EFFECTS OF TIME PRESSURE ON THE USE OF AN AUTOMATED DECISION SUPPORT SYSTEM FOR STRIKE PLANNING

Yves Boussemart, Birsen Donmez, M. L. Cummings, Jonathan Las Fargeas
Humans and Automation Laboratory
Massachusetts Institute of Technology, Cambridge MA

This paper describes the results of an experiment designed to examine the effects of time pressure on behavioral patterns. The main research hypothesis is that people under time pressure tend to increasingly rely on automation in order to cope with the added workload. The context is that of a missile strike planner having to create a set of matches between resources (missiles) and requirements (missions). We introduce time pressure by changing the temporal requirements towards the end of the mission. Overall performance, calls to automation and qualitative strategies are recorded and analyzed using ANOVA and other non-parametric tests. The main finding of this study is that while the number of calls to the automation did significantly increase under time pressure, there did not seem to be a statistically significant shift in problem solving strategies under time pressure. The experimental results show the importance of good automation-human interface design so as to gain maximum benefit from the use of an automated decision support systems.

Introduction

Missile strike planning is a complex example of multivariate optimization, where a set of resources must be paired with a set of goals in a manner that meets constraints and achieves a certain level of quality. In terms of strike planning for Tomahawk Land Attack Missile (TLAM), our representative domain, the missiles and associated missions are characterized by a series of variables. A strike planner's task consists of making sure that no hard constraint on these variables is violated when a specific missile is assigned to a specific mission. In addition, the final solution, that is the set of mission/missile assignments, should optimize soft constraints: it should be as "good" as possible along potentially subjective or dynamic references that may change in the planning process. In addition to such constraints, strike planners usually have to operate under temporal pressure and have a limited amount of time to finalize a strike plan.

Payne defines time pressure as "changing the time available to make a decision" [1]. The effects of time pressure on decision making have been described in the literature extensively and so have the resulting operator coping processes used. The most frequently cited coping processes for dealing with temporal stress are acceleration, filtering and omission [2]. Acceleration is probably the most obvious effect of time pressure and denotes an increased information processing rate. It has been shown, however, that with an increasingly stringent deadline subjects were less likely to rely uniquely on acceleration [3]. Filtering refers to processing some parts of the information more than others; the research has consistently shown that the attributes seen as less important tend to be filtered out first [4, 5]. Omission, also referred to as "shallower search for information" [6], implies ignoring particular parts of the information. In contrast to these coping processes, research has also shown that a common cognitive strategy shift is a tendency to lock into one problem solving strategy under time pressure even if it is suboptimal, a process also known as regression to learnt behaviors [7].

Previous work [8, 9] investigated the creation of decision-support tools aimed at leveraging human-automation collaboration to enhance the quality of the strike mission planning process. The current experiment builds on this previous work by adding temporal constraints to mission planning in order to examine the effects of time pressure on the use of automation during the strike planning process. The main research hypothesis we address is that people under temporal stress will rely more on automated tools in order to cope with the added workload. Time pressure is central to the context of Command and Control (C2) since theaters of operations are inherently dynamic; the conjunction of changes and fixed deadlines tend to put operators under considerable stress due to the time-critical nature and the importance of the decisions they have to make.

Method

Apparatus: StrikeView
StrikeView (Figure 1) is an interface designed to facilitate the process of planning strikes by decreasing the overall workload and improving the quality of the strike [8]. This interface allows the operator to solve the problem, i.e., build a set of mission/missile assignments, either manually or with the help of the computer.

The matching task consists of pairing a set of pre-planned missions with missiles available on different ships. This constitutes a complex, multivariate resource allocation problem, where a human operator must not only satisfy a set of matching

