Data Sources and the Uncertainty of Information (UoI) Experimental Research

Adrienne Raglin[1], *Aleah Emlet*[3], *Justine Caylor*[2], *John Richardson*[2], *Mark Mittrickr*[2], *Somiya Metu*[1]

[1] Army Research Laboratory
Adelphi, MD 20783, USA

[2] Army Research Laboratory
Aberdeen Proving Ground, MD 21005, USA

[2] Pennsylvania College of Technology
Williamsport, PA 17701, USA

ABSTRACT

Information plays a vital role in the courses of action taken on the battlefield. In the era of increased available data and a wide variety of sensors there are various potential sources of rich data that can be collected. However, the uncertainty related to the sources of the data remains a challenge that needs to be addressed for supporting actions and decisions. Sources of data can include categories such as information, network, devices, and visualization. To address this challenge, DEVCOM ARL researchers conducted an experimental user study with 26 soldiers of varying military occupational specialties (MOSs) and years of service (YOS) to get there input. For the user study two scenarios were presented to uncover potential patterns of concerns soldiers' may have related to uncertainty of information (UoI). This allows additional study to better understand soldiers' decision-making processes particularly when considering selected categories of sources of information and associated uncertainty.

Intelligent Human Systems Integration (IHSI 2022): Integrating People and Intelligent Systems
https://openaccess.cms-conferences.org/#/publications/book/978-1-7923-8988-7

The results of the experiment will be used to continue the research into artificial reasoning for creating more adaptable and trustworthy intelligent systems to aid the warfighter.

Keywords: Uncertainty of Information, Data Sources, Artificial Intelligence, Decision-Making, Big Data, User Study

INTRODUCTION

Making decisions is a complex process that relies upon knowledge, understanding, and judgement across many variables. However, any data used to make rational and realistic decisions often is imperfect. Parsons states that "it is necessary to learn how to represent and reason with imperfect information."(Parsons, 1996) Thus this idea is important throughout the decision-making process, including comparing actual to desired outcomes, determining what information is needed, and assessing the risks between decision alternatives.

Current research on the Uncertainty of Information (UoI) concept (Raglin, Metu, & Lott, 2020) includes exploring how to create profiles for the decision makers that can be used for assessment and for guiding recommended decisions. A user study experimentation was created with a series of questionnaires to capture how Soldiers would prioritize various aspects of uncertainty based on the source category of the data and on the uncertainty, taxonomy inspired by Gershon's work on imperfect information (Gershon, 1998).

UNCERTAINTY OF INFORMATION (UOI)

The Uncertainty of Information (UoI) can be expressed as a value that is dependent on descriptors that express uncertainty based on the nature of imperfections within information as presented in Gershon's paper (1998). These descriptors form a taxonomy that attempt to capture the causes and express the type of uncertainty for a given source. Currently the taxonomy consists of any uncertainty that is inconsistent, corrupt, disjoint, incomplete, imprecise, complicated or questionable.

Inspired by the taxonomy the initial approach to the UoI expression is a weighted sum as seen in Equation 1:

$$UoI_{dp} = \sum_{a,b=1}^{k,l} T_{a,b} * S_{a,b} + \sum_{c,d=1}^{m,n} W_{c,d} * D_{c,d} \sum_{a,b=1}^{k,l} G_{a,b} * S_{a,b} + \sum_{c,d=1}^{m,n} W_{c,d} * D_{c,d} \quad (1)$$

where *dp* is a decision point, *D* are variables that express components of the decision making that may be key factors for the task, *W* are the weights associated with the importance of those components, *T* are the categories of taxonomy weights (equivalent to *G*) and *S* are the categories of sources of data and information. The UoI

value represents the contributions of the sources and factors in relationship to the uncertainties categorized.

