Examining the Effectiveness of Storyline Visualisation for Understanding Wargame Scenarios: A Pilot Study

ANDREW CUNNINGHAM, University of South Australia, Australia
STEVEN WARK, Defence Science and Technology Group, Australia
BEN VOLMER, University of South Australia, Australia
MARCIN NOWINA-KROWICKI, Defence Science and Technology Group, Australia
WOLFGANG MAYER, University of South Australia, Australia
SOPHIE UNDERWOOD, Defence Science and Technology Group, Australia
GREG O'KEEFE, Defence Science and Technology Group, Australia
SLAVA SHEKH, Defence Science and Technology Group, Australia

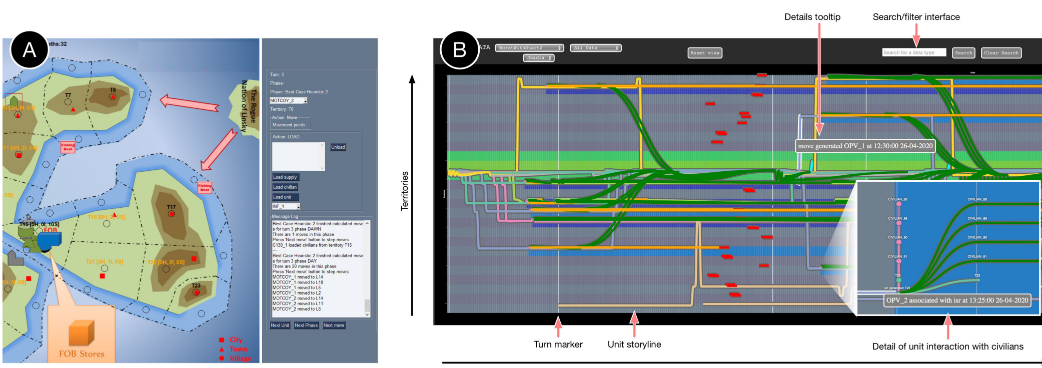

Fig. 1. The two conditions examined in a user study pilot to determine the effectiveness of storyline visualisations for wargame scenarios. A) The Disaster at Joadia Islands wargame developed within the Defence Science and Technology Group. B) A storyline visualisation of a play through of the "Joadia" game. Territories in Joadia are mapped to the vertical axis, time is mapped to the horizontal axis. Units within the game are represented by "storylines" running horizontally.

Authors' addresses: Andrew Cunningham, andrew.cunningham@unisa.edu.au, University of South Australia, Adelaide, South Australia, Australia; Steven Wark, steven.wark@defence.gov.au, Defence Science and Technology Group, Edinburgh, South Australia, Australia; Ben Volmer, ben.volmer@unisa.edu.au, University of South Australia, Adelaide, South Australia, Australia; Marcin Nowina-Krowicki, marcin.nowina-krowicki@defence.gov.au, Defence Science and Technology Group, Edinburgh, South Australia, Australia; Wolfgang Mayer, wolfgang.mayer@unisa.edu.au, University of South Australia, Adelaide, South Australia, Australia; Sophie Underwood, sophie.underwood@defence.gov.au, Defence Science and Technology Group, Edinburgh, South Australia, Australia; Greg O'Keefe, greg.okeefe@defence.gov.au, Defence Science and Technology Group, Edinburgh, South Australia, Australia; Slava Shekh, slava.shekh@defence.gov.au, Defence Science and Technology Group, Edinburgh, South Australia, Australia.

XXXX-XXXX/2024/11-ART $15.00
https://doi.org/XXXXXXX.XXXXXXX

Wargames are a training tool for developing strategic knowledge and approaches to complex situations. Understanding the wargame scenario for review and post-game analysis is an important aspect of the learning process. We have developed a storyline visualisation for a wargame scenario that depicts the units, their actions, and interactions overtime in a single view. We ran a quantitative pilot user study ($N = 4$) comparing storyline visualisation to the more common geospatial map of approached used for wargame scenarios. Our early pilot results suggest storyline visualisation was both faster and more accurate for complex questions. We also found that participants subjectively rated a lower level of effort required when using the storyline visualisation.

CCS Concepts: • **Human-centered computing → Visualization systems and tools**; **Visualization techniques**; Empirical studies in interaction design.

