ARL-TR-8451 ● AUG 2018

US Army Research Laboratory

Using the Value of Information (VoI) Metric to Improve Sensemaking

by Mark Mittrick, John Richardson, Derrik Asher, Alex James, and Timothy Hanratty

NOTICES

Disclaimers

ARL-TR-8451 ● AUG 2018

US Army Research Laboratory

Using the Value of Information (VoI) Metric to Improve Sensemaking

by Mark Mittrick, John Richardson, Derrik Asher, and Timothy Hanratty
Computational Information Sciences Directorate, ARL

Alex James
CUBRC
4455 Genesee St, Ste 106
Buffalo, NY 14225

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
August 2018	Technical Report	February–November 2017

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
Using the Value of Information (VoI) Metric to Improve Sensemaking	
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Mark Mittrick, John Richardson, Derrik E Asher, Alex James, and Timothy Hanratty	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
US Army Research Laboratory Computational Information Sciences Directorate (ATTN: RDRL-CII-T) Aberdeen Proving Ground, MD 21005	ARL-TR-8451
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Presented at 22nd ICCRTS, 2017 November 6–8, Los Angeles, CA. Also available as arXiv: 1807.09837.

14. ABSTRACT

Sensemaking is the cognitive process of creating schemata from extracted information, making decisions from those schemata, and inferring conclusions. Human analysts are essential to exploring and quantifying this process, but they are limited by their inability to process the volume, variety, velocity, and veracity of data. Visualization tools can help this human–computer interaction. For example, analytical tools that use graphical link–node visualization can help sift through vast amounts of information. However, assisting the analyst in making connections with visual tools can be challenging if the information is not presented in an intuitive manner. Experimentally, it has been shown that analysts increase the number of hypotheses formed if they use visual analytic capabilities. Exploring multiple perspectives could increase the diversity of those hypotheses, potentially minimizing cognitive biases. This report presents preliminary research results that indicate an improvement in sensemaking over the traditional link–node visualization tools by incorporating an annotation enhancement that differentiates links connecting nodes. This enhancement provides a visual cue, which represents the perceived value of reported information. Improved sensemaking occurs because the limitations of mentally consolidating, weighing, and highlighting data are removed. This study investigates line thickness as a valid representation of Value of Information.

15. SUBJECT TERMS

value of information, visual analytics, link-node diagram, sensemaking, human computer interaction

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Mark Mittrick
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	16	**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified			(410) 278-4148

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

In 1993, Russell et al. introduced the concept of sensemaking to the human–computer interaction community.[1] They identified it as a common activity in analysis, involving the process of searching for a representation and encoding data in that representation to answer task questions. Following that, researchers sought to begin incorporating sensemaking into their visualization applications.

Around the same time that sensemaking was introduced, Larkin and Simon suggested that analysts could spot anomalies and other patterns if the burden of mentally consolidating information was minimized.[2] This statement identifies the underlying assumption or premise of why many analytical and visualization tools are useful—they enable users to gain insights that are otherwise obscured. Visualization tools affect the amount of cognition needed to solve problems by reducing the difficulty level of finding and comparing data.

The link-node diagram is one prominent visualization tool that combines nodes with connecting links to create a network of associated nodes.[3] While investigating improvements to the human–interactive aspect of link-node analysis, Ware and Bobrow researched techniques for highlighting a small number of nodes to determine whether a large network could be displayed while maintaining the effective visualization power of a small link-node diagram.[4] Their research combined visual highlighting and motion cues to emphasize a small number of nodes and compared the effectiveness of the visual cues to baseline results. Their results showed that analysts could answer questions with undirected graphs having less than 100 nodes, compared to a baseline level of error; however, performance approached chance levels as the undirected graph grew larger. When highlighting was introduced, error levels dropped substantially, in essence demonstrating that pre-attentive cues are effective within the context of large and complex link-node visualizations. Another study found that using a weighting scheme that displayed a link's length in proportion to its weight improved comprehensibility of link-node graphs representing webpage similarity data.[5] Together, these studies provide sufficient evidence for investigating line thickness as an effective means of improving sensemaking in link-node diagrams.