EXPLANATION OF THE DATA SOURCES

As mentioned previously we have selected four main categories for sources. The following describes examples of uncertainty for the categories:

1. Heterogeneous Devices: A heterogeneous device can have an associated uncertainty where there is no guarantee that the device functions properly at all times which can lead to issues of unreliability. In addition, issues with the device performance from device malfunctions or antiquated devices can cause inconsistencies.

2. Information: Associated uncertainty for information can be tied to missing data or corruption whether from a report, sensor or person.

3. Network: Information that is dependent on a networked platform and its ability to transmit that data. Therefore, any instances where the network may be down or intermittent connectivity can create possibilities for data omission or corruption due to its unreliable performance.

4. Visualization - Visualization of the data is an important factor for comprehension. If a visualization mechanism is not the appropriate one and may fail to convey the right message can lead to uncertainties or questions in understanding of information.

USER STUDY

The user study that was mentioned previously consisted of a survey that asked each participant to rank the category sources giving the highest number to the source that would be the most important or concerning if it had any uncertainty. The rankings ranged from 1 (least concerning) to 4 (most concerning). The participants included twenty-six soldiers from Ft Irwin ranging in military occupational specialty (MOS) and years of service (YOS). To provide context for participants' rankings, two predefined scenarios were provided. Scenario 1 portrayed a convoy escort whose task was to protect a supply convoy being transported from the forward operating base (FOB) to a city. A pre-brief indicated few attacks and threats. To accomplish this task, they needed to rely on different sources of information for decisions and they needed to be mindful of the UoI from the four sources. Scenario 2 displayed a more complex situation, in which the Soldier was tasked with defending a valley with several farming villages from surrounding militia. Threat reports incoming from multiple sources were to be used and they must also determine how to place defenses and manage patrols to mitigate the risk.

RESULTS

The following section highlights the observed trends by the participants of the user study for each scenario.

Scenario One

Thirteen participants ranked Information and nine participants ranked Network as four, the most concerning sources. Only three participants ranked Visualization and one participant ranked Device as four. Visualization was ranked as two from ten participants. Device was ranked as one from nine participants. Fig. 1 summarizes the overall rankings for scenario 1.

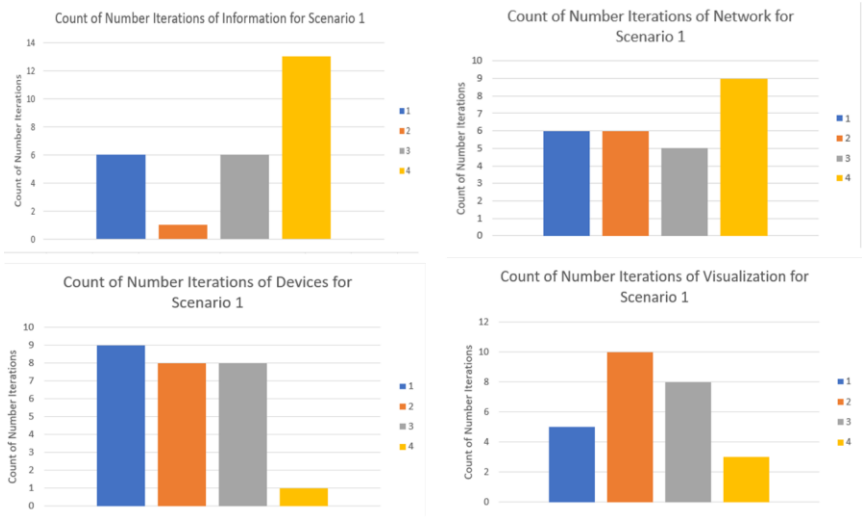

Fig. 1. Summary of overall source rankings for scenario 1

Now that we have explored how the sources were ranked, we will consider how each individual source was ranked by participants' MOS and YOS.

