	Male	100 %
	Female	0 %
Age	Average	46.5 years
	Standard Dev.	10.3 years
Status	Military	76 %
	Civilian	24 %
Nationality	Italian	33.33 %
	France	4.76 %
	Danish	14.28 %
	Norwegian	4.76 %
	Romanian	4.76 %
	United Kingdom	14.28 %
	German	19.04 %
United States	4.76 %	

In this experiment there was not an external observer constantly recording the item movements. Thus, only the final aggregation of beliefs at the end of each round has been recorded (D1), while the shuffling of cards has not been captured. However, this represents

a minor issue as the cards shuffling resulted in a game mechanics seldom used by the players. Moreover, it will be completely superseded in a digital version of the game, currently under development.

Beside the in-game data collection above mentioned, a post-game data collection has been performed in the form of feedback questionnaire. The scope of this questionnaire was to assess participants' understanding of the game and perception with respect of this innovative gaming approach (e.g., relevance with respect to their mission, engagement, facilitation). It is important to note that this questionnaire has been provided as part of a broader feedback questionnaire and only 11 out of the 21 players of the Reliability Game returned their answers.

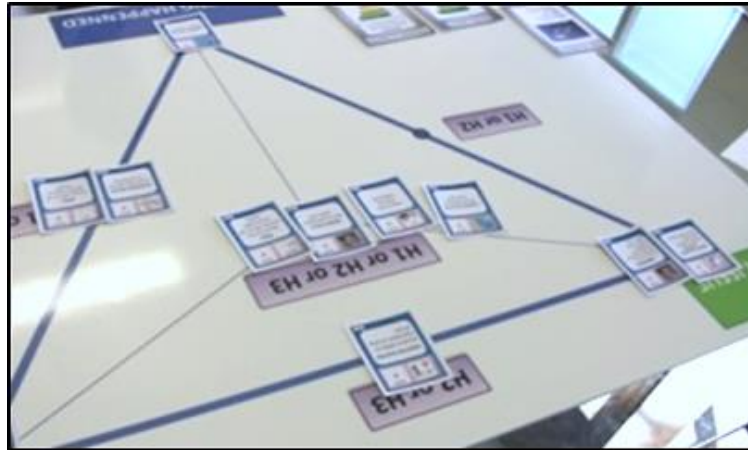


Figure 6. *Example of a picture (D1) collected at the end of a round*

5.3 *Feedbacks and observations on the game design*

The participant survey shows that the players perceived the game as engaging, realistic and relevant with respect to operational needs (Figure 7). From a facilitation point of view, it has been observed that it is important not only to introduce the players to the game rules and to have them familiar with the game dynamics, but also to clearly state and explain the game core to have the players feeling more comfortable and confident about the remaining part of the experiment. Most players actually were explaining their reasoning to the facilitator, which is considered of value for the refinement of next iterations of the game.

Players showed to understand well the purpose of the game and the game mechanics, which appears to be intuitive and requires a low level of pre-experiment training. It has to be underlined that in there is not a proper pre-experiment training session. Instead, the rules are explained and then the facilitator guides the player when providing the first cards by asking after the card is positioned if the player confirms that the card supports the belief associated to the specific card position. In case of a negative answer the facilitator would help the player positioning the card in the corresponding location.