1.1 Value of Information

The Value of Information (VoI) is a metric that computes a likelihood of applicability based on metadata of recorded information. Specifically, the VoI combines source reliability, likelihood that data is true, and timeliness with respect to mission.[6,7] This study tests the perceived value of line thickness within a link-node visualization framework to quantitatively assess the performance gains when utilizing link-node density as an independent variable. If line thickness is determined to help reduce the mental burden on analysts, then it is suitable to be used in VoI paradigms.

1.2 Crowdsourcing

Cialdini and Trost describe crowdsourcing as a process of outsourcing difficult-to-answer questions to a crowd of individuals.[8] The power of crowdsourcing comes from the "wisdom of the crowd" concept, which indicates that a large number of individuals estimating some phenomena will produce an averaged estimate that is as good as, or often better than, that of an expert.[9,10] An explanation for this phenomenon is that noise inherently exists in estimates, and an average over a large amount of these noisy estimates results in a reduction in the overall noise, abiding by the law of large numbers in probability theory.[11,12] Given its power, the crowdsourcing method is an ideal choice for examining whether or not VoI helps to improve sensemaking.

The work presented in this report shows how crowdsourcing informs the VoI paradigm with a non-analyst population, utilizing response time and performance accuracy determined from degree centrality.

2. Methods

The Amazon Mechanical Turk (MTurk) crowdsourcing platform was used to collect data from 303 participants. A simple computerized task required that subjects review a link-node diagram, then select the node they "know most about". Each subject was randomly assigned to one of six conditions, which were derived from three levels (Easy, Medium, and Hard) and two groups (VoI and Control).

2.1 Human Subjects

The research performed in this study falls under the US Army Research Laboratory (ARL) Internal Review Board Exempt Research Determination for Protocol (ARL 17-093), which indicates that it is exempt from regulation 32 CFR 219. The research is exempt because it falls into the exemption criteria defined by the Common Rule, which states that human subjects cannot be identified by the collected data and their responses will not place them at risk of criminal or civil liability or otherwise damage their financial standing, employment, or reputation.

Subjects volunteering for participation in this study were notified that they needed to be familiar with link-node diagrams and that no personally identifiable information would be collected. They would earn $0.25 upon successful completion of the experiment, with a possible additional bonus of $0.25 if they were able to correctly identify the node they know the most about.

Three demographic questions were asked of each qualifying subject, focusing on the subject's occupation, age, and education level. Exclusionary criteria consisted of the subject's knowledge of link-node diagrams—any subject indicating that they were unfamiliar with link-node diagrams was thanked for their interest and program exited (without providing any data).

User bias was minimized by setting the MTurk eligibility criteria to allow subjects to participate only one time. MTurk informed subjects attempting to participate a second time that they were no longer eligible.

2.2 Experimental Setup

Qualified subjects were presented with a sample graph on which to practice and a set of instructions directing them to imagine themselves as analysts studying a link-node diagram. The instructions went on to specify that each link incident upon a node represents the metric to be maximized, with a thicker link representing a more relevant node (Fig. 1). The subject was required to do the following:

1) Assess the diagram to discern the node with the greatest degree centrality, which was modulated by line thickness in the VoI cases.

2) Highlight the node via mouse click to indicate that a selection has been made. (The selected node was also displayed in a list next to the diagram.)

3) If unhappy with the selection, press the Reset button to restart the selection process.

4) Once happy with the selection, press the Submit button to record the answer.

5) The subject was able to drag the nodes and manipulate the graph to optimally assess the degree centrality.

