ARL-TN-1004 ● JAN 2020

Human-Autonomy Teaming Essential Research Program Project 2: Transparent Multimodal Crew Interface Designs
Technical Note 2: Transparency in Mobility Planning

by Brandon S Perelman, Julia L Wright, Gregory A Lieberman, and Shan Lakhmani

NOTICES

Disclaimers

ARL-TN-1004 ● JAN 2020

Human-Autonomy Teaming Essential Research Program Project 2: Transparent Multimodal Crew Interface Designs
Technical Note 2: Transparency in Mobility Planning

Brandon S Perelman, Julia L Wright, Gregory A Lieberman, and Shan Lakhmani
Human Research and Engineering Directorate, CCDC Army Research Laboratory

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE (*DD-MM-YYYY*)	2. REPORT TYPE	3. DATES COVERED (From - To)
January 2020	Technical Note	1 November 2018–30 September 2019

4. TITLE AND SUBTITLE

Human-Autonomy Teaming Essential Research Program Project 2: Transparent Multimodal Crew Interface Designs
Technical Note 2: Transparency in Mobility Planning

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Brandon S Perelman, Julia L Wright, Gregory A Lieberman, and Shan Lakhmani

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

CCDC Army Research Laboratory
ATTN: FCDD-RLH
Aberdeen Proving Ground, MD 21005

8. PERFORMING ORGANIZATION REPORT NUMBER

ARL-TN-1004

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

ORCID ID(s): Julia Wright, 0000-0003-3026-1538; Shan Lakhmani, 0000-0001-6052-439X

14. ABSTRACT

This technical note, the second in its series, describes the integration of interface concepts for improving mobility planning. This work is a subcomponent of the US Army Combat Capabilities Development Command Army Research Laboratory's (ARL's) second project in the Human-Autonomy Teaming Essential Research Program, transparent multimodal crew interface designs, in support of the US Army modernization priority – Next Generation Combat Vehicle (NGCV). The goal of this project was to improve crew members' understanding of autonomous systems' actions, intentions, goals, and general reasoning by at least 25%. To achieve this goal, a team of ARL scientists and contractor engineers implemented iterative improvements to the NGCV Warfighter Machine Interface aimed at addressing multiple stages of the mission, from planning to execution. Interface improvements included a multimodal (tactile and auditory) cueing system designed to improve crew members' local situation awareness while commanding simulated autonomous robotic combat vehicles, a display for evaluating tradeoffs in the Multi Domain Operations battle space, and improvements to existing route-planning functionality. These interface improvements were implemented and tested in ARL's Information for Mixed Squads Laboratory, which permits the simulation of up to platoon-level operations. This note describes the scientific background for, and ideation, implementation, and testing of two technologies: a Transparent Route Planner and Comparator Display.

15. SUBJECT TERMS

Next Generation Combat Vehicle, human–robot interaction, human-autonomy teaming, autonomous systems, human–computer interaction

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Brandon S Perelman
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)
Unclassified	Unclassified	Unclassified	UU	30	(410) 278-5968

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

Contents

List of Figures

1. Introduction

The Human-Autonomy Teaming (HAT) Essential Research Program (ERP) research plan describes several projects that not only will use the capabilities of future intelligent agents to improve team outcome but also will leverage the novel interactions between Soldiers and intelligent agents that serve as teammates rather than as tools. This publication series describes HAT Project 2, Multimodal Transparent Crew Interface Designs, aimed at unburdening and empowering the Soldier by improving crew members' understanding of unmanned/intelligent vehicle actions, intentions, goals, and general reasoning. This aim is accomplished by providing information to the Soldier in a clear, intuitive manner that enables understanding of the underlying unmanned vehicle system's decision processes that have traditionally been opaque, difficult to understand, and not conducive to facilitating trust in autonomy. This technical note, the second in the series, describes the integration of two specific deliverables, the Transparent Route Planner (TRP) and the Comparator Display, into the Next Generation Combat Vehicle (NGCV) Warfighter Machine Interface (WMI). These deliverables are intended to improve human-autonomy teaming during the planning phase of mobility operations, but also have potential future applications for presenting information at the point of need in novel ways during mission execution. For a detailed description of the HAT Project 2 program and the laboratory out of which this effort is based, please refer to the first technical note in this series (Perelman et al. 2020).

2. Implementation

In this iteration of HAT Project 2, two transparency concepts are presented: first, a TRP aimed at improving the crew's understanding of the unmanned vehicle's actions and intended future movements and second, a Comparator Display designed to allow the crew to quickly compare the costs and benefits of several different courses of action (COAs). Both technologies will contribute to improved understanding of the underlying information used by the autonomy to plan routes between mission objectives. These concepts are implemented directly into the WMI and will increase users' comprehension of the actions, intentions, goals, and general reasoning of an unmanned vehicles.

2.1 Designing Interface Elements to Facilitate Transparency

During the course of the mission planning process, commanders must compare several COAs and decide upon one that is appropriate given the mission, enemy, terrain, troop availability, time, and civilian considerations (METT-TC). This is no trivial task, as is seldom one single optimal solution since all potential COAs reflect tradeoffs in the multiple mission-relevant criteria. Often, these high-dimensional tradeoffs interact with individuals' information access and weighting, spatial mental models, or understanding of the problem space, leading them to produce different evaluations of a given situation, and different solutions to spatial planning problems. For example, a study of infantry company commanders by Myer and Lojka (2012) revealed that mission planning is typically conducted using heuristics, and that variability among commanders can produce vastly different assessments of a given situation. This behavioral variability among teammates extends to human-autonomy collaborative route-planning tasks as well. For example, Perelman et al. (2017) studied the behavior of humans and autonomy in a collaborative route-planning task and found that a heterogeneous group of humans and autonomy will generally converge to a small number of solutions (i.e., 2–6 solution *types*), each of which might be reasonable according to individual differences in attitudes or weighting toward tradeoffs of mission-relevant criteria but that still introduce variability into the mission plan. While some of the routes generated by humans in that study were congruent with routes suggested by the autonomy, many were not, and this holds implications for the predictability and trustworthiness of autonomous systems that use different route-planning mechanisms than human teammates. One potential method for mitigating the disparity in spatial decision-making among agents in human-human and human-autonomy teams is to visually present the proposed routes in detail (to reveal agent-proposed actions and intentions), as well as the differences between routes in terms of their scores on mission-relevant criteria so that the tradeoffs can be properly evaluated (to provide more information on agent goals and general reasoning processes). Further, it is essential that detailed information of this nature be presented in a clear, intuitive manner so as not to burden the Soldier and so knowledge of the underlying processes can be leveraged efficiently for maximum improvement to mission planning.

