Evaluating Performance: Is Population-Centric Information Communicated Better Visually or Textually?

Michael N. Geuss

Army Research Lab, Human Research and Engineering Directorate

Abstract. Mission success is improved when decision makers receive population-centric information, like local religious, cultural, and economic centers of power in an efficient and effective manner. However, communicating population-centric information can be difficult because of large numbers of variables and complex relationships among them. This paper reports experimental results evaluating whether objective performance on a resource allocation task is influenced by presenting population-centric information visually or textually. In addition to how the information was presented, the quantity of information (1, 3, or 5 variables) was also manipulated. Participants viewed, or read, information that constrained how they should allocate their resources. Accuracy (appropriately assigning resources), time to make a decision, and subjective ratings of cognitive workload were recorded. Results demonstrated that participants made quicker decisions and subjectively rated the task as easier when population-centric information was presented visually. However, this effect disappeared when larger quantities of information were communicated suggesting that the utility of visualizing population-centric information may be limited.

Keywords: Population-Centric Factors, Mission Command, Visualization.

1 Background

Soldiers operate in complex and quick-paced environments that require equally quick decision-making and decisive action. Complexity in the battlefield is not only created by kinetic factors (e.g., enemy position and capability) but also by characteristics associated with the local population (e.g., religious, economic, & political centers of power). Mission success is improved when decision makers receive and incorporate population-centric information quickly into decision-making. When the population-centric information, or the identification and understanding of factors that influence the local populaces' will and intentions, is *appropriately* incorporated into mission planning, kinetic operations have been shown to be reduced by 60% [1]. However, appropriately integrating population-centric information into decision-making requires effective communication.

Currently within the Army, population centric data is communicated to decision-makers either as an overlay on a geospatial map (e.g., heat-map of water availability) and/or as a textual description (e.g., "neighborhood X has limited access to water"). Importantly, population-centric factors are often – if not always – distributed over space. It is unknown whether decision-making in military-related tasks is altered by the mode with which this information is communicated; that is, textually or visually.

Recent papers have echoed the need for empirical evaluations of human performance to validate the use of various visualization [2]. Visualizations are defined as external visual-spatial representations that are systematically related to the information that they represent [3, 4, 5]. Visualizing complex relationships has been argued as a method to reduce cognitive load (subjective mental and physical demands) by externalizing cognition and capitalizing on the human perceptual system's ability to recognize patterns [3, 6, 7]. However, visualizing multiple variables simultaneously may render any visualization cluttered and incomprehensible. A visualization design rule-of-thumb suggests that no more than three variables be simultaneously displayed because multiple representations can interfere with each other [6]. Therefore, when communicating complex relationships, it may be more advantageous to verbalize the 'bottom-line.'

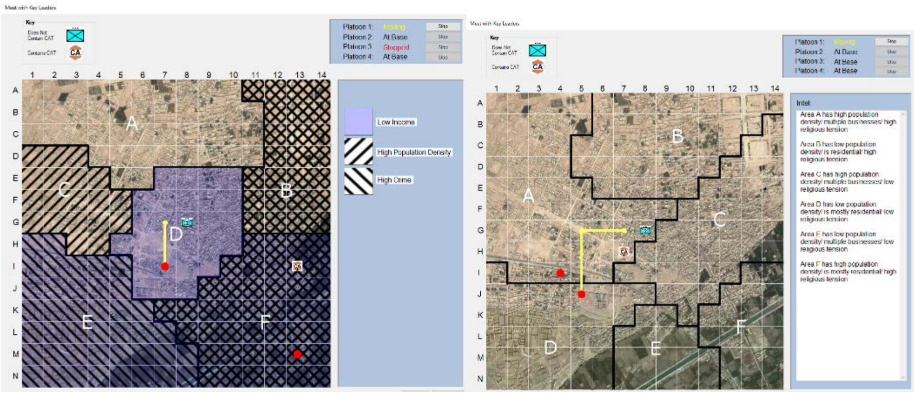

Figure 1. Visual (left) and Text (right) conditions are presented.

The goal of the current study was to determine whether there is a performance benefit to communicating population-centric information visually as overlays on a geospatial map and, if so, quantify the performance benefit as a function of amount of information being communicated. To test this question, participants allocated resources, either infantry units or Civil Affairs Teams (CATs), to meet with local leaders based on information presented to them. They were instructed that the CATs were only to be used in "sensitive neighborhoods." The sensitivity of a neighborhood was defined by being the combination of a number of population-centric variables (See Figure 1). We hypothesized that communicating information visually would provide both a speed and accuracy advantage over textual descriptions. However, we also hypothesized that the utility visualizations would disappear when presenting multiple variables simultaneously.

2 Method

2.1 Participants.

Eighteen (6 Female, 12 Male) volunteers ranging in age from 26 to 55 (m = 35.4, sd = 9.4) recruited from the U.S. Army Research Laboratory (ARL) workforce participated in the experiment. All participants were naïve to the purpose of the experiment and gave written informed consent. The experimental procedure was approved by the Human Research Protection Office, ARL.

