Degree of Automation in Command and Control Decision Support Systems

Ryan M. Robinson
Department of Electrical Engineering
University of Washington
Seattle, WA, USA
Email: rymicro@uw.edu

Michael J. McCourt
Research and Engineering Education Facility
University of Florida
Shalimar, FL, USA

Amar R. Marathe
Human Research & Engineering Directorate
Army Research Laboratory
Aberdeen, MD, USA

William D. Nothwang
Sensors & Electron Devices Directorate
Army Research Laboratory
Adelphi, MD, USA

Emily A. Doucette
and J. Willard Curtis
Munitions Directorate
Air Force Research Laboratory
Eglin, FL, USA

***Abstract*—This paper investigates the effects of integrating automation into the various stages of information processing in a military command and control scenario. Command and control (C2) is an extreme decision-making paradigm characterized by high uncertainty, high risk, and severe time pressure. We introduce a principled approach to decision support system (DSS) design that specifically addresses these issues. Our approach establishes the principles of communicating *confidence* in sensor estimates and *consequence* of actions in an intuitive, timely manner. We hypothesize that automation designed to communicate confidence and/or consequence will improve task performance over systems that neglect these concepts. Toward this end, human-subjects experiments were conducted to compare the effects of displaying confidence/consequence information in a C2 target-tracking and interdiction scenario. Four variations of a decision support interface were designed, each with a distinct "degree of automation": (i) an instantaneous sensor measurement visualization (baseline), (ii) a confidence-based visualization, (iii) a confidence- and consequence-based visualization, and (iv) a confidence- and consequence-based visualization with explicit decision recommendations. While increasing automation generally improved results, the inclusion of consequence information did not have a major effect, perhaps because the scenario was overly-simplified.**

I. Introduction

Human decision-making is increasingly paired with automated decision support in a number of applications, such as healthcare, business management, financial analytics, online shopping/entertainment, anti-terrorism, politics, and athletics. In the development of any such human-machine system, it is critical to specify roles for the human and automation that achieve performance objectives more efficiently and *robustly* (e.g. resilient to human/automation failures) than either in isolation. As automation/autonomy proliferates into increasingly complex application spaces, the clear boundaries between roles become less clear, and designers must work to define principled frameworks that successfully and optimally leverage agent capabilities to offset the limitations of the human and machine elements. One such application space is military command and control (C2), a paradigm uniquely characterized by uncertainty, risk, and time pressure. While ethical standards and current policies [1] require that a human makes the ultimate decision, automation may improve task performance by intelligently supporting various stages of the decision making process.

Command and control distinguishes itself from generalized decision-making scenarios due to the coupling of several factors that impose great cognitive demands on the human operator. First, a wide range of C2 operations deal with high uncertainty in the environment and the outcomes of decisions. Humans are known to act suboptimally when provided with uncertain information [2]. Second, entrusting operators with high-consequence decisions, including the use of lethal force and the potential for collateral damage, increases psychological stress [3]. Third, severe time pressure can drastically decrease task performance by limiting time to gain situational awareness [4] and formulate a response. Fourth, decision-making is often conducted under fatigue, which may degrade task performance [5], [6], and in related military cases has been found to cause operator "burnout" [7], [8].

Automated decision support systems (DSS) for C2 must be tailored to counteract the above tendencies and augment human performance under these unique cognitive demands. Due to relative strengths in computational power and repeatability, DSS have been used to lighten the burden on the operator [6] by filtering and integrating information. However, it is not clear *how* to best communicate the integrated information to the operator to enhance decision-making. Poor interface design can actually degrade performance; for instance, studies of unmanned aerial vehicle pilots have reported that one of the most significant sources of stress is the design of the ground-based interface [3], [8].

Many theories for automated decision support are based on mirroring natural human decision-making processes. While

978-1-5090-1897-0/16/$31.00 ©2016 IEEE

classical behavioral decision theory [9], [10] does not take into account critical aspects of operational settings, "naturalistic" models such as the recognition-primed decision (RPD) model [11], [12], [13] assert that under operational pressure, humans typically do not consider and optimize over many available options, but instead rely on recognition of the situation (*feature matching*) and *mental simulation* to quickly determine a "workable" solution. Inspired by the RPD model, Hutchins et al. developed naturalistic C2 interfaces designed to parallel the cognitive strategies employed by experts [14] to reduce cognitive workload. Their research demonstrated that presenting synthesized graphical information in a naturalistic manner indeed reduces workload and improves performance. Specific improvements over traditional interfaces included (i) displaying tracks (time history) of potential targets, (ii) storing information relevant to identify patterns of tactical activities, and (iii) a visual response management system. The authors also suggest the DSS could be used to highlight missing data and provide alerts.