For the results associated with the information source by MOS for scenario 1, we selected specific ones to consider to begin to explore if there is a difference in decisions and related uncertainty that may be tied with the MOS. For infantry, 11B, three out of the four participants with this MOS ranked information at four. For medic, 68W, three out of the four ranked information at four. For the results by YOS we also selected ones to consider as years of service and indirectly experience may influence decisions and related uncertainty as well. For those with 14 years, two out of the four participants ranked information as four and those with 21 years, both ranked information as three. Fig. 2 summarizes the rankings of the information source by MOS and YOS for scenario 1.

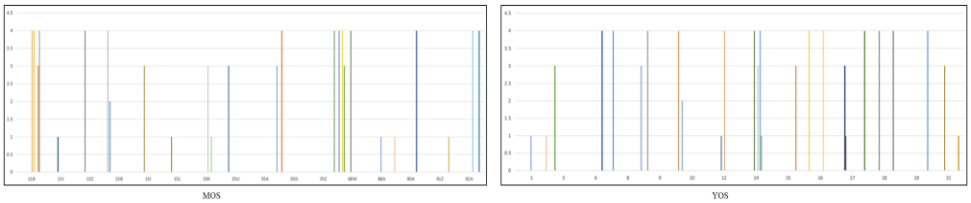

Fig. 2. Information source by MOS and YOS (Scenario 1)

For the network source, the results by MOS for scenario 1 for 11B, one out of the four participants with this MOS ranked network as four. For 68W, one out of the four ranked network as four. The results by YOS for scenario 1 with participants with 14 years, two out of the four participants ranked network as four while those with 21 years, one ranked network as four. Fig. 3 summarizes the rankings of the network source by MOS and YOS for scenario 1.

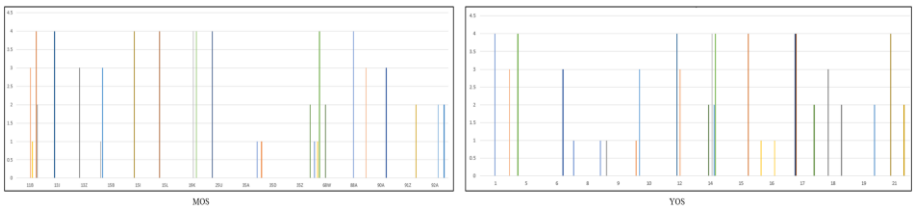

Fig. 3. Network source by MOS and YOS (scenario 1)

For the device source, the results by MOS for scenario 1 for 11B, two out of the four participants ranked devices as two, for 68W, two out of the four ranked devices as three. The results by YOS for scenario 1 with participants with 14 years, one out of the four participants ranked devices as three while those with 21 years, one out of the two ranked devices as three. Fig. 4 summarizes the rankings of the devices source by MOS and YOS for scenario 1.

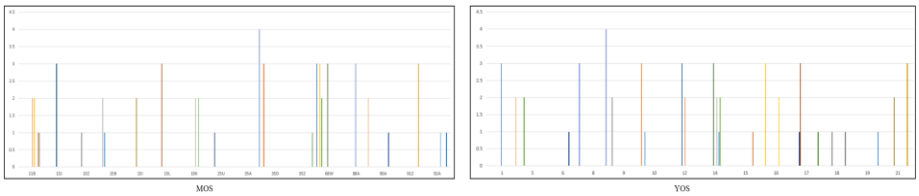

Fig. 4. Device source by MOS and YOS (scenario 1)

For the visualization source, the results by MOS for scenario 1 for 11B, three out of the four participants ranked visualization as three while 68W, two out of the four ranked visualizations as two. The results by YOS for scenario 1 with participants with 14 years, one out of the four participants ranked visualization as three while those with 21 years, one out of the four ranked visualization as four. Fig. 5 summarizes the rankings of the visualization source by MOS and YOS for scenario 1.