2.1.1 Developing the Transparent Route Planner

Prior US Army Combat Capabilities Development Command (CCDC) Army Research Laboratory (ARL) research investigated the breadth of variability in route-planning behaviors in a group of humans and autonomous systems (Perelman et al. 2017), and highlighted the importance of communicating this information among teammates in human-autonomy teams. In particular, autonomous system route planning is frequently accomplished using algorithms based upon optimization (e.g., Dijkstra's algorithm finding the shortest path [Dijkstra 1959]). Planetary rover navigation is a real-world example of this optimization approach, and algorithms have been designed to optimize the rover's exploration of novel regions (Lee and Ahn 2017) or the amount of time that it spends under sunlight (Otten et al. 2017), for example. By contrast, human spatial decision-making is heuristic in nature and well adapted to providing acceptable solutions to high-dimensional problems quickly by leveraging the hierarchical structure of biological systems. Due to these differences, members of human-human and human-autonomy teams frequently produce routes that are unpredictable to teammates.

Visualizing the path that an asset will take to its destination is critical for the control and understanding of that asset's behavior, allowing commanders to evaluate the asset's mobility plans and their potential impact on the mission in advance. This general requirement is particularly well understood in the GitHub gaming community, which has founded the Pathfinding Visualization Project specifically for creating and organizing visualizations that can be used to show an asset's intended future movements for the purpose of increasing the autonomy's transparency (https://github.com/qiao/PathFinding.js). Consistent with the HAT ERP's Decision Point 1 (DP1) goal constructs—demonstrating a 25% increase in a crew's comprehension of unmanned vehicle actions, intentions, goals, and general reasoning—the TRP is designed to improve crew members' understanding of the Robotic Combat Vehicle (RCV) future actions and intentions (Fig. 1) by generating a full route-planning solution for the autonomy across all terrain features and displaying this solution to the crew members.

Fig. 1 Route planning using the baseline WMI (left panel) and TRP-equipped WMI (right panel). Routes generated by a single-click supervisory control scheme are shown in yellow. Using the baseline WMI, which merely shows a connection between the origin and destination, it is unclear how the RCV's local decision-making navigation processes will reach the objective. For example, the vehicle may struggle to navigate entirely within narrow alleys inside city blocks or might even exit and circumvent the city in its entirety, and the baseline WMI does not provide any up-front information for predicting the vehicle's route. The TRP-equipped WMI, on the other hand, clearly displays the planned route that the vehicle will take, allowing for analysis of the impact of the chosen path, or selection of a new path, during mission planning.

2.1.2 Developing the Comparator Display

Selecting an optimal method to represent information visually can be a challenge. Ideally, the representation would assist the reader to frame the problem and potential outcomes, quantitatively compare various factors, and help the reader consider alternative conclusions (Tufte 1997), all while remaining clear, intuitive, and transparent. Matching the manner of presentation to the problem supports effective decision making (Vessey 1991). In the case of a route-planning task, multiple factors are compared both within and between routes to ascertain the path that best satisfies complex mission requirements. In the case of a comparison task, bar graphs are particularly useful, as they allow the reader to directly compare values across two or more factors quickly and efficiently (Gillan et al. 1998). In order to convey the transparency concepts in a way that best suits the information needed and the tasks they support, we use visualization and design research to match the presentation with the problems they solve.

Autonomous systems often use many factors and the interactions between them in their decision-making processes. As situation complexity increases, often the number of relevant factors does as well. However, as the number of factors increases, the agent's decision-making process becomes more complex and less transparent to the human teammate. Behymer and associates (2015) used a Pareto efficiency approach to data visualization to convey relevant decision-making factors to operators. By normalizing these factors on a 0–100 scale and mapping them across a prioritized series of columns, they were not only able to compare multiple factors but also able to compare multiple plans along those factors as well. This plot-based approach allowed participants to compare specific parameters across plans more quickly than a chart or matrix approach to data visualization.

This approach was later used in conjunction with a recommendation agent to help participants compare competing plans in a military defense task (Stowers et al. 2016). Not only did this display (Fig. 2) convey information about the plans themselves, but it also described how heavily the agent weighted each parameter, the projected success of each plan, and whether the estimation of that plan's success was based on certain or uncertain information. The addition of information that supported greater agent transparency was shown to benefit the human's decision-making process, improving overall team performance (Chen et al. 2018). This visualization approach, combining Pareto efficiency information and Situation Awareness-based Agent Transparency-based information, is well suited for the route comparisons and decisions that Soldiers would have to make using the NGCV WMI.

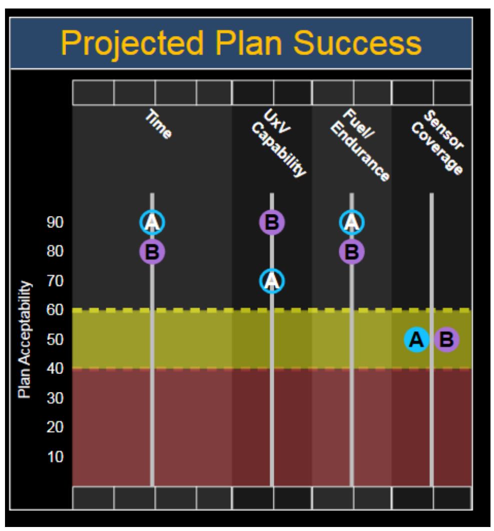

Fig. 2 Projected plan success tile from Stowers and associates (2016)

Comparing and appraising multiple automation-developed plans can be a daunting task for the human teammate. In order to mitigate the difficulty of this task, the work of Behymer (2015), Stowers (2016), and their respective associates was used to inform the design of a comparative display for route-planning, herein referred to as the Comparator Display. The Comparator Display, depicted in Fig. 3, shows crew members how multiple routes differ in terms of mission-relevant criteria. These criteria can include information pertaining to the RCV, such as sensor coverage or fuel (as seen in Stowers et al. 2016), as well as more traditional criteria such as information pertaining to METT-TC. The selected routes' projected ratings according to crewmember-selected criteria are mapped vertically on this plot, with each factor compared to the hypothetical shortest route, which is represented using a horizontal line in the middle of the plot. Similarly, the display permits the autonomy (if available) a means to communicate to the crew the differences between proposed COAs (goals), as well as the environmental and vehicle state information and weighting used during its decision-making process (general reasoning).