However, current studies neglect the importance of transparently communicating *confidence* in the quality of available sensor data in the face of high uncertainty. Methods that do not propagate uncertainty when combining raw sensor measurements may mask critical information, potentially leading to incorrect assumptions, sub-optimal decision-making, and degraded performance. Probabilistic approaches to sensor fusion such as Bayesian fusion [15] and Dempster Shafer Theory [16] hold promise for combining uncertain and conflicting sensor data, and are typically straightforward to compute with automation.

Furthermore, most decision support systems also do not adequately report the *consequence* (i.e. expected risk) of possible actions. The probabilistic nature of fusing sensor measurements may require a probabilistic characterization of the risk in selecting an appropriate action. Automation can be employed to assist in this computationally rigorous information integration step to provide the user with real-time updates on the effects of multiple courses of action. In terms of the RPD model, consequence-based support essentially performs the *mental simulation* step (constrained of course by the complexity of the environmental model and the complexity of the cost functions programmed into the DSS). Gonzalez [17] experimentally demonstrated that "feedforward" decision support (allowing the user to test 'what-if' scenarios before acting) improved performance in a dynamic resource allocation task (an industrial plant simulation), while feedback did not.

While the naturalistic decision support paradigm has undoubtedly led to improvements in military interface design, thus far, no clear principles have been established for communicating uncertainty in sensor information to the human operator, or communicating recommendations for course of action. Our prior work in developing a *privileged sensing framework* for human-autonomy integration [18] indicates that inter-agent communication supplemented with metadata regarding the confidence in communicated information and the possible consequences of acting improves sensor fusion and dynamic assignment of agent responsibilities. We hypothesize that increasing the degree to which confidence and consequence are effectively communicated in a C2 decision support system will improve task performance.

The study described here investigates how automation can provide effective decision support to human decision-makers in the context of a dynamic target tracking and engagement scenario. The main contribution of this work is to compare decision support interfaces that either display or withhold metrics of confidence, consequence, and autonomous recommendations for the human operator. Experiments were performed with human subjects taking the role of commanders tasked with locating and interdicting a target while avoiding/minimizing collateral damage in a simulated environment. Four variations of a decision support interface were designed, each with a distinct "degree of automation": (i) an instantaneous sensor measurement visualization (baseline), (ii) a confidence-based visualization, (iii) a confidence- and consequence-based visualization, and (iv) a confidence- and consequence-based visualization with explicit decision recommendations.

II. Scenario and Decision Support Interface Design

A. Experimental Scenario

To compare the effects of varying degrees of decision support automation on human decision-making in real-time risk-based scenarios, a target tracking and engagement scenario was developed. The scenario designates the participant as a military C2 operator with the responsibility of locating and neutralizing a single target moving around a simulated urban environment while avoiding/minimizing collateral damage. Multiple autonomous search agents are provided; these mobile agents are equipped with noisy sensors to identify and localize the target. The operator must use these sensor measurements (along with any available decision support) to decide *if*, *when*, and *where* to interdict the target, continuously weighing information about target location against the risk of potential collateral damage. Successful engagement requires the target to be within the interdiction's area-of-effect (i.e. blast radius); however, the area-of-effect deviates pseudo-randomly from the operator's desired fire location, potentially causing the target to be missed (when it would have otherwise been hit) and potentially causing unintended collateral damage.

The participant was provided with the locations of the search vehicles in real-time, but was given limited information about the location of the target, in the form of intermittent and sometimes incorrect sensor measurements coming from the search vehicles. To simulate additional uncertainty, false positive sensor readings (2% of all negative readings) were simulated to occur randomly with a certain frequency, and the sensed locations of true positive readings were varied by a normal distribution.