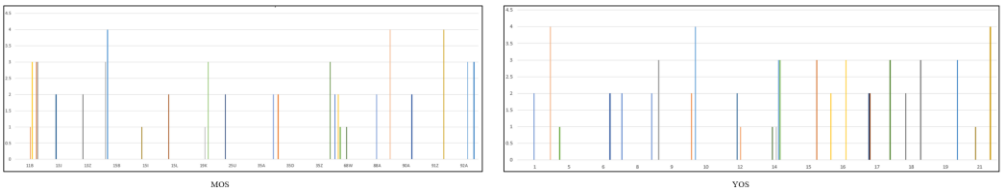

Fig. 5. Visualization source by MOS and YOS (scenario 1)

Scenario Two

For Scenario 2, ten participants ranked Information as three, nine participants ranked Device as two, and ten participants ranked visualization as one, indicating least concerning. Network tied rankings at 2 and 4, highlighting that some participants viewed the Network at different levels of importance or concern. Fig. 6 summarizes the overall rankings for scenario 2.

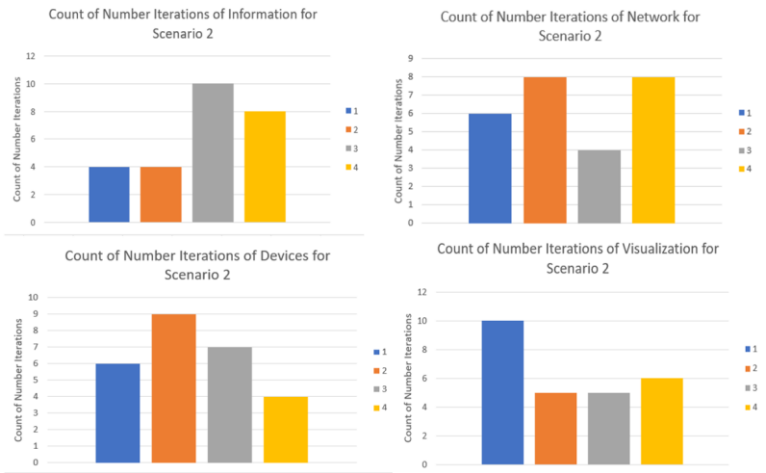

Fig. 6. Summary of overall source rankings for scenario 2

For information source for scenario 2, the 11Bs, one out of the four participants ranked information as four, while for 68W, two out of the four ranked information as four. For the participants with 14 years, two out of the four participants ranked information as three, while those with 21 years, two out of the two ranked information as three. Fig. 7 summarizes the rankings of the information source by MOS and YOS for scenario 2.

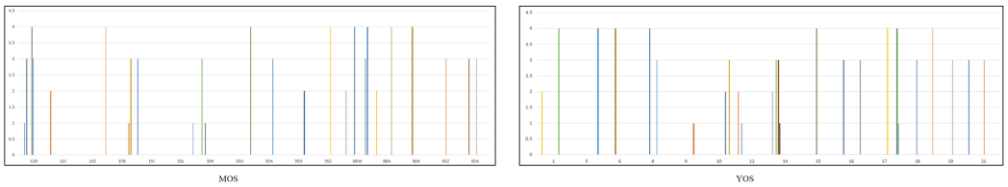

Fig. 7. Information source by MOS and YOS (scenario 2)

For network source the 11Bs, one out of four participants ranked network as four, while for 68W, one out of the four ranked network as four. For participants with 14 years, two out of the four participants ranked network as four, while those with 21 years, one out of the two ranked network as four. Fig. 8 summarizes the rankings of the network source by MOS and YOS for scenario 2.