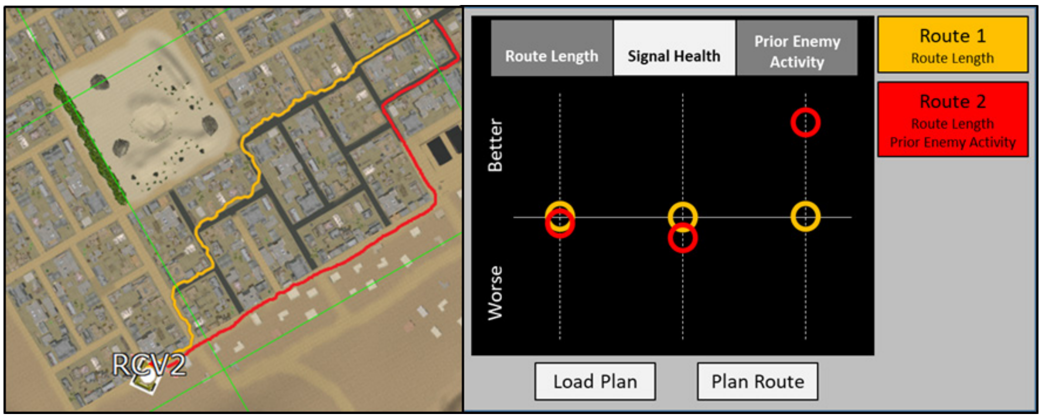

Fig. 3 Route planning using the TRP and prototype Comparator Display. Left panel shows two routes proposed by the RCV's autonomy. Right panel shows the initial prototype of the Comparator Display. Legend on the right indicates color code as well as the information used to generate the route. Each data series (X-axis) indicates a different METT-TC-relevant criterion. Vertical positions of route symbols (circles) indicate estimates of the magnitude of differences between the two routes, clearly showing which route is preferable according to specific criteria. For example, the route shown in red provides slightly worse signal health, but avoids prior enemy activity locations much better than the route in yellow.

2.2 WMI Implementation

The TRP uses a hierarchical optimal search algorithm to generate routes, while the Comparator Display uses the underlying data structures to compute values for the resulting visualization. In order to implement these transparency concepts in the WMI, it was first necessary to generate the required data structures. The underlying algorithm is HPA*, a hierarchical implementation of the A* algorithm (see Botea et al. [2004] for more information), an existing route-planning architecture that requires a graph structure. To generate the base environment used by the TRP, we created a custom simulation environment using the Autonomous Navigation Virtual Environment Laboratory (ANVEL; Quantum Signal), and then transformed the actual simulation environment into rasterized representation of 3-m square voxels, yielding 967- × 967-voxel matrices layered on top of one another to create an array. HPA* was selected because it is a computationally efficient solution that works with larger or higher resolution. Each layer of the array represents a different cost map: the first layer is a mobility cost map consisting of Light Detection and

Ranging (LIDAR)-measured obstacle data in which traversable grid cells have cost = 1 and obstacles have cost = infinite, the second layer contains the cost associated with electromagnetic (EM) threats in the environment, and the third layer contains costs associated with kinetic threats. The data structure used in this experiment was selected to serve as a proof of concept to represent potential sources of threat posed by near-peer adversaries, and the algorithms and data structures used are scalable to much higher complexity in the future. However, as this is intended to be a simple proof of concept, we acknowledge that better algorithmic approaches may exist to accomplish this task. In theory, layers of the input array could represent information about any environmental, team, or vehicle state factor that might impact the cost or impact of various routes to an objective.

To create the mobility layer, we placed a simulated LIDAR sensor in every voxel in the environment that tested whether that voxel contained an obstacle for mobility within 3 m in any direction; voxels containing obstacles were assigned a value of infinity, while clear voxels received a cost value of 1. We further refined this mobility map to encourage the algorithm to generate safe routes that would prefer to keep a distance between the route and obstacles by applying a Gaussian smoothing kernel to the mobility map so that cost would increase with proximity to obstacles in the environment.

The EM and kinetic threat layers (Fig. 4) were created by generating a uniform distribution of discrete threat locations over the problem space (n = 10 and 15, respectively), and then applying a Gaussian smoothing kernel to those values to represent the threat's degrading effect with distance. These data structures required for the TRP are known a priori in simulation, but could be constructed similarly in real-world settings using a combination of geographic information system (GIS) terrain data, database information (e.g., from the Tactical Integrated Ground Reporting system), and information generated in real time by various sensing technologies or even by other Soldiers using the WMI (i.e., by placing Battle Space Objects [BSOs], in the environment). In the future, these data should be modeled on real-world EM and kinetic effects of weapons balanced against the capabilities of the vehicle in the simulation once they are known.