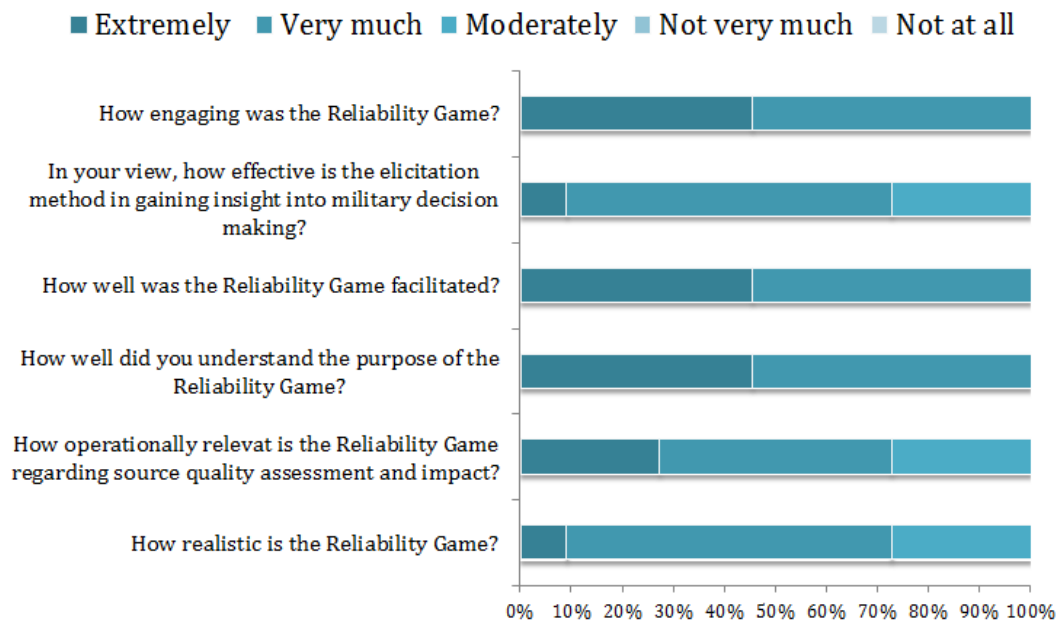


Figure 7. *Players' feedback questionnaire outcomes*

5.4 Outcomes on source quality rating

During the third-round of the game participants were requested to rate the source quality given the meta-information on source type which had been provided (Section 4.2). Figure 8 presents an example of the source quality rating by three different players (red diamonds), which is compared to the source quality values that are provided to the players in Round 2 and Round 4 (blue line). Empty values correspond to an *Unknown* rating. The three players presented different rating profiles. We can observe how Player A has a tendency to rate the source quality higher than the assigned source quality value. Player B demonstrates a tendency to variably rate the source quality higher, lower or equal to the assigned ratings. Finally, Player C shows a tendency to rate the source quality lower than the quality assigned.

Figure 9 depicts the overall source quality assessment for each of the 11 cards, as a percentage of players assigning a given source quality rating by card. It is important to mention that although the source quality values with decimals (3.5 and 4.5) were not included in the original scale, one player requested to use them. From the figure it can be observed that with the exception of the rating of this player, the assessments on Source 1 and Source 8 are identical. This is an important observation as the degree of familiarity of the SMEs with the two sources is considerably different. In fact, Source 1 (Automatic Identification System) is widely available and commonly used in maritime surveillance. On the contrary, Source 8 (Vessel to Route Association algorithm) is more experimental and is still in its early stages of development. Other novel information sources are Source 5 and Source 10, namely a vessel position prediction algorithm and a maritime Patterns-of-Life on ship statistics service. Both sources present a certain degree of variation in the quality rating. However, it can be observed that the one of Source 5 is higher than the one for Source 10. This result suggests that non-conventional information sources are not necessarily considered of low quality. Moreover, from the verbal players' feedback it appeared that the players were drawing comparisons between the source's capacity and their own cognitive abilities (e.g., ability of associating a ship to a route). This observation concurs with some persuasion literature on source factors which has shown the impact of the perceived source similarity on human information assessment [27]. Source 7 (Company Security Officer) is the one exhibiting the highest degree of variability in the quality ratings. This source in Round 2 and Round 4 has an *Unknown* assigned quality. Only three players

rated the source as such, while most of the players assigned a low-quality rating. As explicitly stated by some players, this appears not to be related to the nature of the source (human vs sensor), but rather to a possible conflict of interest of the Company Security Officer who could retain or falsify information due to conflict of interests. This observation is also supported by the fact that Source 2, the human operator, has been rated of high quality.