B. DSS Interface Variants

The decision support interface consisted of (i) a computer monitor display to communicate information to the operator and (ii) a computer mouse to allow the operator to perform actions. As previously stated, key to the interface design problem are: (i) how to filter and integrate information, and (ii) how to best communicate pertinent information.

Past studies [19], [20] segment the decision-making process into four stages: (i) information acquisition/filtering, (ii) information integration/analysis, (iii) action selection, and (iv) action implementation. Theoretically, these stages can each be independently automated, at varying levels. Onnasch et al. [19] performed a meta-analysis of effects of different levels and stages of automation, finding that increasing degree of automation (in terms of levels, stages, or both) increases routine system performance and decreases operator workload, but negatively impacts robustness to failure.

Inspired by this framework, four different versions of a decision support interface were developed. Each version of the interface either displayed or withheld naturalistic communication of confidence in the information filtering stage, consequence in the information integration stage, and autonomous recommendations in the action selection stage. *Assistance level* (AL) was varied from minimal automation (AL1, none of the support modalities) to a high level of automation (AL4, all three support modalities).

The display portion of the interface was primarily composed of two "maps". The first map, known as the *Belief Map*, displays information about target location. The second map, referred to as the *Risk Map*, visualizes the potential for collateral damage when engaging at a particular location.

Assistance level 1 (AL1) provides the human subject with only minimal information about the location of the target on the Belief Map. A circular region believed to contain the target is displayed when one of the search vehicles either accurately identifies the target (true positives) or falsely identifies the target (false positives). The location of accurate identifications were artificially degraded with noise. This is comparable to the traditional "pin" map, where pins of target locations were manually placed in a map. A Risk Map is also displayed which provides a simple, static indication of collateral damage locations and the level of consequence associated with their damage (indicated by color and intensity). The first autonomy level display is detailed in Figure 1a, where small green circles represent sensors and magenta circles represent recently measured target locations when a sensor generates a positive detection. The magenta circles quickly fade ($\sim$1 s).

AL2 provides the user with the same Risk Map as AL1. However, the Belief Map is altered to process the measurement information from the search vehicles (all of which possess the same measurement noise model) into a probabilistic estimate representing the *confidence* in target location, displayed in the form of a heat map. This heat map, which is generated through a particle filtering methodology [21], conveys information about the confidence in target localization and maintains predictions of target location even after the target has left sensor range. When the target location is certain, the heat map condenses to a small region of high certainty, indicated by a yellow/orange/red signature. When the target location is highly uncertain, the heat map diffuses more evenly and the color a dull blue. The Belief Map is depicted in Figure 1b.

AL3 provides the user with the same Belief Map as AL2, but the Risk Map is modified to provide expected risk (*consequence*) calculations based on the current probabilistic location of the target, location of the hazards, and uncertainty in the engagement region-of-effect. The output of this risk calculation is also presented as a heat map, and is dynamically updated as the target location information changes. The improved Risk Map is shown in Figure 1c. Shades of blue indicate high expected risk (low expected score) and shades of red indicate low expected risk (high expected score).

AL4 provides the user with the same two heat maps as AL3. However, there is additional information provided to the subject in the form of autonomous recommendations, based on the minimum expected risk location and value. The autonomy updates a continuous meter that indicates the strength of its current recommendation. It also displays the position of the recommended attack on both maps, and text to indicate a recommendation ranging from "Fire not recommended" to "Fire Now!" The added recommendations can be seen at the bottom of Figure 1d.

According to the *levels and stages* framework presented in [19], [20], these assistance levels represent a monotonically increasing *degree of automation*. Figure 2 illustrates the successive stages of automation that become partially or fully automated as assistance level increases.

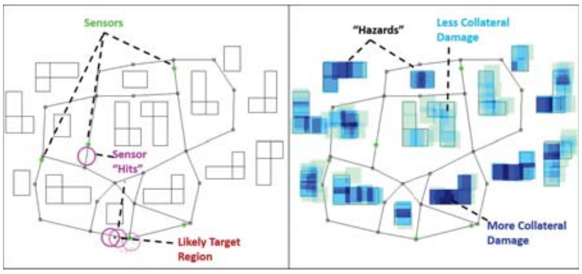
The subject was instructed to use the mouse to click on the desired fire location; this was the same for all ALs. After each individual trial, feedback was provided in form of an assessment of the subjects performance on the trial. A trial score s_t was calculated based on whether a fire decision f was made (yes: -1, no: 0), whether the target was eliminated e (yes: +1, no: 0), and the collateral damage penalty d (from 0 to -24), $s_t = f + 6e + d$. The maximum possible score per trial was +5. An assessment map was also displayed, illustrating the center position of the intended engagement location, the actual engagement region-of-effect, and the true location of the target.