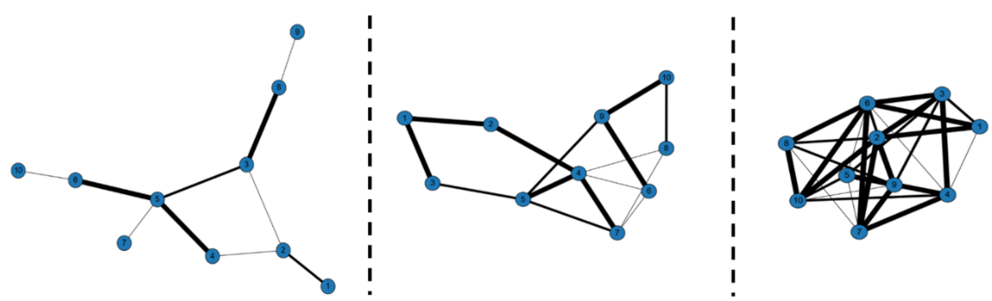

Fig. 1 Link-node diagrams from experiment showing easy–VoI (left), medium–VoI (center), and hard–VoI (right)

2.3 Data Analysis Methods

The Jarque–Bera test indicates whether data comes from normal distribution with an unspecified mean and standard deviation. The Jarque–Bera test was utilized in this study to confirm that the data were not normally distributed, and therefore required nonparametric tests for appropriate analyses.

The Wilcoxon signed-rank test is a nonparametric test. The null hypothesis we used for this test indicates that the distributions of the compared samples are equal. Small P-values reject the null and imply that the distributions and medians are not equal.

The Kruskal–Wallis one-way analysis of variance test is equivalent to the Wilcoxon test and was used to confirm the statistical results.

The P-values generated from the Wilcoxon and Kruskal–Wallis tests represent the statistical significance associated with all compared data samples from the respective conditions. P-values were considered significant if below 0.05 and corrections for multiple comparisons were not necessary.

2.4 Conditions

To keep the experiment consistent across all conditions, the number of nodes was kept constant while varying the number of links. In this study, there are six conditions: Easy-VoI, Easy-Control, Medium-VoI, Medium-Control, Hard-VoI, and Hard-Control. The graph density formula calculates the density of the graph given a set of nodes and edges. The graph density formula is

$$D = \frac{2 \times E}{N(N-1)}, \tag{1}$$

where D is the density of the graph, E is the number of edges in the graph, and N is the number of nodes in the graph (N = 10 for all graphs). In this study, the density was 22% for Easy (10 edges), 33% for Medium (15 edges), and 66% for Hard (30 edges). The initial threshold for each density level was determined by the perceived level of difficulty and confirmed through preliminary data collected. The Easy level threshold (Easy-VoI and Easy-Control) was selected to enable the subjects to count the number of edges instead of estimating in order to maximize their potential to select the correct node. Preliminary data collected from the Medium and Hard levels showed that subjects were estimating rather than counting, since their performance decreased as difficulty level increased.

3. Results

Two metrics were used to evaluate whether line thickness can be considered a valid visual representation of VoI—Situational Awareness (SA) and Response Time (RT). SA is the sum of the node's links (i.e., degree centrality) and RT is the duration between the time a subject first saw the graph and the time they submitted their answer. Results from this study show significant statistical differences when comparing certain VoI and Control (non-VoI) graphs using (SA) and (RT).

Each graph has a deterministic value (SA) that provides a quantifiable metric for performance evaluation. In the Control conditions, all links have a value of 1, whereas in the VoI conditions, each link is weighted depending on the thickness of the line (1 for thinnest, 2 for medium, and 3 for thickest). For example, a node with three links in the Control conditions has an SA of 3 (one for each link), whereas that same node in the VoI conditions with links of medium thickness has an SA of 6 (3 links × 2 medium thickness).

Since the data did not follow a Normal distribution (confirmed with Jarque–Bera tests), nonparametric tests were utilized. The Wilcoxon and Kruskal–Wallis tests were performed pairwise to determine statistical differences between conditions. The study was balanced with an approximately equal number of subjects per condition (Table 1).