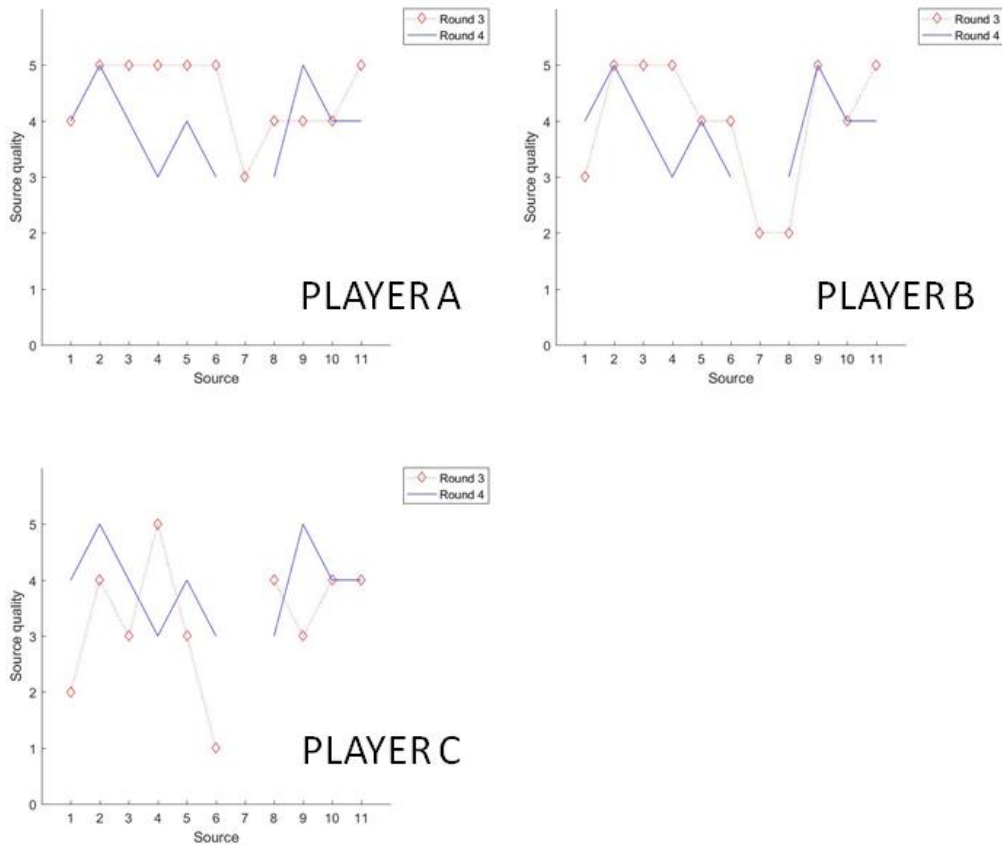


Figure 8. Example of source quality rating (Round 3) by three different players

An interesting result is the one related to the use of the value *Unknown*. In fact, it has been seldom used, even in the case in which the player had no knowledge of the type of source. More specifically, some players did not know the Company Security Officer or the Long Range Identification and Tracking system. They asked for information to the facilitator, who provided basic information, without disclosing details on the quality. Although the players were often reminded of the possibility of using the *Unknown* value, most of them did not do. This suggests that the players tended to estimate a source quality value even if the source is not known.

Moreover, as shown in Figure 8 there was a difference between the participants' source quality ratings and the values provided to them in Round 2 and Round 4. This translates in a variation of the conditions between Round 3 (source type provided, source quality assessed) and Round 4 (source type provided, source quality provided).