III. Experimentation

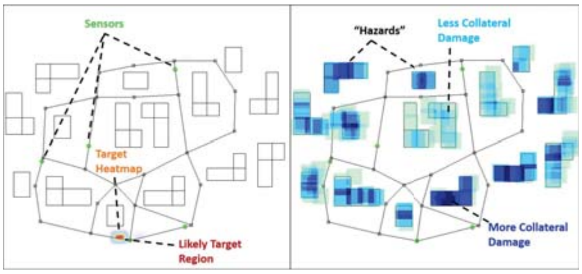
A. Experimental Protocol

A complete experimental session lasted approximately 2.5-3 hours. The subject was brought in and asked to complete a survey concerning demographic, education, and lifestyle factors that were expected to influence task performance.

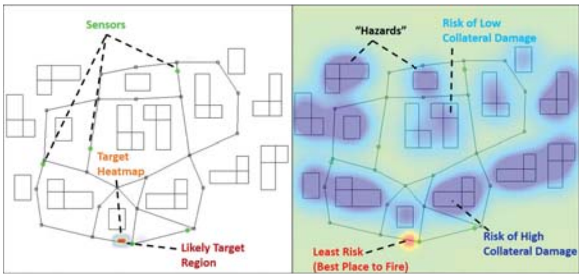
After the initial survey, the subject was seated in front of the interface and experimental equipment was calibrated. Eye gaze data were acquired using a commercially-available eye tracking system (SMI Red 250 Hz [22]) comprised of infrared emitters and two cameras that provide 2 degree-of-freedom gaze tracking, eye blink detection, and pupil diameter



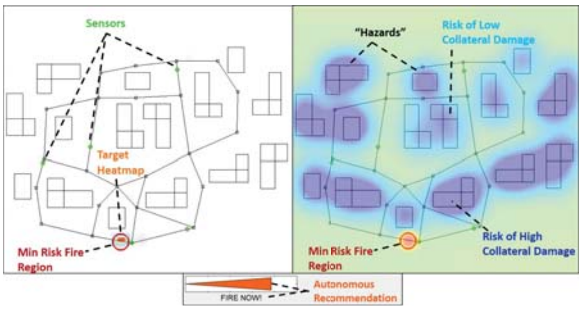
(a) Assistance Level 1



(b) Assistance Level 2



(c) Assistance Level 3



(d) Assistance Level 4

Fig. 1. Belief map and risk map visualizations for each of the four assistance levels (ALs).

monitoring. The subject underwent a very brief calibration session where they were asked to gaze at certain points on the computer screen. A wearable wrist sensor (Affectiva QSensor [23]) was used to collect physiological data from subjects while completing the experimental tasks. Skin conductance (electrodermal activity), motion (3-axis accelerometry), and skin temperature were sensed and recorded. Note that these data sources have not yet been incorporated into our analysis.

Prior to the main experiment, subjects completed a control task. The purpose of the task was to gauge baseline visuo-motor coordination with a computer display and mouse in a

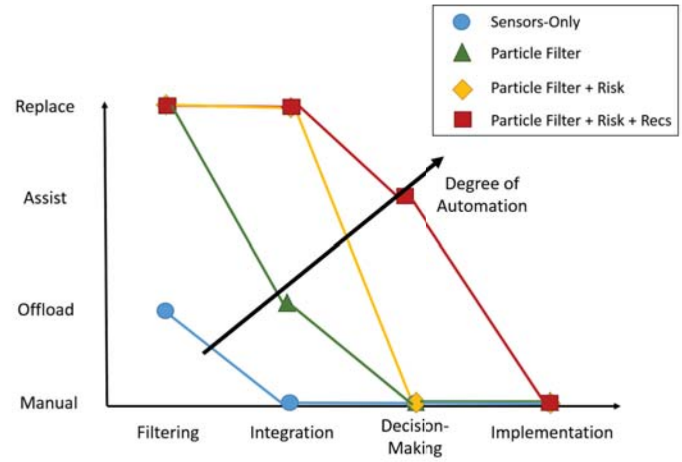


Fig. 2. Approximation of *levels* of automation across the four main *stages* of automation for different DSS assistance levels (ALs) used in the target tracking and engagement scenario