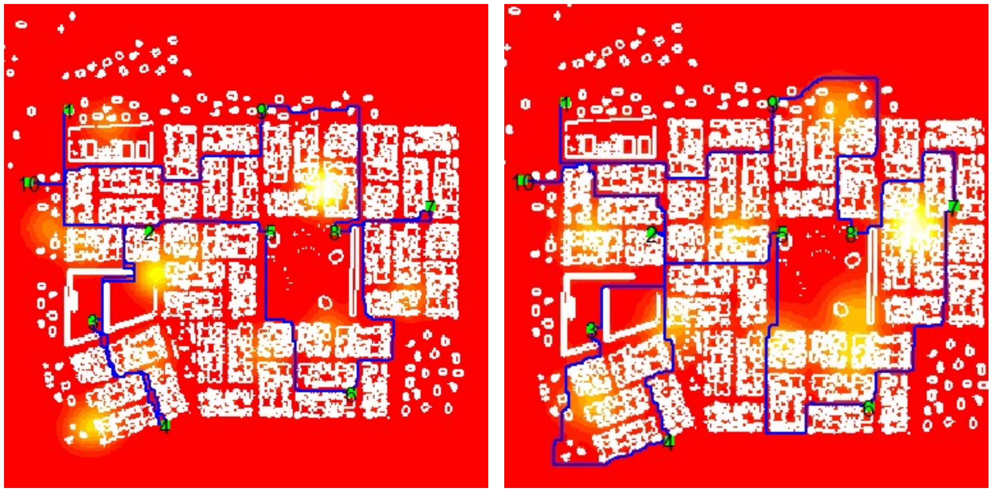

Fig. 4 Heat maps indicating the cost associated with kinetic (left panel) and EM (right panel) threat locations used for the experiment. Rally points are indicated as numbered green boxes on each panel. The red voxels in each panel do not contain obstacles and have an assigned cost of 1. White voxels contain obstacles and are assigned a cost of infinity. Values between 1 and infinity are depicted by color gradient on a spectrum from red (cost = 1) to white (cost = infinite). Blue lines between rally points indicate routes generated by the TRP when provided each cost map, showing the TRP's ability to selectively avoid high-cost areas of the environment.

2.2.1 Transparent Route Planner

To generate routes using the TRP, the crew member first selects a destination and then specifies the list of criteria to consider in the route using the Comparator Display (see Section 2.2.2). Third, the TRP builds an integrated cost map graph based upon the mobility map and those spatialized features by summing across the Z dimension (cost map layers) of the array. Finally, the TRP uses the integrated cost map graph to generate a least-cost route from the vehicle's current location to the intended destination by balancing the costs of each specified feature (see Fig. 4).

2.2.2 Comparator Display

The Comparator Display is used as both a transparency display and a graphical user interface (GUI), allowing the crew member to compare clear, intuitive information about multiple routes on different mission-relevant criteria as well as to select criteria to use during the planning process. The Comparator Display is implemented in the R Statistical Computing Language v. 3.5.2 (R Core Team 2014) for testing, as well as in the WMI using a custom Python v. 3.7 module (Python Software Foundation 2018).

During the route-planning process (see Section 2.2.1), the crew member specifies the list of criteria (currently via Python script; a GUI is under development that will permit this action in real-time in the WMI) that the TRP will use to weight the costs of various navigation options and plan viable routes. Each time a route file is generated between the vehicle's current location and its goal destination, it is cached along with the set of criteria used to build it. In this way, the crew member can plan and display multiple routes simultaneously, using the Comparator Display to determine the quality of each route, relative to one another, on each of the METT-TC-related criteria. The METT-TC characteristics (X-axis) of each route under consideration are displayed as color-coded points on vertical axes (Y-axis) indicating the suitability of that route on each of the mission criteria (see Fig. 5).

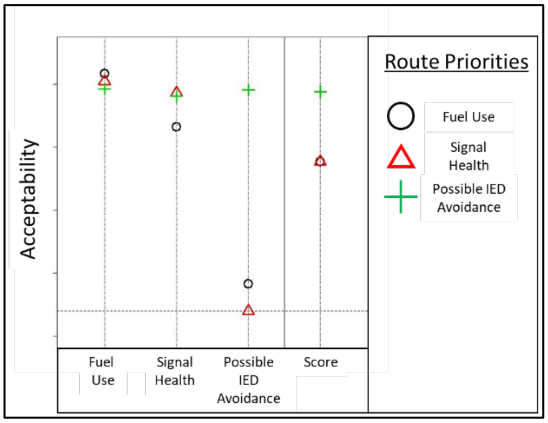

Fig. 5 Comparator Display showing the interface elements and scaling. Each of the three color-coded icons (○, △, +) indicates a single route. Each route's score on each of the criterion values (X-axis data series) is shown as its Y-axis position (higher values are more acceptable outcomes). The Route Priorities panel (right) serves as a legend for each of the three routes created by the user. In the present example, the user has created three routes using each of the following criteria, Fuel Use (black circles; Mobility cost map), Signal Health (red triangles; EM Threat cost map), and Possible IED Avoidance (green crosses; Kinetic Threat cost map). While the TRP used only a single criterion to plan each route, the routes are evaluated on all criteria. In this example, all routes are similar in terms of their fuel use requirements. The routes indicated by red triangles and green crosses are roughly equal in terms of signal health and are better than the route indicated by the black circle, which minimizes fuel use. The route indicated by the green cross is substantially better at avoiding possible improvised explosive devices (IEDs) than the other two routes and is the best route overall (right-most data series; "Score") according to these criteria. Importantly, this figure represents the latest version of the Comparator Display, which differs from the display used in the evaluation described in Section 3 in terms of its scaling and information presented.

The routes can be compared to one another in terms of any quantifiable characteristic, and the display is intended to show magnitude estimates of the differences to be scale–invariant and thus conducive to quick assessment of the relative costs of each route. In order to obtain values on each of the criteria for each of the routes, the route file's grid coordinates are used to pool the cumulative cost associated with each criterion. If required, these values could also be calculated by estimating data over the course of the route (as would be the case for fuel use or speed). These values may be used in any number of ways; here, we implement a simple summation function to compare how much cumulative risk, time, fuel use, and so on is necessitated by each route. Therefore, each route has its own value on each of the criteria. Using the minimum value produced among all routes for each of the criteria as a baseline, the Comparator Display scales these values in terms of their magnitude relative to the baseline,

$$f(x) = \left(2 - \frac{x}{\min(x)}\right) + abs(\min\left(2 - \frac{x}{\min(x)}\right)) \tag{1}$$

Rescaling in this way permits us to generate magnitude values that are interpreted similarly to bar plots. Each data point represents a magnitude estimate of the quality of each route versus the baseline (or worst) route. This section describes the most recent version of the Comparator Display while the evaluation in Section 2.3 employed an older version with different Y-axis scaling.

2.3 Concept Evaluation: Current Study

In order to evaluate the Iteration 1 transparency concepts we experimentally employed a within-subjects design and tested multiple participants' performance levels using both the baseline WMI and a Transparent WMI incorporating both the TRP and Comparator Display. Our goal was to determine the extent to which these augmented versions of the WMI improved crew members' comprehension of the RCV's actions and intentions (TRP-equipped WMI), as well as its goals and general reasoning (Comparator Display-equipped WMI) over baseline.

3. Methods

3.1 Participants

Participants (*n* = 10) for the present study were recruited as volunteers from the CCDC ARL civilian work force. As this test and evaluation were deemed by the CCDC ARL Institutional Review Board to be exempt and benign, participants were not required to complete an Informed Consent form, but were verbally informed of their right to leave the experiment at any time for any reason.