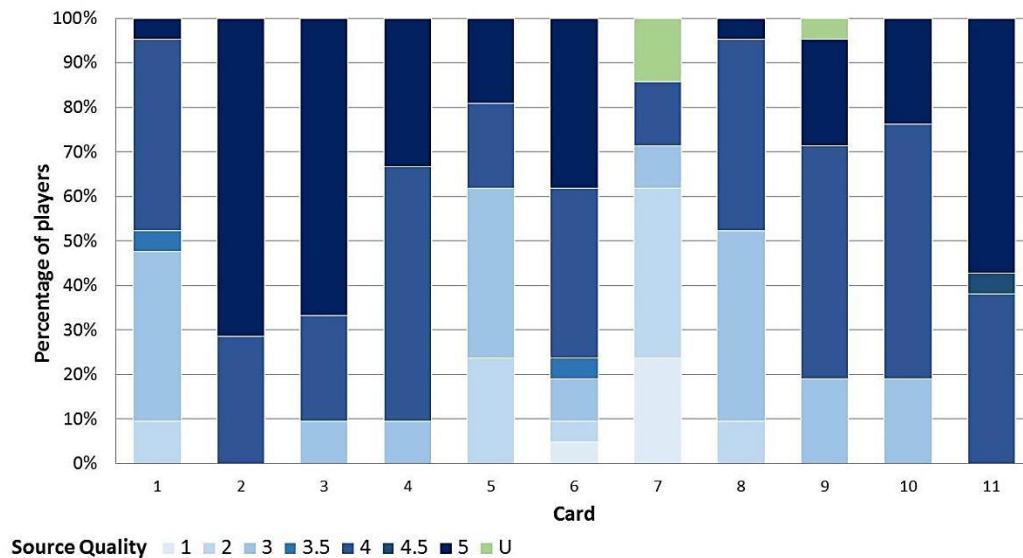


Figure 9. Source quality ratings by card

5.5 Outcomes on confidence rating

At the end of each round participants were requested to rate their confidence in the fact that the correct hypothesis might be H1, H2 or H3. Figure 10 displays the confidence rating of different players. With respect to the relative confidence ratings, it can be observed that the sum of the confidence in the hypotheses is not constant between the rounds and that the variation of the confidence in one of the hypothesis does not imply the variation of the confidence in the others.

From Figure 11, which is reporting a summary of the different confidence ratings of the participants, we can observe interesting results regarding the use of the scale presented. Equivalently to the case of source quality, one player asked to use a value with decimals. More specifically, the participant asked to introduce the value 2.5 as to express the concept of 50%, which is not possible with the original form of the scale. The proposed scale did not include the rating value 0, which conceptually corresponds to the exclusion of the hypothesis with high confidence. However, many players asked to use the value 0. On the contrary the value 5, corresponding to the conceptual opposite (full confidence that the specific hypothesis is the right one), has been rarely used. This result suggests that participants might more easily exclude hypotheses than being certain about them. Another possible interpretation is that they might feel more self-confident in excluding than being certain about the hypotheses. With the term self-confidence the authors refer to the concept of self-assurance in personal judgment.

Contrary to the intelligence scales a sixth level was included in the original scale, namely *Unknown*, to allow the players the possibility to state their inability to draw a conclusion and express their confidence. This value is on purpose excluded from the intelligence standard scales as it forces the analyst to exactly rate his confidence, no matter if high or low, without using the above mentioned value as a solution to avoid liability issues. It is, however, interesting to notice that this value has been used twice by the participants, but this is not reported in the graph as the players soon after asked if they could re-rate the confidence.

Another interesting observation regarding the confidence rating is that some players when requested to express their confidence were stating that it was unchanged with respect to the previous round. The facilitator, however, requested the players to explicitly rate the current confidence levels. The resulting rating was in general not equal to the previous one, suggesting that there had been a change of which the players were not conscious. This also

suggests that the mechanisms to try to minimise the carryover effect were effective on the confidence rating.