Additional Key Words and Phrases: Visualisation, Wargame, Storyline Visualisation, Geospatial

ACM Reference Format:
Andrew Cunningham, Steven Wark, Ben Volmer, Marcin Nowina-Krowicki, Wolfgang Mayer, Sophie Underwood, Greg O'Keefe, and Slava Shekh. 2024. Examining the Effectiveness of Storyline Visualisation for Understanding Wargame Scenarios: A Pilot Study. 1, 1 (November 2024), 12 pages. https://doi.org/XXXXXXX.XXXXXXX

Wargames are an essential tool for training strategies for campaigns, force design, and for teaching fundamental concepts. Spatial wargames simulate a scenario in a multi-dimensional space; units are moved on a *two-dimensional* map over *time*. Whilst these spatial wargames are naturally represented as units moving on a map, we are interested in communicating a higher-order of interactions that may occur over time in a wargame; where the cause is several steps removed from the effect. These higher-order interactions in wargames can be difficult to represent on a typical map as the chronology of events must be communicated through sequential animation or similar techniques that can obfuscate critical aspects of the context. Communicating this context and understanding the lifetime of a particular unit may lead to new insights.

Storyline visualisations are a form of narrative visualisation that depict entity interactions over time by projecting timelines of those entities interactions onto a horizontal axis. Stories aim for coherence and cohesion, making them effective frameworks for integrating multiple dimensions of data—including space, time, and causality—and thereby better support the user's mental model of a complex situation [14], as may be found in wargames. This motivates our research question: *how can storyline visualizations support understanding of the lifespans and interactions of units in wargame scenario data?*

The "Disaster at Joadia Islands" (referred to as "Joadia") is a tabletop wargame developed by the Defence Science and Technology Group [7]. Joadia is an ideal testbed to examine storyline visualisation for wargame scenarios as: 1) the results are non-deterministic; 2) the system has been augmented with intelligent agents that can generate scenario data; and 3) the action space is relatively understandable within a reasonable amount of time. We built a storyline visualisation system for visualising geospatial data, and in particular, wargame scenario data from Joadia. To answer our research question, we designed and ran a pilot study comparing a geospatial view to a storyline visualisation.

The contributions of this work are:

- A question taxonomy for geospatial wargame comprehension;
- A design for a storyline visualisation addressing the specific domain of wargame scenarios; and,
- Results of a pilot evaluation ($N = 4$) suggesting that there is value in further investigating storyline visualisations for certain types of comprehension in spatial scenarios.

The rest of this publication is structured as follows: we first present related work in geospatial and storyline visualisation. We then describe the Joadia wargame at a high-level. We proceed to describe the visualisation we developed to visualise wargame scenarios. We then describe our study design and results, finishing with a discussion and limitations.

1 BACKGROUND

This work and the subsequent user study focuses on *geospatial* wargames. In this background section, we first detail prior work in geospatial visualisation. We also examine more abstract representations of geospatial data, where dimensions in the data are mapped to non-Euclidean mappings. This leads into a discussion of storyline visualisations, which map abstract spatial-relationships (or interactions) between entities onto a single axis and is the primary technique we apply to the wargame scenarios in this work.

Wargames are a broad category of strategic games used by defence (and other domains) as learning tools to understand responses to complex scenarios. They are often team-based adversarial games with at least two teams (referred to as "blue team" and "red team"), but can be non-adversarial, e.g., for bush fire response. Wargames are also often tabletop, physical games, due to the low-cost, the potential for a large action space (since the actions are facilitated by a human) and the support for dynamic interactions. Digital wargames can be used in cases that require complex simulation of systems, or for testing artificial intelligence solutions.

1.1 Geospatial visualisation

Although two-dimensional geotemporal visualizations that use trajectory representations are still regularly employed, the complexity of the information has increased, leading to additional uncertainty when correlating the various datasets [9]. The Space-Time Cube (STC) is a technique using three-dimensional projection to directly represent both 2D space and time [5]. Such techniques can illicit deeper insight into trajectory based data. However, many of these works depend on 3D projections onto 2D displays, which introduce occlusion, depth perception and spatial understanding issues.

Walsh et al. [13] presented an abstracted visualisation technique that represented relative distance from a point-of-reference for a set of geospatial entities over time. The authors compared the technique to the STC and found the proposed technique outperformed the STC for the particular comprehension questions tested. The authors also found a significant effect of question complexity, suggesting that abstracted visualisations may improve comprehension for more complex situations.

1.2 Storyline visualisation

Storyline visualisations are a form of narrative visualisations that depict entity interactions over time by projecting timelines onto an axis. The archetypal storyline visualisation is depicted in Munroe's [8] hand-drawn illustration "Movie Narrative Charts". Munroe visualises the narratives of five distinctive movies. Each entity's (or group of entities, depending on how they are treated in the narrative) storyline within the movie is represented by an edge running from left to right depicting the passage of time. Interactions between entities over time are represented by edges moving adjacent to each other.