Table 1 Distribution of subjects across conditions

Category	Easy	Medium	Hard
VoI	49	44	55
Control	53	50	52
Total	102	94	107

The Wilcoxon results for SA (Table 2) show that Easy-VoI versus Easy-Control and Hard-VoI versus Hard-Control were statistically significant at the (alpha = 0.05) level, suggest that the compared samples come from different underlying distributions. Thus, the Easy and Hard levels show significant improvement in performance of choosing the node with the highest degree centrality. In contrast, Medium-VoI versus Medium-Control was not found to be statistically significant. These results suggest that line thickness for the Easy and Hard levels of difficulty are a valid representation of VoI (Fig. 2).

Table 2 **Wilcoxon signed-rank test results for Situational Awareness**

Wilcoxon	P-value
Easy-VoI vs. Easy-Control	0.0404[a]
Medium-VoI vs. Medium-Control	0.6361
Hard-VoI vs. Hard-Control	0.0056[a]

[a] Statistically significant at the alpha = 0.05 level

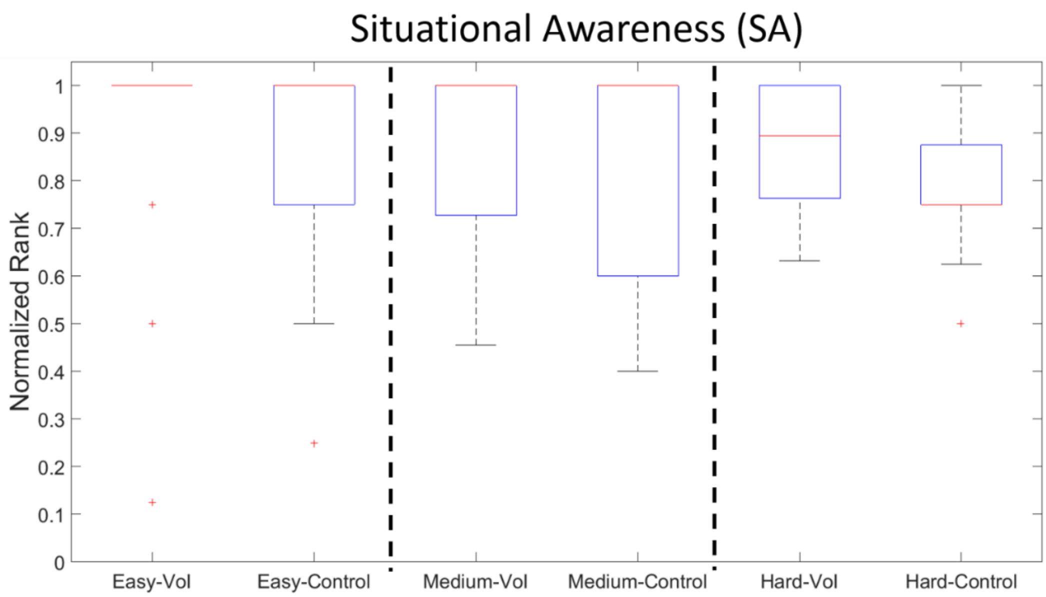

Fig. 2 Situational Awareness across conditions. The boxplot shows the distribution of data, with Normalized Rank ranging from 0–1 (y-axis) and the six conditions (x-axis). The red bars show the medians, the blue boxes show the upper and lower quartiles, the whiskers extend 2.7 standard deviations from the median, and the red plus signs represent outliers. The vertical dashed lines separate the conditions according to difficulty level (left segment is easy, middle segment is medium, and right segment is hard).

The distribution of data was compared pairwise across the levels of difficulty. In the Easy level, the Easy-VoI condition's median, upper and lower quartiles, and whiskers are all the same, since the majority of the subjects performed perfectly. A total of 49 subjects were placed in the Easy-VoI condition (Table 1), 40 (82%) of which earned a perfect score. The nine subjects who chose incorrectly are represented by the three visibly distinct outliers (denoted by red plus signs in Fig. 2). The nine outlier data points overlap because normalized rank performance collapsed into three distinct values. Medium level results suggest that there is no visible statistical significance in the middle pairwise comparison shown in Fig. 2. (Additional data are necessary to fully examine this phenomena.) Finally, in the Hard level, we have observed statistical significance with the VoI condition (VoI-Hard) subjects outperforming the control condition subjects (Hard-Control).