manner relevant to the main experiment. On the display, a red dot (approximately 1/40th the width of the screen) would appear in a pre-selected location. The subject will be instructed to move the cursor, shown as a blue circle (with a radius 2.5 times that of the red dot) over the dot, and click when the dot is fully contained within the blue circle. After a successful click, a new red dot will appear in another pre-selected location, and the task will be repeated for 100 trials. Response time and distance between consecutive targets will be measured for each red dot.

At the beginning of the main experiment and after each of the five experimental blocks, subjects were instructed to record their subjective stress and fatigue levels on a stress visual analogue scale (SVAS) and fatigue visual analogue scale (FVAS), respectively. The SVAS and FVAS have been used in patient populations [24], [25] as well with healthy individuals [26] to quantify changing stress/fatigue levels. These data too have not yet been incorporated into our analysis.

B. Main Experimental Design

The main experiment was composed of five distinct blocks of 120 trials (12 sets of 10 trials), where the first four blocks cycled between the four different assistance levels, and the fifth block was a repeat of the first block, to compare the same assistance level during initial learning and post-learning on the same subject. The order of assistance levels was varied between subjects in order to (i) reduce ordering bias so that conclusions regarding the overall performance of each assistance level are justified, and (ii) compare initial learning (Block 1) results for all four assistance levels. 24 subjects were tested, corresponding to the 24 possible permutations of the order of assistance level in the first 4 blocks (e.g. [1,2,3,4], [1,2,4,3], . . . , [4,3,2,1]).

At the beginning of each set, 10 simulation clips (corresponding to the 10 trials) were loaded, and then presented in sequence. Each individual trial lasted up to 10 s, but would

(a) Age (b) Occupation (c) Education

(d) College Major (e) Compensation (f) Gaming hours/wk

Fig. 3. Pie charts describing subject pool.

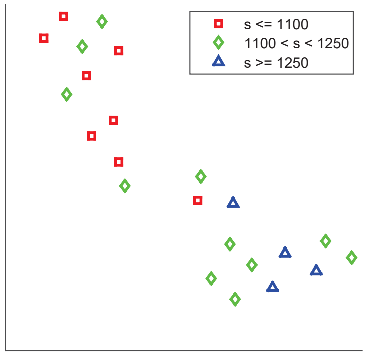


Fig. 4. Two-dimensional visualization of 6 critical subject pool variables using t-distributed stochastic neighbor embedding (t-SNE) [27]. Age, compensation, engineer vs. non-engineer, control task performance, pre-experiment fatigue, and pre-experiment stress variables from the 24 subjects were embedded onto a plane, and color-coded by total score s_T. The general clustering of high-scoring ($s_T \geq 1250$) and low-scoring ($s_T < 1100$) subjects suggests a correlation between score and particular combinations of these variables.

immediately end if and when the subject chose to engage the target. After each trial, the subject had up to 5 s to review their performance assessment from the previous trial. Based on the decision that the subject makes, the target may or may not be neutralized. If the subject did not attempt to engage the target during the alloted 10 s, the inaction was recorded and scored accordingly. Specifically, no action was given a score of zero. This is slightly better performance than an attack that misses the target without causing collateral damage (scored as -1).

The same group of 120 trials was shown to each subject in each block, in order to eliminate the bias from the infinite number of possible simulation conditions. The 120 trials were split up into 12 sets of 10 trials each; the order of trials in each set was randomized to minimize the chance that subjects would recall the same trial from a previous block, but the set order was consistent across all blocks (e.g. Set j of Block i consisted of the same 10 trials as Set j of Block $i + 1$), enabling an unbiased assessment of improvement over time.

C. Subject Pool

Among the pool of 24 subjects, there was a nearly even split between ages 20-29, 30-39, and 40-49 (21 was the lowest age, 49 was the highest). All subjects received some level of higher education, and a majority of subjects received their doctorate. A plurality of subjects majored and have a career in engineering. Due to compensation policies, there was a nearly even split between subjects who were compensated and those who were not. Subjects who were compensated recieved a \$50 base payment for participation and up to a \$50 bonus based on score; these subjects were notified of the payment structure prior to the main experiment.