Recent storyline visualisation work has examined how to automatically generate these storyline visualisations from curated data, using Monroe's Movie Narrative Charts as a benchmark for readability. Tanahashi and Ma [10] presented a set of codified design considerations for creating these storyline visualisations as well as a layout algorithm based on evolutionary computation that realises these design considerations.

There are few evaluations of the effectiveness of storyline visualisation, and the majority of them are qualitative studies garnering subjective feedback from users. Tang et al. [12] found that automatic storyline visualisations often produce inferior results compared to hand-drawn storyline visualisations, and proceeded to define a design space for storyline visualisation. They created a tool to author storylines using the design space, and then qualitatively evaluated the system with positive results. Later, Tang and Li et al. [11] evaluated a reinforcement learning layout for storyline visualisations and found their technique produced favourable results according to design experts. Baumgartl et al. [3] applied storylines to a professional application (e.g. pathogen outbreaks in hospitals) and evaluated the system through expert review. They found that the visualisation afforded experts the ability to "see more" of their data, and experts acknowledged that the tool would be useful in their day-to-day work. Arendt and Pirrung [2] performed a quantitative user study comparing two storyline layouts and found that layouts that prioritise reducing edge-crossings performed worse than those that provide a consistent interpretation of the vertical axis.

2 WARGAME SCENARIO

To answer our research question, we required a suitable wargame from which we could extract a scenario for visualisation. The "Disaster at Joadia Islands" (referred to as "Joadia") is an in-house tabletop wargame developed by the Defence Science and Technology (DST) Group. Joadia is a turn based game simulating humanitarian assistance in the aftermath of a crisis in which a fictitious set of islands (the eponymous Joadia islands) has been affected by a tsunami. Players control either the Blue team, representing the Australian Defence Force (ADF), or the Red team, representing an antagonistic force attempting to disrupt the Blue team's effort. As the Blue team, the player's goal is to both maximise the number evacuees and minimise the number of fatalities while dealing with uncertain information and limited resources. Blue team has a number of army, air force, and navy platforms under their control that can perform warfighting tasks including airlifts, intelligence intelligence surveillance reconnaissance (ISR) missions, and engaging enemy units. During a turn, the Blue team moves units, reveals civilian populations, heals injured civilians, and ensures civilians have food, while attempting to evacuate healthy refugees on limited capacity C130 aircraft. Players on the Red team attempt to disrupt the Blue teams' efforts by using undercover boats to land Militia zealots onto Joadia.

Internal efforts at the DST have ported the Joadia wargame to a computer program developed in Python that also serves as a platform for testing intelligent agents [7]. This program serves as our baseline for wargame visualisation. Joadia is a compelling platform for testing intelligent agents as it is non-deterministic and is partially-observable, which could indicate that the actions taken by the intelligent agents are viable candidates to visualise in order to understand why certain actions happened.

3 WARGAME STORYLINE VISUALISATION

A workshop was held with experts in wargaming using an early concept demonstrator of the storyline visualisation system. The purpose of the workshop was to elicit expert feedback on requirements for such a system. From the workshop results, we focused on the following requirements specific to the wargame visualisation:

RQ1 — Game turns must be clearly represented.
RQ2 — Users must be able to identify territories within the visualisation.
RQ3 — The system must support understanding large datasets.
RQ4 — The system must be able to show the end state for certain entities (e.g. civilians).

RQ5 — Units must have continuous storyline renderings without breaks. A user must be able to follow a storyline of a unit from start to finish.

RQ6 — Surviving civilians (versus injured or dead) determine the outcome of the game. Therefore, civilians and their state should be clearly communicated.

In the following section, we describe the visualisation design at a high-level and identify how these requirements are addressed.

3.1 Visualisation design

We took a node-link approach to rendering the storylines. Units and actions are internally represented as nodes, the relationship between units and actions are the links. The storyline visualisation developed is shown in Figure 1(B). The visualisation was developed using D3 [4] and a significantly adapted D3-sankey library[1]. Units are represented by filled-circle glyphs. Units are uniquely coloured on a distinct colour palette to support quick identification. Healthy civilians are uniquely coloured dark-green and unhealthy are coloured orange for quick identification (requirement **RQ6**). Actions taken by the units are distinguished by square-glyphs labelled with the activity's behaviour. To support requirement **RQ3**, labels on activities change depending on the zoom scale of the visualisation; at a 0.5 zoom factor or more, the labels display a full description of the action (e.g. "entity1 generatedBy load-1"), at smaller zoom factors the label only displays the action name (e.g. "load-1").