The Wilcoxon results for RT (Table 3) show that Easy-VoI versus Easy-Control was found to be statistically significant at the (alpha = 0.05) level. This suggests that 1) the compared samples come from different underlying distributions, and 2) subjects took considerably less time in the VoI case (Fig. 3). In contrast, Medium-VoI versus Medium-Control and Hard-VoI versus Hard-Control were not found to be statistically significant. Additional data may result in significant differences at the Medium and Hard levels.

Table 3 Wilcoxon signed-rank test results for Response Time

Wilcoxon	P-value
Easy-VoI vs. Easy-Control	0.0432[a]
Medium-VoI vs. Medium-Control	0.0873
Hard-VoI vs. Hard-Control	0.7317

[a] Statistically significant at the alpha = 0.05 level.

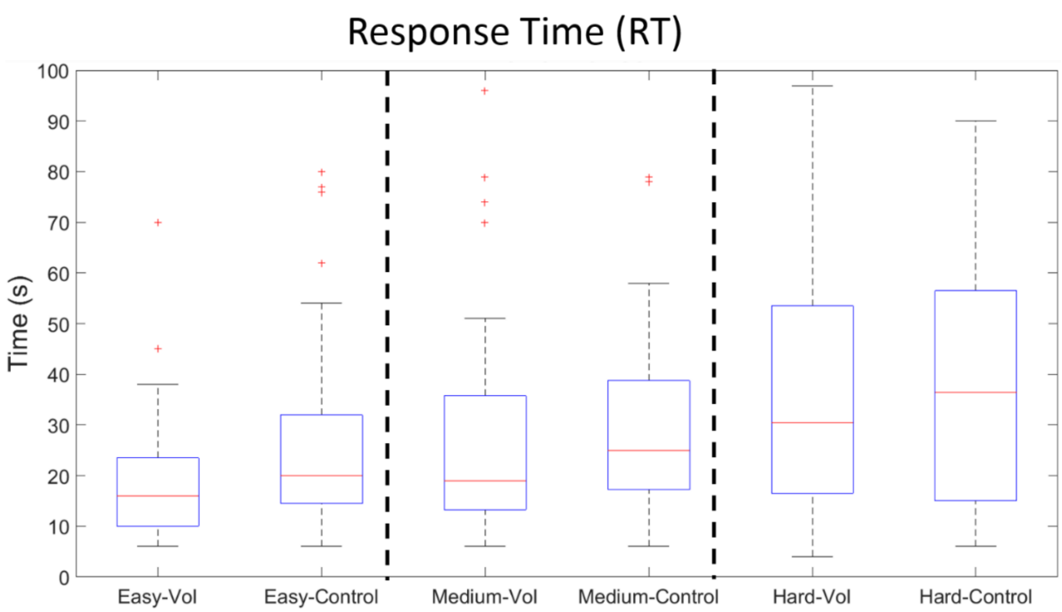

Fig. 3 Response Time across conditions. The boxplot shows the distribution of data, with Time (y-axis) ranging from 0–100 (seconds) across the y-axis and the six conditions (x-axis). The red bars show the medians, the blue boxes show the upper and lower quartiles, the whiskers extend 2.7 standard deviations from the median, and the red plus signs represent outliers. The vertical dashed lines separate the conditions according to difficulty level (left segment is easy, middle segment is medium, and right segment is hard).

Again, the distribution of data was compared pairwise across the levels of difficulty. In the Easy level, the results suggest that there is statistical significance with the Easy-VoI subjects outperforming the Easy-Control condition subjects. Although no statistical significance was found in the Medium and Hard levels, the medians are trending in a direction consistent with the SA analysis of SA (Figs. 2 and 3).