3.2 Experimenter Equipment

Participants completed their tasks using a standard desktop computer equipped with a keyboard and mouse. Stimuli were presented using custom experiment software created in the Psychology Experiment Building Language (PEBL) v. 0.14 (Mueller 2014). Routes were generated through a custom simulation environment created using the ANVEL Quantum Signal for use as stimuli using the TRP implemented in the WMI. Accompanying Comparator Display stimuli were created in R.

3.3 Experiment Design

In this iteration, we compared two interface elements to the WMI baseline, the TRP and the Comparator Display. In order to examine the effects that these interface elements had on participants, we employed a within-subjects design with two successive comparisons. In the first block, participants use both the baseline WMI and the TRP. In the second block, participants use the baseline WMI and the Comparator Display.

3.4 Procedure and Analysis Plan

The experiment software was designed to present the TRP evaluation first, followed by the Comparator Display evaluation, as separate experimental blocks.

In the first experimental block, participants completed 25 trials, alternating between one where they were shown a route produced by the baseline WMI and one where they were shown the route produced by the TRP (see Fig. 1 for a visual example). In each trial, after participants were shown the RCV's route, they were asked to draw the route they believed the RCV would take by clicking on the screen; each click added another segment to the route, which was connected to the previous segments. If participants made a mistake, they could remove the previously placed segment by clicking the Back Up button. When finished creating the route, participants clicked the Done button to advance to the next trial. In order to judge the effectiveness of the TRP on participants' comprehension of the RCV's actions and intentions, participants' route predictions when they used the TRP were compared with the predictions made when they used the baseline WMI. Data were analyzed on a trial-wise basis by using the Algorithm for finding the Least Cost Areal Mapping between Paths (ALCAMP; Mueller et al. 2016), which compared the participant's response (i.e., their drawn route) to the route that the RCV would actually execute in that situation (i.e., the ground truth). The ALCAMP algorithm creates a polygon from two arbitrary paths, the area of which corresponds to the dissimilarity between those paths. In the present case, this area corresponds to the amount of error in the response; if the participant's response was very similar to the ground truth, this area would be very small compared to if the response was very different. The TRP shows crew members the actual route that the RCV will execute, in detail, so error in this condition should be interpreted as error in the drawing task.

The second experimental block tested the Comparator Display. Participants were presented with a map containing EM and kinetic threats represented as MIL-STD

2525D (2014) BSOs, and routes created by the TRP (Fig. 6) using different cost maps. While viewing the map, participants completed a 4-alternate forced choice (4-AFC) task in which they attempted to deduce which information was used by the RCV to create the route (i.e., goals and general reasoning). The criteria in this task included 1) mobility, 2) EM threats, 3) kinetic threats, or 4) all threats. Participants clicked the appropriate button to submit their response, and then clicked the Done button when they had made their decision. After completing 44 trials in the baseline condition (i.e., without the Comparator Display; Fig. 6), participants completed the same 44 trials using the Comparator Display (Fig. 7). Trial presentation order was randomized within each condition to control for systematic order effects. In order to evaluate the Comparator Display in the 4-AFC task, only a single route was shown at a time, and the Y-axis criterion values were scaled using the shortest route (i.e., Mobility) as a baseline. This initial version of the Comparator Display differed from the current implementation in two ways: first, it displayed criterion values for only one route at a time (necessary for the evaluation used here), and second, it used the shortest route's criterion values as baseline (i.e., the Y-axis is correctly interpreted as the route's score on that criterion value relative to the shortest path). Participants reported that this scaling was difficult to understand, thus the modifications made to Y-axis scaling described in Section 2.2.2 are the result of qualitative user feedback.

Fig. 6 Data collection tool designed to test the TRP. In this baseline WMI example, the participant is attempting to draw the route that they believe the RCV will take (red line) given the plan offered by the WMI (yellow line).

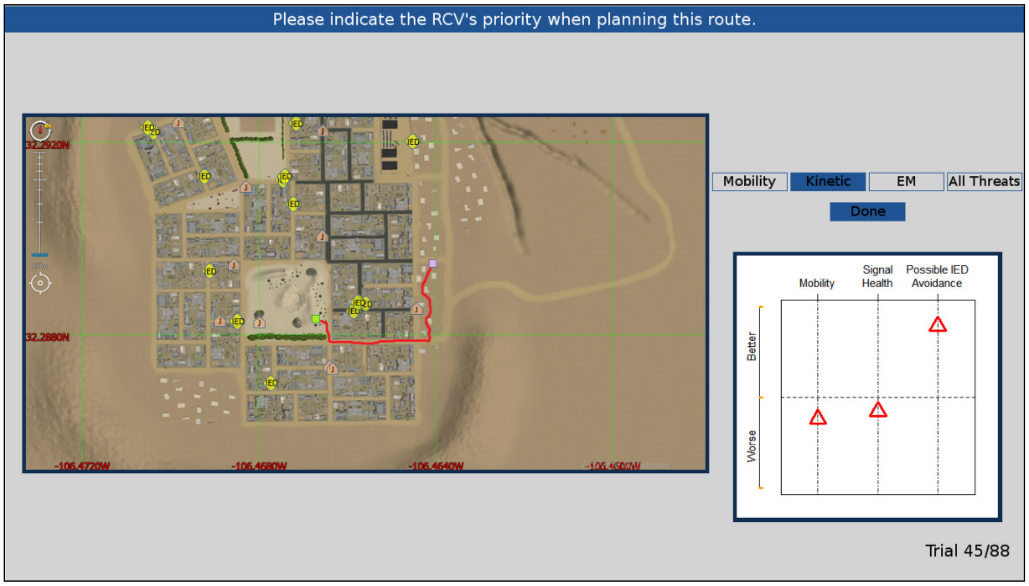

Fig. 7 Comparator Display evaluation, shown in the Comparator Display condition. The map is shown in the left panel. EM threats are depicted as red BSOs with a J in them, for signal jammer. Kinetic threats are depicted as yellow BSOs labeled yellow with the letters IED in them, for possible IEDs. The Comparator Display on the right shows the quality of the route on three criteria: mobility, signal health, and possible IED avoidance. In this example, the Comparator Display clearly shows that the route avoids kinetic threats.