constraints, usually part of the rules of engagement (ROE), but also optimize the mission-missile assignments to minimize operational costs or enhance the quality of the overall plan. Generally it is left to the strike planner to manually assign missiles to missions, taking into account the different mission and missiles characteristics, as well as the constraints *du jour* included in the ROE. Given the scope of this experiment, we consider two hard-constraints based on the features of the missiles: navigation equipment (GPS[1], DSMAC[2] or both) and warhead type (penetrating, unitary, submunition). We also consider three so- called "soft-constraints": mission priority (low, medium, high), firing rate (probability of hitting a target) and days to port (number of days until the ship is due back to the harbor). The automated decision support provides the user with a heuristic-based computer-generated solution that only takes into account hard constraints along with a limited set of additional criteria. The solution provided usually is not optimal, but always exhibit correctness with respect to hard constraints. Finally, a time bar gives subjects a visual indication of how much time they have left to generate their solution. There also is a message box where information from Central Command can be relayed.
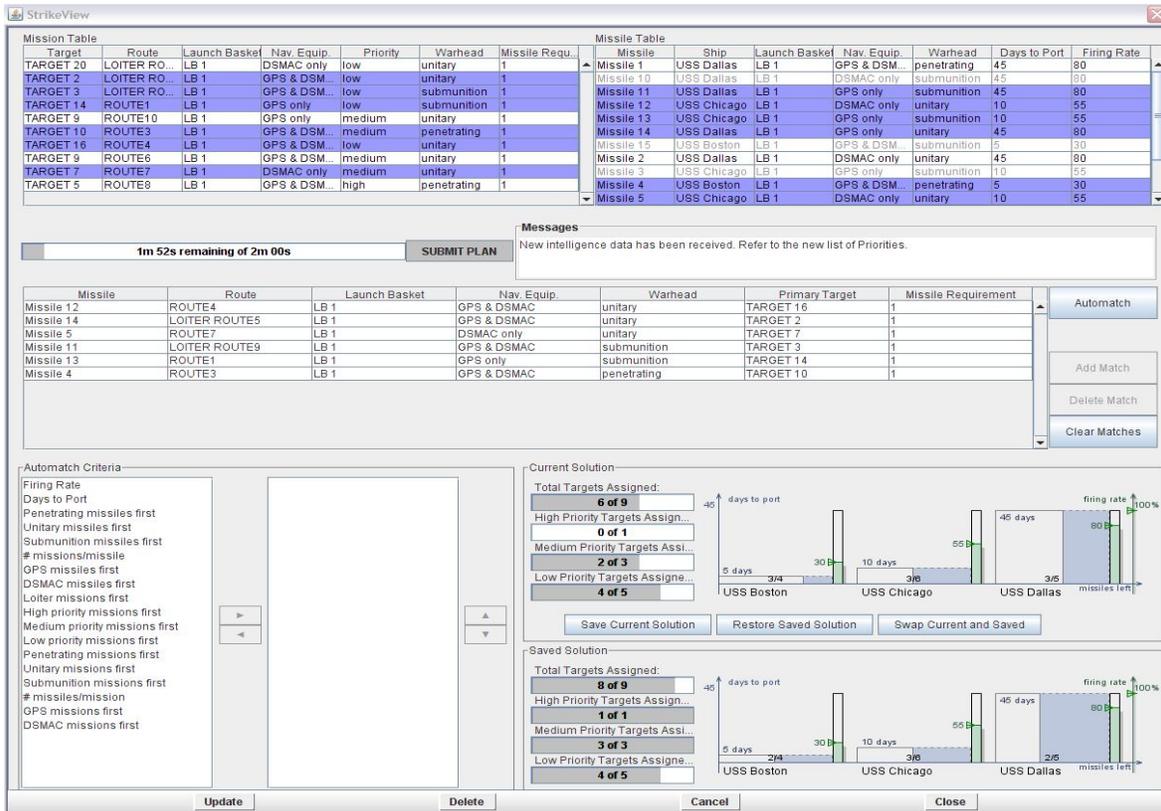

Figure 1. StrikeView Interface

Experimental Design
The experiment was a 2x2 mixed factorial design with time pressure (Low Time Pressure – LTP, High Time Pressure - HTP) as a within subjects variable and the order of presentation as a between subjects variable (LTP first, HTP first). The order of presentation was counterbalanced and randomly assigned to subjects.

Participants
18 participants were recruited, mostly from the MIT student population. Due to software glitches, the data obtained from two participants had to be dropped, therefore data from 16 participants were analyzed. In these 16, the male/female split was 11/5, 9 started with the LTP scenario and 7 with the HTP scenario. Finally, 8 were undergraduates and 8 were postgraduates (either graduate students of professionals). Each participant was paid $10 for the hour-long experiment, with a prospect of earning

---

[1] GPS: Global Positioning System

[2] DSMAC: Digital Scene-Mapping Area Correlator, is a high resolution satellite radar image of the target area which the Tomahawk follows to within feet of the intended target

an additional $60 gift certificate awarded to the best performer on the task. This monetary incentive was used to promote participants' involvement along with a drive to achieve an optimal solution.

Experimental Scenarios
In the LTP scenario, participants were given five minutes to complete the matching task, a duration that was determined to be comfortable during pilot studies. The HTP scenario started just like the LTP scenario with a five minute deadline. However, three and a half minutes into the experiment, the participants received new orders from Central Command to invert the priorities of the missions (that is low priority missions should now be regarded as high priority and vice versa). The participants had to re-plan the strike in the remaining one and a half minutes: this corresponds to the increased time pressure phase of the experiment. For both scenarios, the number of missions were greater than the number of missiles available, thus, it was not possible to assign a missile to every single mission.