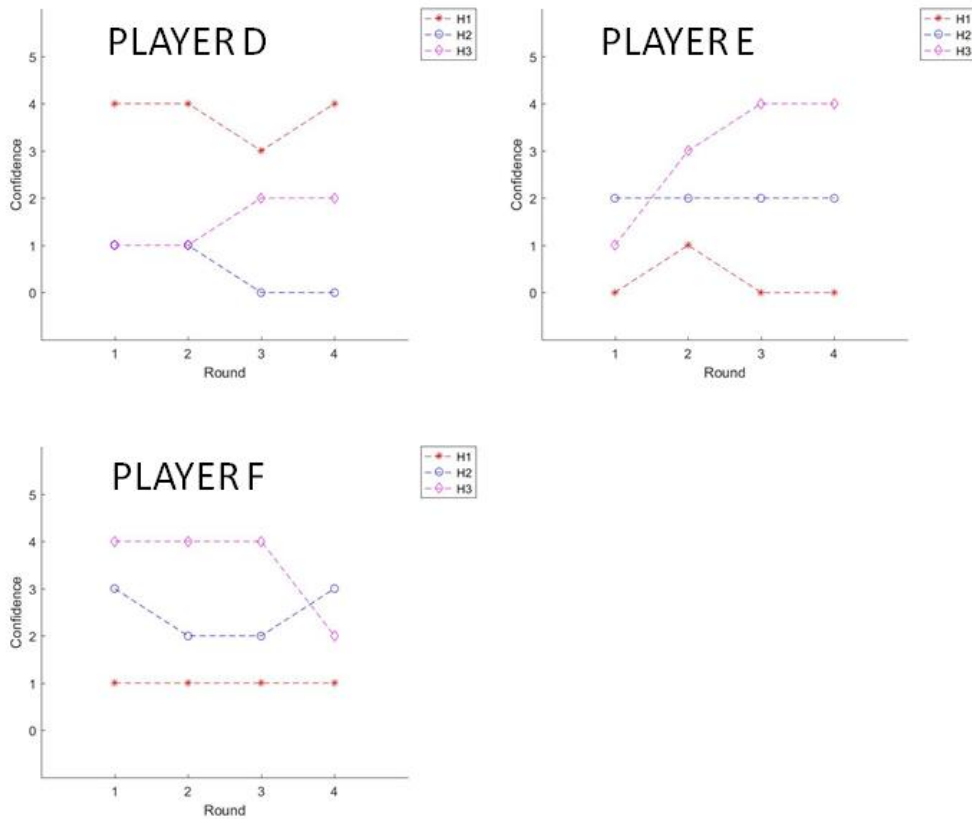


Figure 10. Example of confidence rating by different players

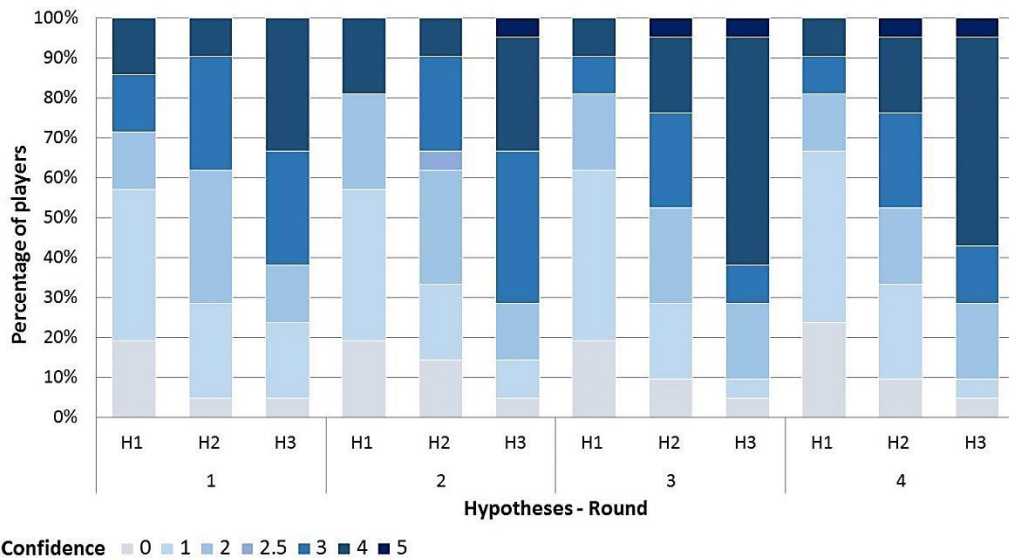


Figure 11. Confidence ratings by hypothesis in the different rounds.

5.6 Outcomes on card positions

At the end of each round a picture of the board has been taken. From the pictures the authors have been able to record all the data regarding the single card assessment. Table 5 reports

an example of the card positions in the different rounds played by one participant. It is important to underline that although this example shows only cards positioned in the points specified on the board (corners, mid of axes and center of the triangle) the players in general used the full spectrum of the possible positions, more specifically the axes displayed on the board (Figure 2).