Performance in this task was measured as percent correct on the 4-AFC task, corresponding to the participant's comprehension of the RCV's goals and general reasoning in terms of which criteria its underlying route-planning algorithm prioritized during the planning process.

4. Results and Discussion

Results are presented here in terms of percent improvement, congruent with the DP1 goal constructs described in Section 1. We report both the results of traditional null hypothesis significance tests as well as Cohen's *d* as a measure of effect size.

4.1 TRP Evaluation Results

Participants' ALCAMP-derived error values provided a means of testing their ability to predict the RCV's future actions and intentions (Fig. 8).

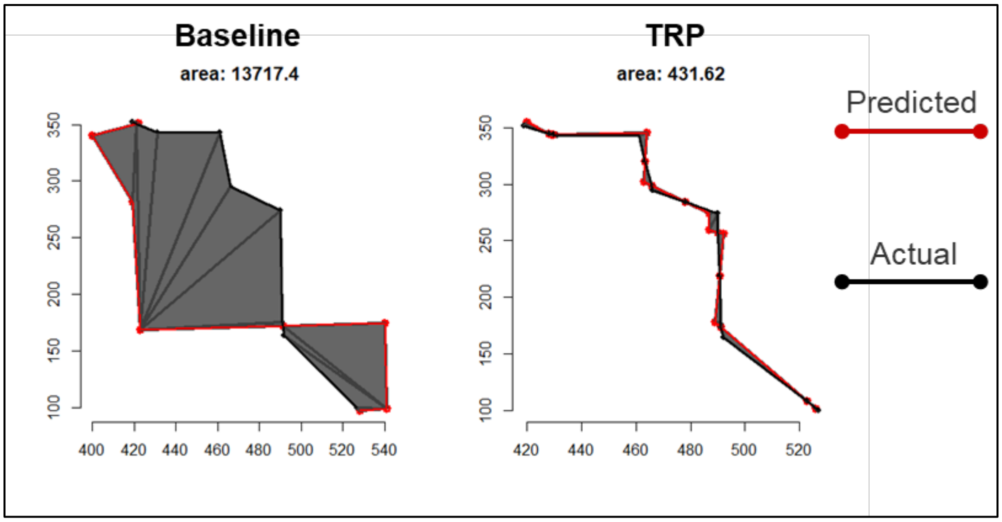

Fig. 8 Example ALCAMP analysis output for the baseline and TRP conditions. Red lines in each pane correspond to the route drawn by the participant. Black lines correspond to the RCV's actual route (ground truth). The area of the polygon in voxels created by ALCAMP is shown above each plot and constitutes our error metric.

Participants' performance on the route-planning task was compared using paired-samples T-tests. On average, participants' drawn routes were significantly more similar to the RCV's actual routes when using the TRP ($m = 1195.36$, $SD = 1830.30$ voxels error) versus the baseline WMI ($m = 6866.46$, $SD = 6960.23$ voxels error), $t(9) = 8.60$, $p < 0.001$, Cohen's $d = 1.78$. That is, participants were able to predict the RCV's future mobility with much greater accuracy using the TRP, relative to baseline. This difference corresponded to a 319% improvement over baseline when using the TRP. These results are shown both as a histogram, for all of the trials, and a boxplot in Fig. 9.

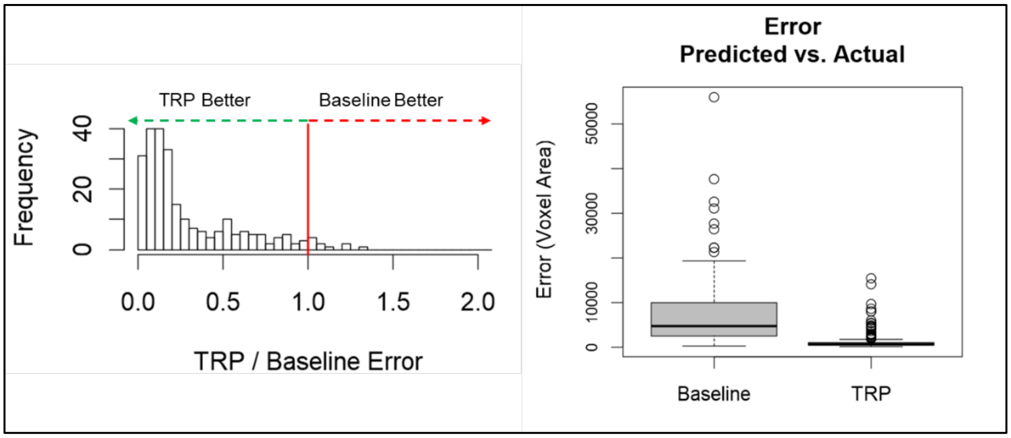

Fig. 9 Results of TRP evaluation. Left panel shows trial-wise ALCAMP-derived error values as a proportion, TRP vs. baseline. If TRP and baseline error was equal, the trial received a value of 1; if TRP was better than baseline was <1. In nearly all trials, participants' knowledge of RCV future mobility actions and intentions was superior using the TRP. The panel on the right shows the pooled error in each of the conditions depicted as box and whisker plots; each box represents the range between the first and third quartiles, the median is depicted as a black line, and the top and bottom whiskers indicate maximum and minimum values excluding outliers (hollow circles).

4.2 Comparator Display Evaluation Results

The Comparator Display evaluation results are reported as percent correct on the 4-AFC task, indicating their ability to infer the RCV's goals and general reasoning process. All but one participant showed a net improvement when using the Comparator Display (Fig. 10). On average, participants showed a 15.89% net improvement, $t(9) = 4.17$, $p = 0.002$, Cohen's $d = 1.34$, with the maximum improvement being 43.49% over baseline.