IV. Results

Here we present the results of experiments demonstrating the relationships between performance and assistance level, as well as other variables that could not be controlled. We plan to further analyze physiological metrics and assess learning rate in subsequent publications.

From a preliminary analysis of survey results and overall performance, a subset of subject characteristics was identified as correlating with performance. Figure 4 depicts a reduced-dimensional visualization of 6 characteristics, categorized according to total score. The high-performing subject group generally consisted of engineers with low self-reported stress/fatigue (prior to the experiment) that were compensated for their participation and performed well on the preliminary control task, whereas the low-performing subject group generally consisted of non-engineers with high stress/fatigue that volunteered to participate.

Subject performance for individual blocks, and thus different assistance levels, was also recorded. Figure 5 displays boxplots separating experimental blocks by order and assistance level. The clearest conclusion is that AL1 (mean blockscore $\hat{s}_b = 148.2$) performance was considerably lower than the other assistance levels, regardless of the block in which subjects performed AL1.

One could reasonably assume that equivalent "difficult" sets of simulation clips will depress scores regardless of assistance level or block number; therefore it is reasonable to expect set number to have a significant effect on set score. A three-way ANOVA test was used to analyze performance, with assistance level, block number, and set number (1-12) as factors. Main effects were obtained for assistance level [$F_{(3,1436)} = 188.82$, $p < 0.0001$], block number [$F_{(4,1435)} = 14.04$, $p < 0.0001$], and set number [$F_{(11,1428)} = 99.73$, $p < 0.0001$], and there was significant interaction between assistance level and block score [$F_{(12,1427)} = 3.21$, $p = 0.0001$] and between assistance level and set score [$F_{(33,1406)} = 3.27$, $p < 0.0001$], but not quite between block number and set score [$F_{(44,1395)} = 1.36$, $p = 0.0587$]. Subsequent multiple comparisons tests indicated several important results. First, subject performance using AL1 is significantly lower than the other three assistance levels.

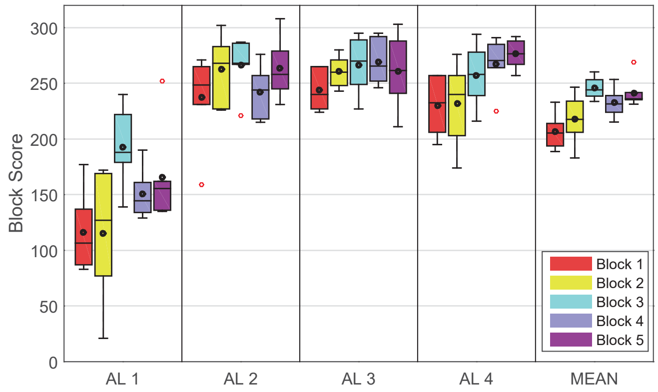


Fig. 5. Boxplots of block score for different assistance levels (ALs 1-4) and different blocks (1-5).

Second, Set 2 generated significantly lower scores than all other sets, and Set 12 generated higher significantly higher scores than all other sets. Third, AL1, Block 3 (the best-performing AL1 block) performance was only significantly different (lower) than: (i) AL2, Blocks 1-5, (ii) AL3, Blocks 1-5, and (iii) AL4, Blocks 3-5. Fourth, AL1 and AL4 were the only assistance levels to demonstrate significant improvement between the beginning and end of the experiment, as AL1, Block 5 and AL4, Block 5 produced significantly greater scores than AL1, Blocks 1-2 and AL4, Blocks 1-2, respectively.

It is noteworthy that AL4 performance is initially lower than AL2 and AL3, but by the end of the experiment (Block 5) produces the highest performance. This would suggest AL4 requires additional training to reach optimal performance, and/or it improves resistance to fatigue. A follow-on three factor ANOVA *excluding AL1* (main factors of assistance level, block number, and set) was employed to gauge significant differences between the last 3 assistance levels, as well as compare "experienced" (Block 5) performance. Main effects were *not* obtained for assistance level [$F_{(2,1077)} = 1.2$, $p = 0.3013$], and a multiple comparisons test did not show that AL4, Block 5 was significantly greater than AL2, Block 5 or AL3, Block 5. Therefore, no conclusions about AL4's superiority could be drawn.