Table 5. *Example of card positions collected for each player*

Card	Position Round 1	Position Round 2	Position Round 3	Position Round 4
1	H1	H1	H3	H3
2	H1 or H3	H1	H1	H3
3	H2	H1 or H2	H1 or H2	H1 or H2
4	H3	H1 or H3	H3	H1 or H3
5	H2	H3	H1	H1
6	H2	H1 or H3	H1 or H3	H1 or H3
7	H1 or H2 or H3	H1 or H2 or H3	H1	H1 or H2 or H3
8	H3	H1 or H3	H1 or H3	H1
9	H3	H1 or H2 or H3	H1 or H2 or H3	H3
10	H3	H3	H3	H3
11	H3	H3	H3	H3

Table 5 shows how position variations between the rounds have been consistently observed. This table reports only the final position for each round, while the card shuffling is not reported. This is because although the players have been allowed to shuffle cards during the game (GM6), this GM has been used only twice during the experiment run.

From a qualitative analysis it has been possible to observe the impact of source quality and source type on players' assessment by means of the change of the card positions on the board between the different rounds played by the same participant. However, a more in depth analysis is required to be able to quantify this impact and draw connections between those factors, the player SA and final confidence. Such an analysis requires a formalisation of belief assessment together with a proper encoding of the players' cards positions. The mathematical framework should be rich enough to capture the uncertainty expressed by the players and conveyed through the board game. Evidence theory [40] will be the favored framework for its ability to express total ignorance and non-additive assessments.

6 Conclusions

To take full advantage of the variety of information within systems that support SA, the underlying fusion processes should properly account for source factors. In order to enable this capability, research is still required with respect to the characterisation and quantification of those factors on SA. To this end, we developed a data exchange game, called the Reliability Game. The purpose of this game is to collect data regarding players' belief changes as a function of source factors, more specifically source type and quality. To gather such data each player is presented with a scenario and plays several rounds of the game. The only variation between rounds consists in his knowledge regarding source type and quality. The corresponding belief changes are captured through the variation of game items position (cards) and final confidence ratings.

We performed a qualitative analysis on the data gathered through an experiment run with a non-digital version of the Reliability Game in order to evaluate the effectiveness of the game design and game mechanics. The variations of the players' belief assessments

between the different rounds demonstrate that the proposed methodology effectively captures elements of source factors impact on SA. Moreover, the analysis allows assessing important aspects of the use of the rating scales, which might be relevant to the standardisation efforts in communication of uncertainty (e.g., confidence, reliability).

In addition to the collection of in-game data, a post-game data collection has been performed in the form of a feedback questionnaire. The results show that the game is perceived both as engaging and relevant. Moreover, the game scope and game mechanics were easily understood.

Finally, in order to be able to collect data on a larger scale, enhancing the statistical power of the designed experiments, the digital version of the Reliability Game is currently under development. However, it is expected that with such a digital version the rich interaction with the experts (players) who provided invaluable information about their steps of reasoning and decision, will be lost.

Acknowledgments

This research was supported by NATO Allied Command Transformation (NATO-ACT). The authors wish to thank Cdr. Andrea Iacono for advice and feedback during the development of the game. Moreover, the authors would like to thank Dr. David Mandel for the fruitful technical discussions on the concept of source reliability.