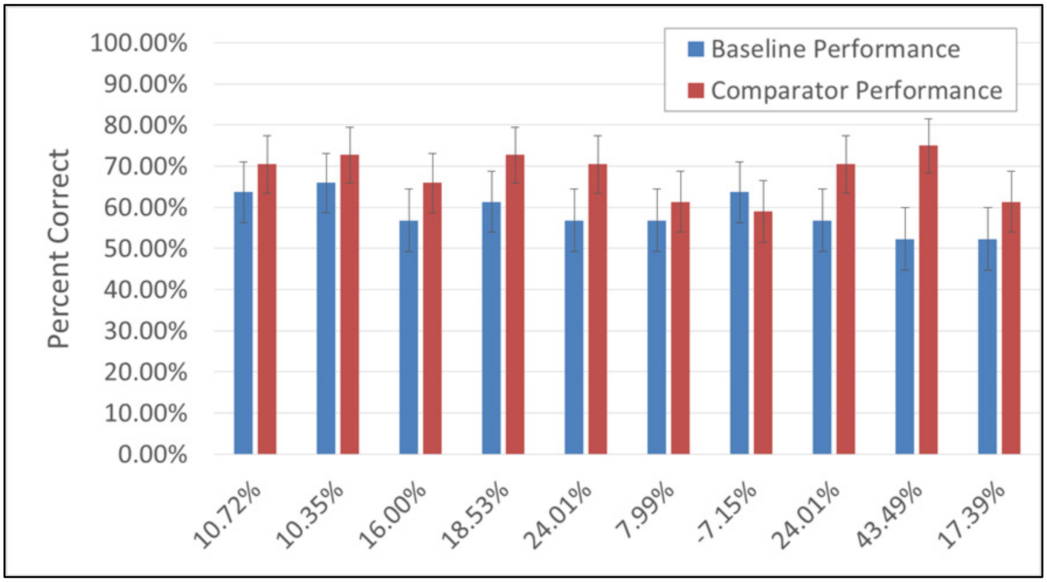

Fig. 10 Each participant's performance on the 4-AFC task using the baseline WMI (blue bars) and Comparator Display (red bars). Whiskers on each bar indicate standard error. Percentages below each pair of bars indicate the net change in performance when using the Comparator Display vs. the baseline WMI. All but one participant showed a net benefit in comprehension of the RCV's goals and general reasoning when using the Comparator Display.

5. Conclusions

Successful deployment of the RCV in the NGCV platform will depend upon the crew members' ability to make decisions using high-dimensional information at fast-paced optempo while working in unstructured, complex operational environments. The crew must be able to accomplish these tasks rapidly, while simultaneously maintaining comprehension of and trust in the actions, intentions, goals, and general reasoning processes of autonomous agents under their control. The technologies evaluated herein are intended to improve that comprehension, as well as to improve the speed at which crew members are able to accomplish their goals. As the autonomous capabilities of RCVs in the NGCV platform increase, we further expect that the Comparator Display's ability to present complex, high-level information in a clear and intuitive manner that will empower rather than burden the crew member will become more and more essential, thereby increasing situation awareness, calibrating appropriate trust in the system, and improving the efficiency and outcomes of future mission planning.

5.1 Refining the Transparent Route Planner

The goal of the TRP was to provide a transparent single-click route-planning solution that improves crew members' understanding of the autonomy's future actions and intentions. A secondary goal of the TRP was to enable an intelligent supervisory control scheme so that crew members could direct the vehicle to destinations in the environment with minimal interaction (i.e., without overloading the crew member by requiring them to plan the entire route manually). The results of this evaluation showed that the TRP dramatically improved participants' understanding of the unmanned vehicle's future actions and intentions relative to the baseline system. Further, based on the current findings, additional modifications will be made the TRP to improve its integration with the Comparator Display for planning (see Section 5.3, Future Directions).

5.2 Refining the Comparator Display

The goal of the Comparator Display was to provide the crew with insight into the autonomy's goals and general reasoning processes. The initial findings were promising; however, the experiment design used to evaluate the Comparator Display was necessarily simplified from its current version shown in Section 2.2.2. In this evaluation case, the Comparator Display used the shortest route calculation to generate the values it displayed. Without the Comparator Display, participants in this study needed to infer the system's goals and general reasoning processes from the values shown on the Comparator Display alone; they did not have access to other candidate routes, so would not be able to make generalized comparisons using physical attributes of the routes themselves.

In the WMI implementation, the Comparator Display is intended to be used to compare multiple routes that are all visible simultaneously, allowing the operator to compare several competing routes on factors pertinent to mission success. Based upon qualitative user feedback, we revised the scaling so that the criterion values would use the lowest value in the series as a baseline rather than using the shortest route's criterion scores. This change is intended to improve the usability of the Comparator Display by making the Y-axis more intuitive and interpretable. Furthermore, the display aggregates the individual factor scores into one overall acceptability score for each route, giving crew members the opportunity to make rapid at-a-glance comparisons. Converting the acceptability scale (Y-axis) to an overall, separate, and consistent scale is vital to user understanding and efficacy over repeated uses (Gillan et al. 1998). This vertical axis must be consistent across uses; otherwise, the user will have to take additional processing time to ensure they understand how the ranking is being reported, and it could result in inflating (or underestimating) factor differences that could, in turn, impact effective use of the Comparator Display.

5.3 Future Directions

The use of the Comparator Display for route-planning applications is still in its early stages, yet it shows much promise in improving operator understanding of agent intent, reasoning, and planned outcomes. Future directions for the development of the Comparator Display should allow for more flexibility, both in which factors are considered and how each factor is weighted in the final decision, as well as converting the acceptability scale (Y-axis) to other potential scores, such as overall mission success rather than compared to other factor scores.

Specific factors to consider will vary depending on mission requirements; as such, the operator should have the ability to include or exclude factors that are unsuitable to current tasking. Some predefined factors for mission type can be available by default (e.g., most missions in contested areas would prioritize speed, stealth, and communications stability). However, there may be instances where the predefined factors are not needed, are unsuitable (e.g., the area is known to be free of kinetic threats to that unit, so another factor may be more informative as to mission success), or are not available because the TRP does not have the necessary information.

Additionally, as novel technologies become more complex, adaptive, and individualized (e.g., see DeCostanza et al. 2018), novel measures and predictions of team performance under specific operational circumstances may also be leveraged for optimal and transparent route planning: for example, rather than predicting which route will be shortest, fastest, or most fuel-efficient, future applications of the TRP and Comparator Display may communicate which routes will be most conducive to a specific team or specific crew member's capabilities. This will allow routes to be planned and presented that optimize the crew's capability to meet and exceed mission demands or that will allow crews to predict and allocate resources optimally for specific situations. As intelligent, autonomous systems and machine learning become more common in operational situations, technologies like the TRP and Comparator Display will be absolutely essential to fielding novel technologies quickly, efficiently, and in a way that Soldiers will be empowered rather than burdened by rapidly increasing informational load.