Layout: Units and the actions they take are represented as nodes. Storylines for a unit are rendered as Bezier curve "links" between the source and destination nodes (requirement **RQ5**). To address requirement **RQ2**, the data is faceted by territory; each territory is afford a vertical band in the visualisation and unit and action nodes are positioned vertically within the territory that they take place. Furthermore, units are allocated unique rows within the territory which helps ensure continuity and reduces distracting line movement. Nodes are then positioned horizontally by the time that they took place. Times are tightly packed, so that if nothing occurred during a particular timebox, it does not take up screen space. Turns were clearly represented by vertical white dividers (requirement **RQ1**).

3.2 Interactions

Several interactions are present within the visualisation system. We categorise these interactions into view transformation, details on demand, and filtering. The system integrated the following view transformations:

Pan — Allows the user to view different parts of the graph by clicking and dragging in any direction.

Zoom — The view can be zoomed by scrolling the mouse wheel. Nodes display different text based on how far the user is zoomed.

Click to navigate — When zoomed, links can span the view, hiding start and end nodes. Clicking on the links moves the user's view to different nodes of the link. This allows the user to follow the story of any node they wish. Left-clicking a link will move the view to the source node; right clicking a link will move the view to the target node.

Reset view — With large datasets it is possible for the user to lose their frame of reference. The reset view button will reset the pan and zoom to the original position of the graph.

Details on demand is provided by popup information panels that displays information to the user about that data element:

[1]https://github.com/d3/d3-sankey

Links — Displays the source node, relationship type, target node and if available, the time of the task that happened.

Nodes — Fades every other node and link aside from the highlighted and the links the node is related to. The tooltip will also display information based on the type of node.

Entity — The tooltip displays the location of the data element alongside any additional information based on the entity. For example, a civilian will display their health status and a vehicle might display the civilians which are located inside it.

Activity — The activity tooltip displays the turn number and start/end time the activity took place. Evacuate also displays the civilians who evacuated.

Filtering took one primary method:

Search Filter — The user can type a string to search in the search box in the top-right of the window. Entities (or activities) with matching names will appear in the visualisation, while other entities and activities will be hidden. We do not recalculate the layout so as to ensure the user has a consistent visual frame of reference between the filtered and non-filtered view.

4 PILOT STUDY

We designed and ran a user study to examine how effective storyline visualisations are for understanding the events and interactions of units within a wargame scenario with performance based on comprehension and recall. The study is designed as a within-participants study, with a participant viewing both the default Joadia user interface system (referred to as the Baseline) and the storyline visualisation system. The study design was approved by the university Human Research Ethics Committee (approval #203580).

4.1 Conditions

Two conditions were examined in the study:

Baseline: The baseline condition replicates a typical rendition of a geospatial wargame. A map of the game world is presented with units moving on the map. The different types of units in the gameworld are visually distinguishable (Figure 3).

Storyline: This condition is the storyline visualisation described in section 3 and depicted in Figure 1(B). This condition is as described in section 3 with one exception, we did not teach the "click to navigate" interaction—we wanted to keep interactions to a minimum so as to not bias towards the storyline visualisation.

4.2 Tasks

To test comprehension and recall, we designed a question taxonomy based on Amini et al.'s [1] geospatial question taxonomy. Amini et al. present a taxonomy that encodes a question of a spatial dataset (space, time, entity) into a measure of complexity based on the plurality and unknown aspects of the question. For example, "which blue unit was in region 14 at turn 7" is a relatively simple question as it requires observing a particular region at a particular time to produce an answer. A more complex question would be "how did the refugees at zone 14 survive until evacuation?", as this question requires reviewing the entire timeline of zone 14 and observing interactions over that time. By being able to measure a questions effective complexity, we seek to identify whether there are interaction effects in our experiment variables.

4.3 Hypothesis

Since storyline visualisations represent the entire lifetime of an entity visually in a single image, we expect storyline visualisations to be better for tasks where the user must comprehend and search the lifespan of entities. So for answering questions covering the lifespan of entities within a scenario, we will test the following hypotheses:

H1: **Speed** — Storyline visualisation will have significantly faster comprehension time than Baseline for complex questions.
H2: **Accuracy** — Storyline visualisation will have significantly higher comprehension accuracy than Baseline for complex questions.

We also expect that the storyline visualisation will help users build a better mental-model of the narrative and therefore improve recall:

H3: **Recall** — Storyline visualisation will have significantly higher recall for questions and answers.

We expect to see the storyline visualisation perform better than the standard visualisation on more complex questions as the user's mental model (being a situational model encoding space, time, entities, and cause) is better supported.