Experimental Procedure
After signing consent forms, the participants were given a quick overview of the interface and experimental procedure via a slide-based presentation. A short training phase with the real interface and a mockup scenario was then provided in order to allow users to familiarize themselves with the task. There was no specific time limit on this hands-on training, and participants were asked to tell the experimenter when they felt they had achieved sufficient level of proficiency with the experimental setup. During the subsequent testing phase, each participant was presented with LTP and HTP scenarios in a random order. Before starting each scenario, the participants were given an identical set of ROE, which stated that the missions should be treated in the "normal" order of priority (high, normal, then low), and that they should try to maximize the firing rate attribute while minimizing the days to port attribute. In the HTP case, these priorities were reversed three and a half minutes into the scenario. After completing the two scenarios, the subjects were debriefed orally while a screen capture of their behavior was replayed. Questions were specifically asked to determine what type of strategy was used to solve the problem, how they reacted to the change of ROE and if they felt that time pressure affected their decision making process.

The data was recorded through built-in non-intrusive logging of the user interactions. The data consisted of mouse clicks, hovers and other interactions with UI features. We used TRACS2.5D to record participants' behavioral patterns. TRACS2.5D takes each triplet of successive mouse actions which was then fed through a parser which determines what category of action and what level of information detail was involved [9, 10]. Finally, as a failsafe, all trials were recorded using screen capture software.

Dependent Variables
The first dependent variable is a performance metric based both on the percentage of missions covered by a missile and on the optimization of the soft constraints (eq. 1).

$$perf = \sum_{all\ priorities} \Pr\left(\frac{1}{2}pm + \frac{1}{2n}\left(\frac{FR - \frac{DTP}{5*DTP_{max}}}{100} \right) \right)$$

$\Pr(H, M, L) = \{60, 30, 10\}: priority\ factor$

$pm \in [0, 100]: percentage\ of\ matches\ per\ priority$ Eq.1.

$n: number\ of\ missiles\ per\ priority$

$FR: Firing\ Rate$

$DTP: Days\ to\ Port$

$DTP_{max}: Max.\ Days\ to\ Port$

$0 < perf < 100$

The second dependent variable is the number of calls to the automated help. Finally, the last dependent variable is aimed at providing a finer grained analysis of the participant's behavior by examining specific sequences, or chains, of user events. This provides information regarding the succession of actions that are most likely to be undertaken by the strike planner. The specific features of sequences of interest were determined by using the TRACS2.5D tool [9, 10] and manually noting the most strongly-recurring chains of events. Three recurring chains were identified as strongly recurring using this method. Following the TRACS2.5D nomenclature, the patterns of interest were: (1) browse, evaluate, select, (2) browse, select, create a match, and (3) select criterion, call automatch, evaluate match. These patterns covered on average about 85% of all interactions.

Results

Number of Calls to Automation

Figure 2 shows the observation frequency for different number of automation calls. Number of automation calls greater than or equal to two are grouped under one category as there were a few observations that were greater than two. The figure shows a general trend of increased use of automation for the high time pressure condition. An ordered logit model, specifically proportional odds, was developed to compare the level of automation calls for the two different time pressure conditions adjusted for order of presentation, and order – level of time pressure interaction. A proportional odds model takes into account the ordinality of the data [11], in this case the three bins for the number of automation calls. Repeated measures were accounted for by creating a population-average model. Because the data consists of repeated measures, generalized estimating equations (GEE) was used for estimation.
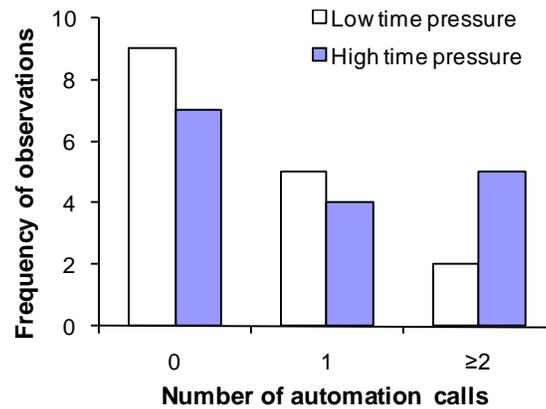

Figure 2. Observation Frequencies for Different Number of Automation Calls

Wald statistics for type GEE analysis revealed that time pressure ($\chi^2(1)=6.26$, $p=.01$) was statistically significant. Order, and order – level of time pressure interaction were not significant ($p>.05$), and hence were dropped from the model. High time pressure had 2 times higher odds of automation call than low time pressure (95% CI: 1.16, 3.42).