References

- [1] <https://www.igi-global.com/dictionary/serious-games/26549> retrieved on 24/04/2018
- [2] [2] Carvalho, M. B., Bellotti, F., Berta, R., Gloria, A. D., Sedano, C. I., Baalsrud Hauge, J., Hu, J., & Rauterberg, M., "An activity theory-based model for serious games analysis and conceptual design", *Computers and Education*, 87, pp. 166–181, 2015. <https://doi.org/10.1016/j.compedu.2015.03.023>
- [3] Rongas, B., "A Model-driven Framework for Educational Game Design", *International Journal of Serious Games*, 3(3), 2016. <https://doi.org/10.17083/ijsg.v3i3.126>
- [4] Vermillion, S., Malak, R., Smallman, R., Becker, B., Sferra, M., & Fields, S., "An investigation on using serious gaming to study human decision-making in engineering contexts", *Design Science*, 3, E15, 2017. <https://doi.org/10.1017/dsj.2017.14>
- [5] Djaouti, D., Alvarez, J. & Jessel, J.-P., "Classifying serious games: the G/P/S model", In *Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches*, vol. 1, pp. 118–136, 2011. <https://doi.org/10.4018/978-1-60960-495-0.ch006>
- [6] Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay A, Baker D, Popović Z, "Predicting protein structures with a multiplayer online game", *Nature* 466 (7307), 756–760, 2010. <https://doi.org/10.1038/nature09304>
- [7] Ren, Y., Bayrak, A. E. & Papalambros, P. Y. "EcoRacer: game-based optimal electric vehicle design and driver control using human players", *Journal of Mechanical Design* 138 (6), 2016. <https://doi.org/10.1115/1.4033426>
- [8] Endsley, R. M., "The application of human factors to the development of expert systems for advanced cockpits", In *Proceedings of the Human Factors Society 31st Annual Meeting*, pp. 1388-1392. Human Factor Society (Santa Monica, CA), 1987. <https://doi.org/10.1177/154193128703101219>
- [9] Graafland M., Schijven M., "A serious game to improve situation awareness in laparoscopic surgery", In: Schouten B., Fedtke S., Bekker T., Schijven M., Gekker A. (eds) *Games for Health*. Springer Vieweg, Wiesbaden, 2013. https://doi.org/10.1007/978-3-658-02897-8_1
- [10] Sawaragi, T., Fujii, K., Horiguchi, Y., Nakanishi, H., "Analysis of Team Situation Awareness Using Serious Game and Constructive Model-Based Simulation", *13th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems HMS 2016*, Volume 49, Issue 19, 2016, pp. 537-542 Elsevier, 2016. <https://doi.org/10.1016/j.ifacol.2016.10.617>