4.4 Questions

Given the question taxonomy, we established the following questions to test our hypotheses:

Q1: Which unit visited the most unique regions in the scenario?
Q2: How many times did both MOTCOY_1 enter territory 15?
Q3: How many civilians did the MHR save during the scenario?
Q4: How many unique regions did MOTCOY_1 visit?
Q5: How many unique regions did MOTCOY_2 visit?
Q6: Which regions did MRH_1 visit between turns 1 and 5 inclusive?
Q7: Which units perform ISR in turn 3?
Q8:Which territories does FishBoat_6 visit during turn 3?

These questions covered a range of complexity on the question taxonomy depicted in Figure 2.

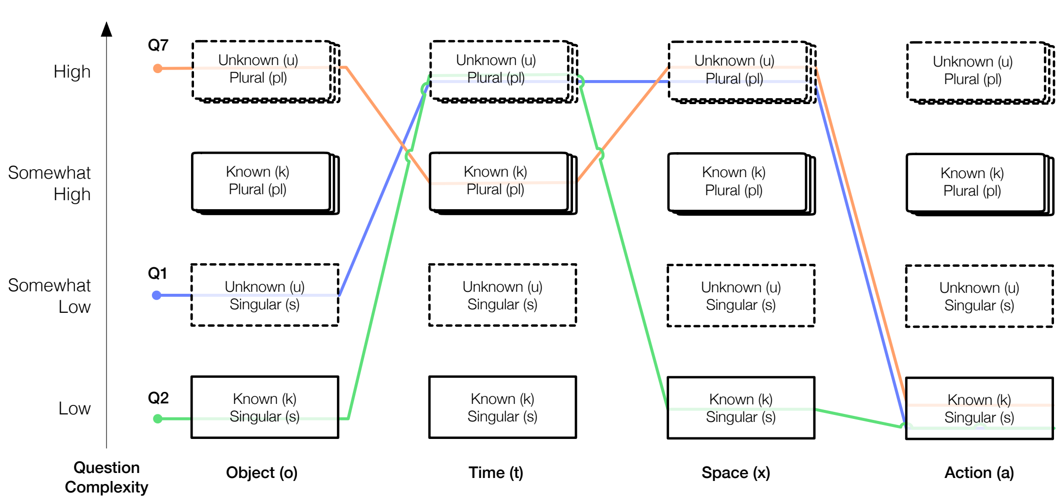

Fig. 2. The final question taxonomy developed for this study. Three exemplar questions identified in sec:questions (Q1, Q2, and Q7) are indicated by lines on the diagram.

4.5 Study Design and Measures

The study was designed as a full factorial within-participants study, with participants completing repeated trials that covered all system conditions and questions. Each trial required the participant to answer a question about the presented scenario. The experimental design was 2 (visualisations) × 8 (questions). The ordering of the visualisations was counterbalanced. Learning effects from question ordering was mitigated using a partially balanced latin square.

We recorded four dependent variables from the trials:

- *Task Completion Time*: the total seconds it took to answer the question.
- *Trial Attempt Error*: the number of attempts made to to answer the question (successful or otherwise). Participants were allowed three attempts before being moved to the next question. As a result, the minimum value was 1 and the maximum value was 4 (representing that they never answered correctly).
- *Task Failure*: whether the question was answered correctly within the maximum of three attempts.
- *Recall*: Short-term recall was measured 10 minutes after each condition. Participants were asked to freely recall the questions and their associated answers. Participants were prompted with any questions they could not recall. The final score was the sum of the following: free-recall of a question and answer was scored 2, prompted recall of an answer was scored 1. Partial scores were awarded for remember parts of a question or answer.

We also recorded two subjective measures. Participants answered a Single Ease Question (SEQ) after each condition, which captures task difficulty on a scale of 1 (Very difficult) to 7 (Very easy). Participants also completed a NASA Task Load Index (NASA TLX) questionnaire to measure perceived workload of the two conditions [6].

4.6 Apparatus

The user study was run on a MacBook Pro 2017 with 2.6Ghz Intel Core i7 and 16GB RAM. The MacBook Pro was connected to a Razer Core X with an AMD Radeon RX Vega 56 driving two Dell 1440P displays at 60Hz. A third display was setup so that the facilitator could monitor the participants' actions.

We developed a tool for visualising the baseline condition (Figure 3). The tool was developed using Unity 2018.4.4f1. In this tool, a geospatial map could show the units and their location for a given point of focus in the scenario. A scrollbar below the maps could change the point of focus, allowing participants to "scrub" through the entire scenario. Buttons were also available to step accurately forward or backward through the scenario.