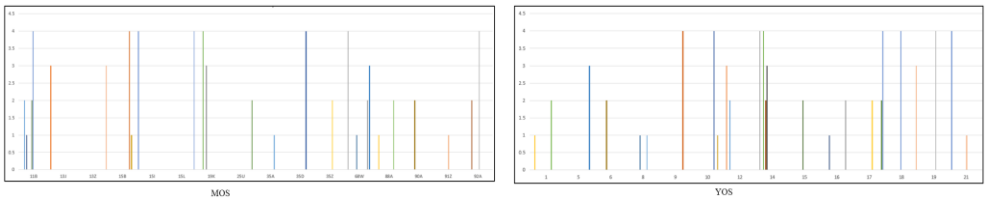

Fig. 8. Network source by MOS and YOS (scenario 2)

For device source, the 11Bs, one out of the four participants with this MOS ranked devices as four, while for 68Ws, two out of the four ranked devices as two. For participants with 14 years, one out of the four participants with this YOS ranked devices as four while for 21 years, two out of the two ranked devices as two. Fig. 9 summarizes the rankings of the devices source by MOS and YOS for scenario 2.

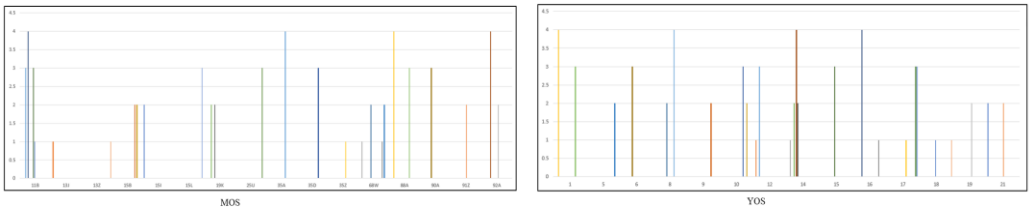

Fig. 9. Devices source by MOS and YOS (scenario 2)

For visualization source, the 11Bs, one out of the four participants with this MOS ranked visualization as four. For 68Ws, one out of the four ranked visualizations as four also. For participants with 14 YOS, one of the four participants ranked visualizations as four while for 21 years, one of the two participants ranked devices as four. Fig. 10 summarizes the rankings of the visualization source by MOS and YOS for scenario 2.

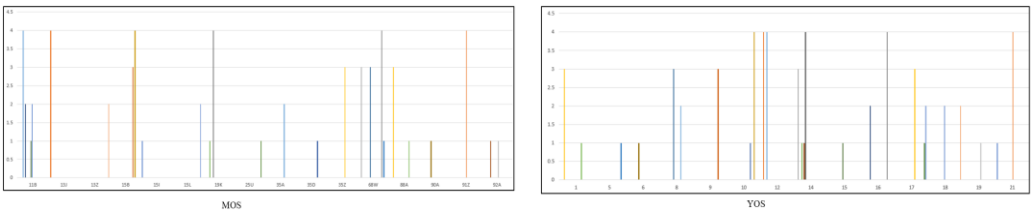

Fig. 10. Visualization source by MOS and YOS (scenario 2)

CONCLUSIONS

From this initial analysis of the results of this experimental user study, information is ranked the most important by the most participants. This was expected as the other source categories were more tied with performance. However, while performance has an impact on the data, the data itself significantly drives decisions. Due to the interaction of data and sources, we will explore how the UoI computational model can address this dependent influences. As we move forward, we will explore defining subcategories under information to begin to address this challenge and those related to modalities such as image, text or audio. In conclusion, for the results between scenarios, we noticed that information followed by devices, networks, and visualization were the general rankings of the sources. As MOS and YOS was considered, information again was ranked the most important. Additional analysis will continue to uncover any potential trends across MOS and YOS as well as utilizing this data for assigning profiles for agents that will be proxy decision makers in simulations.

ACKNOWLEDGMENTS

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-19-2-0062. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

Gershon, N. (1998). Visualization of an imperfect world. IEEE Computer Graphics and Applications, 18(4), 43-45.

Parsons, S. (1996). Current approaches to handling imperfect information in data and knowledge bases. IEEE Transactions on knowledge and data engineering, 8(3), 353-372.

Raglin, A., Metu, S., & Lott, D. (2020, July). Challenges of Simulating Uncertainty of Information. In International Conference on Human-Computer Interaction (pp. 255-261). Springer, Cham.