2.2 Stimuli.

The experiment was conducted on a standard 19” desktop monitor with 1920x1080 resolution and a refresh rate of 60Hz. Participants viewed both the Text and Visualization conditions on the monitor. In the text condition, information describing neighborhoods (marked A-F on the maps) was provided textually in a sidebar. In the visual condition, the information describing the neighborhoods was provided as a visual overlay (hash marks, color, etc.) directly on the map with the legend provided on the sidebar.

Key-leaders would appear on the map as a red dot. Key-leaders are people from the local community that have influence over their community. To move the Infantry or CAT, participants clicked on the respective icon with a computer mouse and then selected the grid location where the key-leader appeared. They would then be asked to confirm their decision. Once confirmed, the unit would move to the target grid and be unusable until it arrived or was instructed to stop. Participants had 2 Infantry and 1 CAT at their disposal to meet with key-leaders.

After completing the key-leader engagement experiment, participants completed a number of surveys including: The NASA Task Load Index [8], the n-back (n = 3) task [9], and the Santa-Barbara Sense-of-Direction Scale (SBSOD) [10]. The NASA TLX requires participants to subjectively rate the workload required to complete the task on a number of subscales (e.g., frustration, physical demands, mental demands, etc.). The n-back task assesses participants working memory capacity by requiring participants to respond whether the currently viewed letter is the same as the one 3 prior in a series of continuously presented letters. In the current study, we implemented a 3-back test with a total stimulus set of 15 letters, 20 trials, and a stimulus presentation time of 2 seconds. The SBSOD requires participants to subjectively rate their sense of direction by rating how much they agree with statements about spatial and navigational abilities, preferences, and experiences.

2.3 Design.

A repeated-measures design was implemented with every participant completing all 3 levels of Quantity of Information (1, 3, and 5) for both Conditions (Text, Visual). In total, participants completed 6 blocks of trials. Within each block, 18 key-leaders appeared in semi-random locations on the map. The only limitations placed on where the key-leaders would appear was that 1/3rd of the trials, or 6 in total, must appear within the "sensitive neighborhood." The order of condition was counterbalanced across participants with half of the participants completing all Visual blocks first and then all Text blocks second and the other half in reverse order. Within each condition, the amount of information was always displayed in increasing order. This was done to reduce variability between conditions.

The information presented in the Visual condition for half of the participants was identical to that presented in the Text condition for the other half of participants. To prevent the same person from seeing the same information, we created two different maps with identical information and key-leader locations. Half of the participants would see map 1 for the Visual conditions and map 2 for the Text conditions and the other half of the participants would see map 1 for the Text conditions and map 2 for the Visual conditions. This created identical Text and Visual conditions across participants to reduce variability that could be due to key-leaders locations or the description of the information (e.g., residential neighborhood) that may have introduced bias if not controlled.

2.4 Procedure.

After signing the informed consent document, participants were instructed that their task was to allocate units to meet with key-leaders. They were instructed that CATs were highly trained to interact with leaders in sensitive areas and that only the CAT should be used when a key-leader appeared in the sensitive zone. Infantry teams should interact with leaders outside the sensitive zone. These were the only restrictions placed on participants. Each block of trials (Quantity of Information by Condition) had a unique sensitive zone that was provided to the participants in written format so that they could reference it during the experiment. Once participants indicated that they understood the task, they completed 5 practice trials where they moved a unit to interact with a key-leader. Participants then completed each of the 6 blocks of trials with self-regulated breaks between blocks to allow participants to familiarize themselves with the information that defined the sensitive area on the next map. The NASA TLX was administered after participants completed the Text and the Visual conditions. The N-back and SBSOD were completed after all experimental conditions were completed. After completing the surveys, participants were debriefed. The entire procedure lasted about 45 minutes.

3 Results

The goal of the analysis was to determine whether Condition (Text or Visual) and the Quantity of Information being communicated influenced performance on the resource allocation task. One, three, or five pieces of information were communicated either textually or visually. Performance was defined in terms of both assigning units as instructed (Accuracy) and the quickness with which a decision was made (Speed). N-back and SBSOD scores did not predict performance or interact with either independent variable in predicting performance.

3.1 Accuracy.

Accurately performing the task was defined as assigning the CAT to meet with leaders who appeared in the sensitive neighborhood. Since 6 out of 18 leaders appeared in a sensitive zone, accuracy scores were defined as the number of leaders out of six where a CAT was used. Using a regular infantry team or failing to assign a CAT to interact with a leader in a sensitive zone was counted as incorrect. If a participant utilized their CAT to interact with all 6 leaders in the sensitive neighborhood, their score was 100%.

Regardless of the amount of information being communicated or mode of communicating, participants' performance was very close to perfect (Ceiling effect). Participants sent the CAT to meet with leaders in the sensitive neighborhood on 95.68% of the trials. In Table 1, the average percent correct and standard deviation in percent correct is presented for each condition. Because there was a ceiling effect and there was no variability in performance on some conditions, no inferential statistical tests were conducted. However, the lowest performance, although still good, was seen when one and five pieces of information were being communicated textually. There were zero instances of false alarms (i.e., using a CAT in a non-sensitive area).