We developed a second tool for facilitating the questions and answers, also using Unity 2018.4.4f1. The tool presented participants with the question. Each question was tagged as either a number, unit, or territory answer. For number answers, participants could enter free text. For unit or territory answers, the tool would present participants with checkboxes of all the named entities. The tool identified how many chances the participant had left on a particular question. If the participant answered incorrectly, the window would pulse red and the participant would get another turn to answer (up to three times). Between turns, the tool presented a pause window to provide participants a break if needed.

4.7 Procedure

The experiment procedure consisted of five stages: (1) pre-study questionnaire and introduction, (2) training, (3) exploration, (4) comprehension, and (5) post-condition questionnaire and recall. In Stage 1, participants answered a pre-study questionnaire covering basic demographics and prior

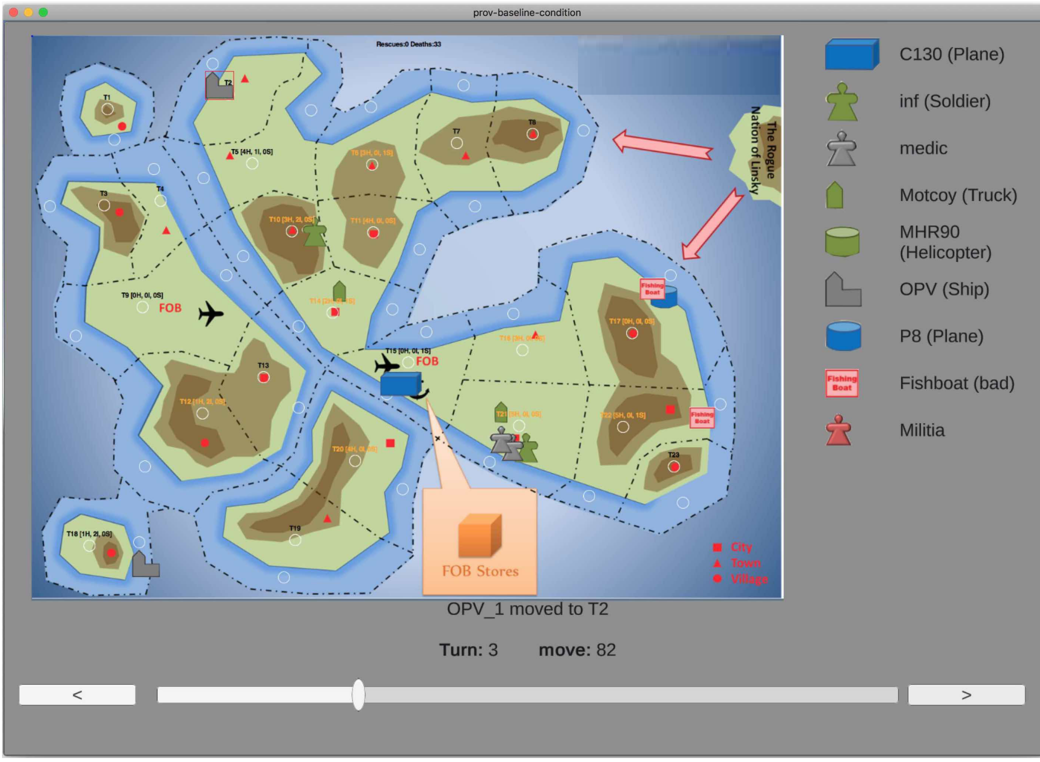

Fig. 3. The baseline visualisation tool developed for the study. Users could step forward and backwards through turns or scrub through using the slider. The activity taking place on a turn was listed below the map of the game and the changes would be reflected on the map (e.g. the OPV moved to territory 2).

experience with wargames (including tabletop and computer-based games). During this stage, participants also watched a 5 minute video describing the domain (i.e. the wargame) that they would be operating in. In Stage 2, participants were trained to use the system using mock data. For the baseline visualisation, the training covered stepping through turns in the game. For the storyline visualisation, training covered navigation and interactions. During training, participants were presented a set of practice comprehension questions. In Stage 3, participants had 3 minutes to explore the system and scenario. Participants were asked to step through the entire scenario and were encouraged to remember as many aspects of the scenario as possible for later recall. In Stage 4, participants performed the main task. Participants answered a set of 8 comprehension questions. Participants were asked to answer questions as accurately as possible. In Stage 5, participants filled a post-study questionnaire. This questionnaire was composed of a NASA Task Load Index, the SEQ and a preference survey, as well as general feedback. Participants were then asked to recall the questions and answers they gave. Stages 2–5 were then repeated for the other condition. A session with a participant lasted approximately 75 minutes.