V. Discussion

The significant improvement between AL1 and AL2 supports the argument that providing confidence information in a naturalistic manner (in the form of a particle filter heatmap) improves the abilities of the human operator to track and successfully engage the target. The manner in which the particle filter propagates uncertainty is simple to understand and is robust to noisy sensor measurements. We note that a potential confound affecting the performance divide between AL1 and AL2 was the difference in information provided about target velocity: while the particle filter (present in AL2/AL3/AL4) specifically modeled the target as having a velocity range between 7 and 12 m/s (and randomly generated particles accordingly), no information about this target velocity range was provided to the subject prior to the experiment, and therefore the subject would have needed to develop an internal model to infer target velocity when operating with AL1. The general rise in AL1 performance as block number increased may be due in part to learning that target velocity from prior blocks using AL2/AL3/AL4. However, as Figure 5 shows, the difference between AL1 and AL2 in later blocks remained substantial, indicating that the main variable differentiating these assistance levels—the visualization of uncertainty—was also likely an important factor.

The increase in performance from AL2 to AL3 was not shown to be significant; this may be due to the semi-static nature of the risk structure in the scenario. The location of hazards was easy to memorize prior to the start of the trial, and estimate on the belief map. Despite the additional information contained in the Risk Map in AL3 during the trial, eye tracking data revealed that subjects performing AL3 focused almost exclusively on the Belief Map, which is identical to that in AL2. Thus, since most subjects did not utilize the dynamic information in the Risk Map, AL3 provided no extra information over AL2.

In its current implementation, AL4 did not provide any additional assistance and actually caused a minor decrease in average performance (when compared over all blocks). However, Figure 5 and the previous ANOVA results show an increase in performance with increasing block number which suggests that with extended training/experience performing the task, AL4 may lead to improved performance. Future experiments where subjects are given more time to train would be required to confirm this conjecture. Alternatively, a redesigned AL4, that addresses the following issues, may also provide better performance: (i) the recommendation indicator (red circle) is present even when the minimum risk is high; due to the nature of the particle filter, its movement can be erratic and distracting; (ii) the recommendation meter is located at the bottom of the screen, which may require the operator to give up focus on the map to read it. Given these observations, a redesign of this assistance level may be warranted.

Future iterations of this C2 scenario/interface will consider (i) more complex environments with multiple hostile and non-hostile contacts, dynamic hazards, and "soft" information given to the human that the automation is not aware of, and (ii) creating a dialogue between the human and automation so that confidence/consequence information can be shared bidirectionally to improve automated recommendations.

VI. Conclusion

A human-subjects experiment was performed to evaluate a novel approach to decision support in a command and control scenario. The inherent uncertainty/ambiguity in command and control operations implies that the human operator should benefit from a concise, efficient representation of situational uncertainty. This comes in the form of confidence-based representation of filtered sensor data, and consequence-based integration of filtered data with known costs. In our target tracking and interdiction simulations, confidence information was manifested as a particle-filter-based heatmap of target location. Furthermore, the space of potential risk in such scenarios may be sufficiently complex that the operator may benefit from a visualization of risk. The effects of communicating confidence and consequence were experimentally evaluated by comparing performance in human subjects experiments. Confidence visualizations (AL2) improved results significantly over visualizations of baseline sensor information. However, supplementing confidence with consequence (AL3) and automated recommendations (AL4) were not shown to improve performance to a significant degree over confidence alone. Some evidence suggests that prolonged experience on the task may be required for consequence and automated recommendations to improve performance. Future work will explore whether increased training time or modifying the manner in which consequence or automated recommendations are presented may lead to enhanced performance.

Acknowledgment

This effort was funded through the Office of the Secretary of Defense Autonomy Research Pilot Initiative (ARPI), MIPR DWAM31168.

References

[1] Dept. of Defense, "Autonomy in weapon systems (dod directive 3000.09)," 2012, Nov. 21.

[2] M. L. Platt and S. A. Huettel, "Risky business: the neuroeconomics of decision making under uncertainty," *Nature neuroscience*, vol. 11, no. 4, pp. 398–403, 2008.