Table 1. Descriptive Statistics for Accuracy Measure

	One	Three	Five	Total:
Text	92.58 (10.2)	100 (0)	89.82 (16.3)	94.14 (7.28)
Visual	100 (0)	95.37 (9.5)	96.28 (9.1)	97.22 (4.7)
Total:	96.28 (5.1)	97.68 (4.8)	93.05 (2.5)	95.68 (5.2)

3.2 Speed.

Speed was defined as the amount of time between the appearance of a key-leader and a unit being assigned to move toward that key-leader. Speed was averaged across the 18 trials per scenario. A 3 Quantity of Information (1, 3, or 5) x 2 Condition (Textual, Visual) x 2 Order repeated measures analysis of variance (rmANOVA) was conducted on Speed scores. Order was the only between-participants factor. All results reported are Greenhouse-Geisser corrected because the data failed to meet the assumption of equal variance.

The analysis revealed a main effect of Quantity of Information such that participants made quicker decisions when less information was communicated, F (1.32, 21.16) = 6.54, p = 0.01, η^2 = 0.29. The main effect of Condition failed to reach significance, F (1, 16) = 2.64, p = 0.12, η^2 = 0.14. However, there was a significant interaction between Condition and Quantity of Information, F (1.6, 25.59) = 3.42, p = 0.04, η^2 = 0.18 (See Figure 2).

Planned comparisons tested whether there was an effect of Condition at each level of the Quantity of Information variable. There was only a significant effect of Condition when one piece of information was being communicated, t (17) = 2.73, p = 0.01 (see Figure 3). Participants were quicker when viewing information Visually (m = 6.03 s, se = 0.21) than Textually (m = 7.85 s, se = 0.61).

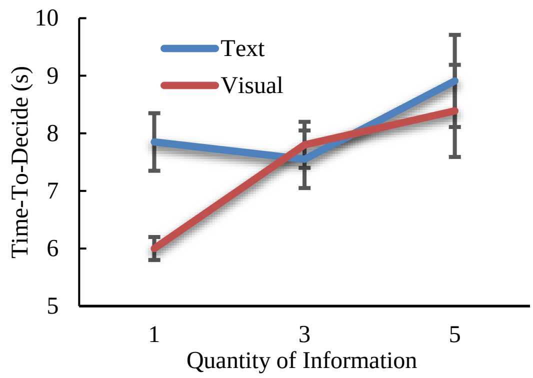

Figure 2. How quickly participants allocated resources is plotted against number of variables communicated for each condition. Participants were faster with visual presentation, but only when one variable was being communicated.

3.3 Cognitive Workload.

Cognitive workload was calculated as the sum of the scores from each sub-scale for both the Text condition and Visual condition. Higher scores indicate more cognitive workload. Participants reported requiring more cognitive workload when completing the Text condition (m = 49.08, sd = 18.22) than when completing the Visual condition (m = 38.53, sd = 12.67), t (17) = 3.01, p = 0.01. That is, participants subjectively rated the Text condition as being more cognitively demanding than the Visual condition.

4 Conclusions

Overall, the results of the current study suggest that visual presentation of information results in quicker decisions and subjectively easier processing of information. However, the objectively measured benefit of communicating information visually disappeared when more than 1 piece of information was presented. These results provide evidence for recommendations to limit visualizations to less than 3 pieces of information [6]. While the current study presents a methodology for evaluating visualization methods, future studies would benefit from making the task more difficult in order to find differences in accuracy, if such differences exist.

References

1. Benoit, T. S., Graffeo, C., Carter, M. A., & Swisher, C. R.: The Sixty Percent Mission: An Introduction to High-Risk Ethnography Process and Protocol in Support of the US Army's Civil Affairs Humanitarian Mission. In: Modeling Sociocultural Influences on Decision Making: Understanding Conflict, Enabling Stability, pp. 111-134. CRC Press (2016).
2. Isenberg, T., Isenberg, P., Chen, J., Sedlmair, M. and Möller, T.: A systematic review on the practice of evaluating visualization. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2818-2827 (2013).
3. Bertin, J.: Semiology of graphics: diagrams, networks, maps. University of Wisconsin Press (1983).
4. Hegarty, M.: The cognitive science of visual-spatial displays: Implications for design. Topics in cognitive science 3(3), 446-474 (2011).
5. Stenning, K. & Oberlander.J.: A cognitive theory of graphical and linguistic reasoning: Logic and implementation. Cognitive science, 19(1), 97-140 (1995).
6. Munzner, T.: Visualization analysis and design. CRC press (2014).
7. Ware, C. Information visualization: perception for design. Elsevier (2012).
8. Hart, S. G. & Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P. A. and Meshkati, N. (eds.) *Human Mental Workload*. Amsterdam, North Holland Press (1988).
9. Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. Journal of Experimental Psychology, 55, 352-358.
10. Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I.: Development of a self-report measure of environmental spatial ability. Intelligence, 30, 425-448 (2002).

Acknowledgments.

The author would like to acknowledge Joe Campanelli for creating experimental stimuli and to thank Jon Bakdash and Jock Grynovicki for their feedback.