- [11] Endsley, R. M., "Toward a theory of situation awareness in Dynamic Systems", *Human Factors*, Vol. 37, Nr. 1, pp. 32-64, 1995. <https://doi.org/10.1518/001872095779049543>
- [12] Stikeleather, J., "Big data's human component", *Harvard Business Review*, 2012, <https://hbr.org/2012/09/big-datas-human-component> retrieved on 10/10/2017.
- [13] Endsley, R. M., "Automation and situation awareness". In Parasuraman, R., Mouloua, M. (Eds.), *Human factors in transportation. Automation and human performance: Theory and applications*, pp. 163-181, Hillsdale, New Jersey USA: Lawrence Erlbaum Associates, 1996.
- [14] Pew, R., "Human Skills and Their Utilization". In *Human Factors Engineering: Engineering Summer Conferences*. University of Michigan, Ann Arbor, 1985.
- [15] Explainable AI Systems: Understanding the Decisions of the Machines. <https://www.bbvaopenmind.com/en/explainable-ai-systems-understanding-the-decisions-of-the-machines/#.WkEZgAJbI-A.twitterHall>
- [16] D. L., Jordan, J. M., "Human-centered Information Fusion", Artech House, Boston, 2010.
- [17] Wiener, E. L., "Cockpit Automation", In *Human Factors in Aviation (Cognition and Perception)*, Wiener, E. L., Nagel, D. C. (Eds.), Ch. 13, pp. 433-461, 1988. <https://doi.org/10.1016/B978-0-08-057090-7.50019-9>
- [18] Marusich, L. R., Jonathan Z. Bakdash, J. Z., Onal, E., Yu, M. S., Schaffer, J., O'Donovan, J., Höllerer, T., Buchler, N., Gonzalez, C., *Effects of Information Availability on Command-and-Control Decision Making: Performance, Trust, and Situation Awareness*, *Human Factors*, Vol 58, Nr.2, pp. 301 – 321, January 2016. <https://doi.org/10.1177/0018720815619515>
- [19] Rogova, G., Nimier, V., "Reliability in information fusion: literature survey", In *Proceedings of the seventh International Conference on Information Fusion*, pp. 1158–1165, Stockholm, Sweden, July 2004.
- [20] Pichon, F., Mercier, D., Lefèvre, É., Delmotte, F., "Proposition and learning of some belief function contextual correction mechanisms", *International Journal of Approximate Reasoning*, Nr. 72, pp. 4–42, 2016. <https://doi.org/10.1016/j.ijar.2015.12.012>
- [21] Pichon, F., Dubois, D., Denoëux, T., "Relevance and truthfulness in information correction and fusion", *International Journal of Approximate Reasoning*, Nr. 53, pp. 159–175, 2012. <https://doi.org/10.1016/j.ijar.2011.02.006>
- [22] Haenni, R., Hartmann, S., "Modeling partially reliable information sources: A general approach based on Dempster–Shafer theory", *Journal of Information Fusion*, Vol. 7, Nr. 4, pp.361-379, December 2006. <https://doi.org/10.1016/j.inffus.2005.06.005>
- [23] Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., Jenkins, D. P.: *Human Factors Methods: A Practical Guide for Engineering And Design*. Ashgate Publishing Company, Brookfield, 2006.
- [24] de Rosa, F., Joussetme, A.-L., De Gloria, A.: *Gamified Approach in the Context of Situational Assessment: a Comparison of Human Factors Methods*. *Advances in Artificial Intelligence, Software and Systems Engineering, Proceedings of the 9th Int. Conference on Applied Human Factors and Ergonomics*, Springer (in press).
- [25] de Rosa, F., Joussetme, A.-L., "Critical review of uncertainty communication standards in support to Maritime Situational Awareness", *Technical Report*, NATO STO Centre for Maritime Research and Experimentation (in press).
- [26] "Reliability", <https://en.oxforddictionaries.com/definition/reliability>
- [27] Briñol, P., Petty, R. E., "Source factors in persuasion: A self-validation approach", *European Review of Social Psychology*, 20, 49–96, 2009. <https://doi.org/10.1080/10463280802643640>
- [28] Brathwaite, B., Schreiber, I., "Challenges for Game Designers", Charles River Media, 2008
- [29] Joussetme, A.-L., Pallotta, G., Locke, J., "A Risk Game to study the impact of information quality on human threat assessment and decision making", *Technical Rep. CMRE-FR-2015-009*, NATO STO Centre for Maritime Research and Experimentation, La Spezia, 2015.
- [30] International Organization for Standardization, *ISO 15725 Accuracy (trueness and precision) of measurement methods and results — Part 1: Introduction and basic principles*, 2011.
- [31] Yankova, K., "The Influence of Information Order Effects and Trait Professional Skepticism on Auditors' Belief Revisions", Gabler Verlag, 2015. <https://doi.org/10.1007/978-3-658-08871-2>
- [32] "Confidence", <https://en.oxforddictionaries.com/definition/confidence>
- [33] "Energy Performance Certificate", https://en.wikipedia.org/wiki/Energy_Performance_Certificate
- [34] Private conversation with intelligence analyst

- [35] Osborne, J. W., Blanchard, M. R., "Random Responding from Participants is a Threat to the Validity of Social Science Research Results", *Frontiers in Psychology*, Vol. 1, Article 220, January 2010.
- [36] Brooks, J. L., "Counterbalancing for serial order carryover effects in experimental condition orders", *Psychological Methods*, Vol. 17, Nr. 4, pp. 600–614, December 2012.. <https://doi.org/10.1037/a0029310>
- [37] Millar, A.G., "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", *Psychological Review*, Vol. 101, Nr. 2, pp. 343–352, 1956. <https://doi.org/10.1037/h0043158>
- [38] McVay, J. C., Kane, M. J., "Conducting the Train of Thought: Working Memory Capacity, Goal Neglect, and Mind Wandering in an Executive-Control Task", *Journal of Experimental Psychology, learning, Memory, and Cognition*, Vol. 35, Nr.1, pp. 196-204, January 2009.. <https://doi.org/10.1037/a0014104>
- [39] Wilson, D. W., Jenkins, J., Twyman, N., Jensen, M., Valacich, J., Dunbar, N., Wilson, S., Miller, C., Adame, B., Lee, Y.-H., Burgoon, J., Nunamaker, J. F., "Serious Games: An Evaluation Framework and Case Study", 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, 2016.. <https://doi.org/10.1109/HICSS.2016.85>
- [40] Shafer, G., "A Mathematical Theory of Evidence", Princeton University Press, 1976.