How each factor is weighted in the agent's decision should be transparent to the operator, and the operator should have the ability to redefine weights if required. Mission requirements may update during execution, and so changes to mission parameters may need to be input manually by the operator.

The experimental evaluation presented herein tested the TRP and Comparator Display independently in order to evaluate their individual effectiveness with respect to the DP1 goal constructs. In the WMI, however, the two are intended for use as an integrated route-planning system. Future testing will evaluate the TRP and Comparator Display together in the context of mission planning, as well as dynamic replanning, during the mission execution phase.

6. References

Behymer KJ, Mersch EM, Ruff HA, Calhoun GL, Spriggs SE. Unmanned vehicle plan comparison visualizations for effective human-autonomy teaming. Procedia Manufacturing. 2015;3:1022–1029.

Botea A, Müller M, Schaeffer J. Near optimal hierarchical path-finding. Journal of Game Development. 2004;1(1):7–28.

Chen JY, Lakhmani SG, Stowers K, Selkowitz AR, Wright JL, Barnes M. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. Theoretical Issues in Ergonomics Science. 2018;19(3):259–282.

DeCostanza AH, Marathe AR, Bohannon A, Evans III AW, Palazzolo ET, Metcalfe JS, McDowell K. Enhancing human-agent teaming with individualized, adaptive technologies: a discussion of critical scientific questions. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2018 May. Report No.: ARL-TR-8359.

Dijkstra EW. A note on two problems in connexion with graphs. Numerische Mathematik 1. 1959:269–271

Gillan DJ, Wickens CD, Hollands JG, Carswell CM. Guidelines for presenting quantitative data in HFES publications. Human Factors. 1998;40(1):28–41.

Lee D, Ahn J. Integrated optimization of planetary rover layout and exploration routes. Engineering Optimization. 2017;50(1):164–182.

Lyons JB. Being transparent about transparency: a model for human-robot interaction. Paper presented at the 2013 AAAI Spring Symposium Series; 2013; Palo Alto, CA.

MIL-STD 2525D. Department of Defense interface standard: joint symbology standard. Fort Meade (MD): Defense Information Systems Agency Standards Management Branch; 2014 June 10.

Mueller ST. (2014). PEBL: the psychology experiment building language. Ver. 0.14. Houghton (MI): Michigan Technological University; 2014 [accessed 2014 June]. [Computer experiment programming language]. http://pebl.sourceforge.net.

Mueller ST, Perelman BS, Veinott ES. An optimization approach for mapping and measuring the divergence and correspondence between paths. Behavior Research Methods. 2016;48(1):53–71.

Myer, MR, Lojka, JR. On risk: Risk and decision making in military combat and training environments. Naval Postgraduate School, Monterey CA, Defense Analysis Department, 2012.

Otten N, Wettergreen D, Whittaker W. Strategic autonomy for reducing risk of sun-synchronous lunar polar exploration. In: Hutter M, Siegwart R, editors. Field and Service Robotics; 2017 Sep. Cham (Switzerland): Springer International Publishing. p. 465–479.

Perelman BS, Evans AW III, Scheafer KE. Mental model consensus and shifts during navigation system-assisted route planning. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2017;61(1):1183-1187. doi: 10.1177/1541931213601779.

Perelman BS, Lakhmani S, Wright JL. Human-autonomy teaming essential research program project 2: transparent multimodal crew interface designs. Technical note 1: general overview. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); 2020 Jan. Report No.: ARL-TN-1003.

Python Software Foundation. Python Ver. 3.7.3. San Diego (CA): Python Software Foundation; 2018. http://www.python.org/.

R Core Team. *R:* a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing; 2014. ISBN 3-900051-07-0. http://www.R-project.org/.

Stowers K, Kasdaglis N, Newton O, Lakhmani S, Wohleber R, Chen J. Intelligent agent transparency: the design and evaluation of an interface to facilitate human and intelligent agent collaboration. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2016;60(1):1706-1710. doi: 10.1177/1541931213601392.

Tufte ER, Robins D. Visual explanations. Cheshire (CT): Graphics Press; 1997. p. 52.

Vessey I. Cognitive fit: a theory-based analysis of the graphs versus tables literature. Decision Sciences. 1991;22(2):219–240.

List of Symbols, Abbreviations, and Acronyms

4-AFC	4-alternate forced choice
ALCAMP	Algorithm for finding the Least Cost Areal Mapping between Paths
ANVEL	Autonomous Navigation Virtual Environment Laboratory
ARL	Army Research Laboratory
BSO	Battle Space Object
CCDC	US Army Combat Capabilities Development Command
COA	course of action
DP	Decision Point
EM	electromagnetic
ERP	Essential Research Program
GIS	geographic information system
GUI	graphical user interface
HAT	Human-Autonomy Teaming
IED	improvised explosive device
LIDAR	Light Detection and Ranging
METT-TC	mission, enemy, terrain, troop availability, time, and civilian
NGCV	Next Generation Combat Vehicle
PEBL	Psychology Experiment Building Language
RCV	robotic combat vehicle
TRP	Transparent Route Planner
WMI	Warfighter Machine Interface

1 (PDF) DEFENSE TECHNICAL
INFORMATION CTR
DTIC OCA

1 (PDF) CCDC ARL
FCDD RLD CL
TECH LIB

1 (PDF) CCDC ARL
FCDD RLH B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

1 (PDF) CCDC ARL
FCDD HSI
J THOMAS
6662 GUNNER CIRCLE
ABERDEEN PROVING GROUND MD
21005-5201

1 (PDF) USAF 711 HPW
711 HPW/RH K GEISS
2698 G ST BLDG 190
WRIGHT PATTERSON AFB OH 45433-
7604

1 (PDF) USN ONR
ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 (PDF) USA NSRDEC
RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 (PDF) OSD OUSD ATL
HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

12 (PDF) ARL
FCDD RLH
J LANE
Y CHEN
P FRANASZCZUK
K MCDOWELL
A MARATHE
B PERELMAN
FCDD RLH BD
D HEADLEY
FCDD RLH FA
A DECOSTANZA
FCDD RLH FB
A EVANS
FCDD RLH FC
J GASTON
FCDD DAH C
L GARRETT
FCDD DAH W
F MORELLI