COVID-19 Considerations: Participants were provided a pre-screening questionnaire before beginning the study to screen for risks and symptoms of COVID-19. No participants were rejected in pre-screening. Equipment and participants hands were sanitised before the experiment.

4.8 Participants

Four participants (three male, one female) were recruited from the student body of the university or via social networks. Participants were between 25 and 39 years of age ($M = 31.5$, $SD = 7.04$). All participants were right handed.

5 RESULTS

This section describes the results of the pilot study. It must be noted that—as it is a pilot—these results come from a relatively small sample size ($N = 4$). An alpha value of 0.05 was used for all statistical tests. Unless otherwise stated, a linear mixed effects model was used to examine differences in measures. The model was specified with fixed effects of visualisation and question, and full-factorial interaction effects. A random effect of the participant was specified on the intercept.

5.1 Speed

Analysing task completion time, the linear mixed effects model revealed a significant main effect of condition ($\chi^2(1) = 5.866, p = 0.015$) and question ($\chi^2(1) = 71.378, p < 0.001$). The model also revealed a condition × question effect ($\chi^2(1) = 30.566, p < 0.001$). Examining the interaction effect,we performed post-hoc pairwise comparisons of question and condition using the Tukey HSD test. The post-hoc test showed that Storyline ($M = 141.69s, SD = 47.67$) was significantly faster than Baseline ($M = 272.99s, SD = 65.44$) for Q1 ($p = 0.0279$), supporting hypothesis **H1**.

5.2 Accuracy

To analyse accuracy, we examined the trial attempt error. Again, the linear mixed effects model revealed a significant main effect of condition ($\chi^2(1) = 11.068, p < 0.001$) and question ($\chi^2(7) = 34.308, p < 0.001$). No interaction effect was found. Participants took significantly less attempts to answer the questions using the Storyline visualisation ($M = 1.28, SD = 0.63$) than Baseline visualisation ($M = 1.88, SD = 1.18$), supporting hypothesis **H2**.

There were not enough results from the task failure measure to perform a meaningful analysis, however, we report the observations here for completeness. Across all participants, Baseline had 3 complete task failures, Storyline had 1. Participants were asked to be as accurate as possible, so this result is not surprising as participants would spend extra time ensuring they were correct on the second or third attempt.

5.3 Recall

Mean recall scores for Baseline were 5.25 ($SD = 2.5$) and 7 for Storyline ($SD = 0.70$). A two-tailed paired sample t-test showed no significant difference in recall ($t(3) = -1.12, p = 0.343$) failing to support hypothesis **H3**.

5.4 Subjective Measures

We analysed the two subjective measures:

SEQ: Single ease question (i.e. how difficult was the task) was analysed using a one-tail paired sample t-test. Storyline ($M = 4.66, SD = 1.25$) was significantly easier than Baseline ($M = 3.33, SD = 1.25$) for the tasks presented; $t(3) = -4, p = 0.028$.

NASA TLX: Each participant's total workload score was combined into a single measure and a paired sample t-test was conducted to compare the results. Storyline ($M = 203.33, SD = 59.58$) had a significantly lower subjective workload score than Baseline ($M = 305, SD = 62.5$); $t(3) = 10.95, p = 0.008$.

6 DISCUSSION

Our early results suggest that Storyline visualisation may be significantly faster than Baseline for question Q1, supporting hypothesis **H1**. Q1 was a complex question, asking "Which unit moved to the most unique territories in the scenario?". Answering this question required the participant to observe every unit and make a mental note of the unique territories visited. For lower complexity questions in the Baseline condition, participants would step through the scenario and read the activity text only, ignoring the map entirely. The more complex questions required participants to use the map and visually scan all the entities.

Hypothesis **H2**—that storyline visualisation would be more accurate for complex questions—was also partially supported. In our small sample size, Storyline visualisation was more accurate for *all* questions, not just complex questions; we found no interaction effect between condition and accuracy. We still expect that with a larger sample size we would observe some interaction effect.

We found no effect on recall, failing to support hypothesis **H3**. One participant noted that scanning through and repeatedly reading the textual descriptions of activities in the scenario potentially helped them learn aspects of the questions and answers for later recall by rote. This suggest that recall in Baseline was supported through rote-learning rather than building a strong mental model of the game.

On both the NASA TLX and SEQ, participants subjectively rated the Storyline visualisation easier for performing the comprehension tasks presented.