[3] P. M. Asaro, "The labor of surveillance and bureaucratized killing: New subjectivities of military drone operators," *Social semiotics*, vol. 23, no. 2, pp. 196–224, 2013.

[4] J. Kerstholt, "Time pressure effects on decision making in a dynamic task environment (het effect van tijdsdruk op beslissingsgedrag in een dynamische taaksituatie)," DTIC Document, Tech. Rep., 1991.

[5] Y. Harrison and J. A. Horne, "The impact of sleep deprivation on decision making: a review." *Journal of Experimental Psychology: Applied*, vol. 6, no. 3, p. 236, 2000.

[6] L. A. Breslow, D. Gartenberg, J. M. McCurry, and J. G. Trafton, "Dynamic operator overload: A model for predicting workload during supervisory control," *Human-Machine Systems, IEEE Transactions on*, vol. 44, no. 1, pp. 30–40, 2014.

[7] W. L. Chappelle, K. McDonald, and R. E. King, "Psychological attributes critical to the performance of mq-1 predator and mq-9 reaper us air force sensor operators," DTIC Document, Tech. Rep., 2010.

[8] J. A. Ouma, W. L. Chappelle, and A. Salinas, "Facets of occupational burnout among us air force active duty and national guard/reserve mq-1 predator and mq-9 reaper operators," DTIC Document, Tech. Rep., 2011.

[9] P. Slovic, B. Fischhoff, and S. Lichtenstein, "Behavioral decision theory," *Annual review of psychology*, vol. 28, no. 1, pp. 1–39, 1977.

[10] H. J. Einhorn and R. M. Hogarth, "Behavioral decision theory: Processes of judgment and choice," *Journal of Accounting Research*, pp. 1–31, 1981.

[11] G. A. Klein, "A recognition-primed decision (rpd) model of rapid decision making," in *Decision making in action: Models and methods*. Ablex Publishing Corporation, 1993, pp. 138–147.

[12] K. G. Ross, G. A. Klein, P. Thunholm, J. F. Schmitt, and H. C. Baxter, "The recognition-primed decision model," DTIC Document, Tech. Rep., 2004.

[13] G. Klein, "Naturalistic decision making," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50, no. 3, pp. 456–460, 2008.

[14] S. G. Hutchins, *Principles for intelligent decision aiding*. Springer, 1996.

[15] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.

[16] G. Shafer *et al.*, *A mathematical theory of evidence*. Princeton university press Princeton, 1976, vol. 1.

[17] C. Gonzalez, "Decision support for real-time, dynamic decision-making tasks," *Organizational Behavior and Human Decision Processes*, vol. 96, no. 2, pp. 142–154, 2005.

[18] A. R. Marathe, B. J. Lance, K. McDowell, W. D. Nothwang, and J. S. Metcalfe, "Confidence metrics improve human-autonomy integration," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 2014, pp. 240–241.

[19] L. Onnasch, C. D. Wickens, H. Li, and D. Manzey, "Human performance consequences of stages and levels of automation an integrated meta-analysis," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, p. 0018720813501549, 2013.

[20] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 30, no. 3, pp. 286–297, 2000.

[21] E. A. Doucette, "The application of a particle filter to urban ground target localization, tracking, and intercept," Ph.D. dissertation, Auburn University, 2012.

[22] SensoMotoric Instruments. Red250/red500. [Online]. Available: http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/red250-red-500.html

[23] IMOTIONS Products. Affectiva gsr qsensor. [Online]. Available: https://imotions.com/portfolio-item/affectiva-gsr-qsensor/

[24] F.-X. Lesage, S. Berjot, and F. Deschamps, "Clinical stress assessment using a visual analogue scale," *Occupational medicine*, p. kqs140, 2012.

[25] K. H. Kaplan, D. L. Goldenberg, and M. Galvin-Nadeau, "The impact of a meditation-based stress reduction program on fibromyalgia," *General hospital psychiatry*, vol. 15, no. 5, pp. 284–289, 1993.

[26] T. Ritvanen, V. Louhevaara, P. Helin, T. Halonen, and O. Hänninen, "Effect of aerobic fitness on the physiological stress responses at work," *International journal of occupational medicine and environmental health*, vol. 20, no. 1, pp. 1–8, 2007.

[27] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 2579-2605, p. 85, 2008.