4. Discussion and Conclusion

This study aimed to determine whether line thickness can be used as a valid representation of VoI. The SA analysis showed that line thickness for the Easy and Hard difficulty levels, determined through degree centrality, provided a significant improvement in subject performance over the Control conditions (see Fig. 2). In addition, the RT analysis showed that subjects performed significantly faster in the Easy level due to line thickness (see Fig. 3). Furthermore, the median RT was found to be greater in Control conditions per difficulty level, although not significant for Medium and Hard levels. Together, these results suggest that line thickness might be a viable option to represent VoI.

The results from this experiment are a first step toward lessening the mental burden to improve sensemaking. The ability of VoI to provide a visual cue is paramount to quickly understanding the information presented, as well as making important decisions in a quick and timely matter. Our work shows that using line thickness as a visual cue to the value of a node significantly improves selection performance in the experiment. Building upon previous work by Ware and Bobrow, this result shows how cues generated from the perceived value of the data used to create the link-node diagram can aid the analyst under certain conditions.[4]

Additional data and experiments are needed to explore how further graph density manipulations might influence the perceived value of line thickness and help clarify the nonsignificant findings (see Tables 2 and 3). We propose a set of supplementary experiments with a greater number of participants, different node quantities, and additional graph density levels to determine how best to utilize line thickness and its perceived value. We currently use random number generation to determine the line thickness and placement of the graphs. We are unable to assess how this impacts our study at this point. In the future, we hope to find a way to normalize this assignment to further control our observations.

5. References

1. Russell DM, Stefik MJ, Pirolli P, Card SK. The cost structure of sensemaking. Proceedings of the SIGCHI conference on human factors in computing systems. 1993 Apr 24–29; Amsterdam, The Netherlands.
2. Larkin J, Simon H. Why a diagram is (sometimes) worth ten thousand words. Cogn Sci. 1987;11(1):65–100.
3. Holten D, Van Wijk JJ. A user study on visualizing directed edges in graphs. In: CHI 2009. Proceedings of the 27th International Conference on Human Factors in Computing Systems. 2009 Apr 4–9; Boston. c2009. p. 2299–2308.
4. Ware C, Bobrow R. Supporting visual queries on medium-sized node–link diagrams. Inf Vis. 2005;4(1):49–58.
5. Rodgers P, Mutton P. Visualizing weighted edges in graphs. Proceedings of the Seventh International Conference on Information Visualization; 2003 July 18; London, UK. Los Alamitos (CA): IEEE Computer Society; c2003. p. 258–263.
6. Hanratty T, Hammell RJ, Heilman E. A fuzzy-based approach to the value of information in complex military environments. In: Benferhat S, Grant J, editors. Lecture notes in computer science. Vol 6929. Berlin, Heidelberg (Germany): Springer; 2011. p. 539–546.
7. Hanratty TP, Newcomb EA, Hammell RJ 2nd, Richardson JT, Mittrick MR. A fuzzy-based approach to support decision making in complex military environments. Int J Intell Inf Technol. 2016;12(1):1–30.
8. Cialdini RB, Trost MR. Social influence: social norms, conformity and compliance. In: Gilbert DT, Fiske ST, Lindzey G, editors. The handbook of social psychology. Vol. 2. New York (NY): McGraw-Hill; 1998. p. 151–192.
9. 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. 2010;466(7307):756–760.
10. Cialdini RB, Goldstein NJ. Social influence: compliance and conformity. Annu Rev Psychol. 2004;55(1):591–621.
11. Yi SKM, Steyvers M, Lee MD, Dry MJ. The wisdom of the crowd in combinatorial problems. Cogn Sci. 2012;36(3):452–470.
12. Lee MD, Zhang S, Shi J. The wisdom of the crowd playing The Price Is Right. Mem Cognit. 2011;39(5):914–923.

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIR ARL IMAL HRA RECORDS MGMT RDRL DCL TECH LIB
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
4 (PDF)	ARL RDRL CII T M MITRRICK J RICHARDSON D ASHER T HANRATTY