7 LIMITATIONS AND CHANGES

The study presented was a pilot study. While a complete study with a larger number of participants is necessary to draw stronger conclusions, some trends were still apparent (as presented in the discussion). Furthermore, the study focused on comprehension of the lifetime of units within the wargame, rather than comprehension of complex spatial relationships which should perform better on a 2D map. It is worth noting that we do not consider storyline visualisations as a replacement to a map visualisation, but rather as a complement to provide a different view and insight into the data.

Participants had some difficulty reading the territory names, and suggested having the territory names clearly marked at any zoom factor, as well as having the territory name "stick" to the side of the screen when scrolling. Participants also suggested that typing in the search string was onerous when the names of the units and activities are all pre-determined. One participant suggested having a menu to quickly select and filter particular entities, rather than a search box.

8 CONCLUSION

We presented the motivation and design of a storyline visualisation for wargame data. We consider storyline visualisations to be a complementary tool that experts, from analysts to commanders, may use to better understand the results of a wargame scenario.

We evaluated the storyline visualisation in a comparative pilot user study, finding that the storyline visualisation was faster and more accurate for the particular set of questions we asked with our relatively small sample size ($N = 4$). As we expected, more complex questions that had multiple unknown factors performed particularly well in storyline visualisation. We also found that participants rated the workload lower for the storyline visualisation compared to the baseline map visualisation.

While these are only pilot results, we are encouraged by what we have found. We plan to adapt the tool taking into account feedback from participants as well as our observations of the system's use. In the future, we will run the full study to support more definitive conclusions and subsequently examine the system for plan monitoring.

REFERENCES

[1] Fereshteh Amini, Sébastien Rufiange, Zahid Hossain, Quentin Ventura, Pourang Irani, and Michael J McGuffin. 2015. The impact of interactivity on comprehending 2D and 3D visualizations of movement data. *IEEE transactions on visualization and computer graphics* 21, 1 (2015), 122–135. https://doi.org/10.1109/TVCG.2014.2329308

[2] Dustin Arendt and Meg Pirrung. 2017. The "y" of it Matters, Even for Storyline Visualization. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 81–91.

[3] T Baumgartl, M Petzold, M Wunderlich, M Höhn, D Archambault, M Lieser, A Dalpke, S Scheithauer, M Marschollek, VM Eichel, et al. 2020. In Search of Patient Zero: Visual Analytics of Pathogen Transmission Pathways in Hospitals. *arXiv preprint arXiv:2008.09552* (2020).

[4] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2301–2309.

[5] Peter Gatalsky, Natalia Andrienko, and Gennady Andrienko. 2004. Interactive analysis of event data using space-time cube. In *International Conference on Information Visualisation*. IEEE, 145–152. https://doi.org/10.1109/IV.2004.1320137

[6] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[7] D. Kamenetsky M. Wong and P. Williams. 2019. Towards an intelligent agent for a Humanitarian Aid Disaster Relief wargame. (2019).

[8] R Munroe. 2009. Xkcd# 657: Movie narrative charts. *URL: http://xkcd. com/657* 2 (2009).

[9] Kay L O'Halloran, Sabine Tan, Duc-Son Pham, John Bateman, and Andrew Vande Moere. 2016. A digital mixed methods research design: integrating multimodal analysis with data mining and information visualization for big data analytics. *Journal of Mixed Methods Research* (2016), 1558689816651015. https://doi.org/10.1177%2F1558689816651015

[10] Yuzuru Tanahashi and Kwan-Liu Ma. 2012. Design considerations for optimizing storyline visualizations. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2679–2688.

[11] Tan Tang, Renzhong Li, Xinke Wu, Shuhan Liu, Johannes Knittel, Steffen Koch, Thomas Ertl, Lingyun Yu, Peiran Ren, and Yingcai Wu. 2020. PlotThread: Creating Expressive Storyline Visualizations using Reinforcement Learning. *IEEE Transactions on Visualization and Computer Graphics* (2020).

[12] Tan Tang, Sadia Rubab, Jiewen Lai, Weiwei Cui, Lingyun Yu, and Yingcai Wu. 2018. iStoryline: Effective convergence to hand-drawn storylines. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 769–778.

[13] J. A. Walsh, J. Zucco, R. T. Smith, and B. H. Thomas. 2016. Temporal-Geospatial Cooperative Visual Analysis. In *2016 Big Data Visual Analytics (BDVA)*. 1–8. https://doi.org/10.1109/BDVA.2016.7787050

[14] Rolf A Zwaan and Gabriel A Radvansky. 1998. Situation models in language comprehension and memory. *Psychological bulletin* 123, 2 (1998), 162.

Received 1 September 2024