Performance

(a)   (b)
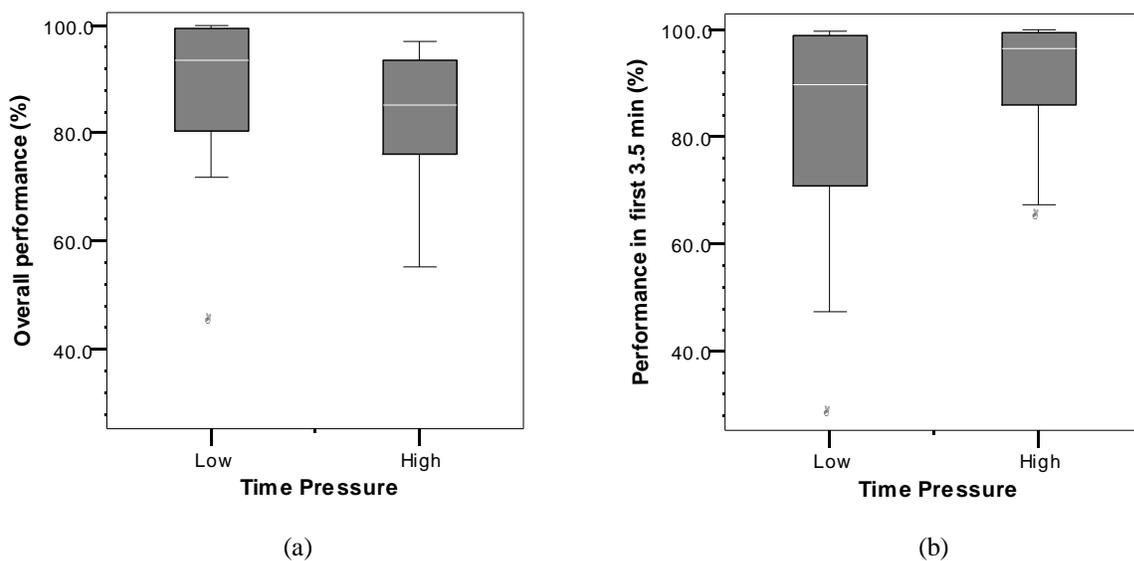
Figure 3. Performance for Different Time Pressure Conditions

A repeated measures ANOVA was conducted on overall performance (Figure 3a). Time pressure, order, and their interaction were not significant ($p>.05$). This suggests that the time pressure may not have been severe enough to affect the overall performance, or the increased use of automation under higher time pressure may have compensated for the otherwise diminished performance.

Homogeneity of Scenario Difficulty
LTP and HTP scenario were designed to be as similar as possible in terms of difficulty. However, because the scenarios were not precisely identical, we wanted to ensure that the previous results were due to the difference in time pressure and not to uncontrolled variation in scenario difficulty. In order to show that the increased odds of automation use under the high time pressure condition was in fact due to the dynamic taskload increase rather than the different Strike View tasks performed in the two conditions, number of automation calls and performance in the first three and a half minute portion of each condition was analyzed (Figure 3b). No difference was expected since the first three and a half minute of each condition had similar taskload. The results showed a significant order effect ($F(1,28) = 4.74$, $p = .04$), but non-significant time pressure effect. The interaction between time pressure and order interaction was also not significant ($p>.05$). The number of automation calls (zero, one, or more) were analyzed with an ordered logit model. As expected, time pressure, and time pressure-order interaction were not significant ($p>.05$). These results suggest that the significant increase in automation use reported in the previous section is indeed due to the higher time pressure induced as opposed to variation in scenario difficulty.

Overall Strategy Switch
The question of whether time pressure would have impacts other than usage of automation in the participant's behavior was approached by measuring the presence of three strongly recurrent patterns in the user's behavior. Each pattern can be seen as a chain of three sub-events and can be scored by using pattern matching algorithms that count the number of time a specific pattern appears in the data. As discussed previously, the three patterns of interest were: (1) browse, evaluate, select, (2) browse, select, create a match, and (3) select criterion, call automatch, evaluate match. The analysis of the scores for the different pattern did not show any significant difference between the different scenarios or between the different phases of each scenario.

Discussion

The results of this experiment show that time pressure did lead to an increase use of automated help and thus verify our main research question; however, temporal stress did not seem to produce any significant differences in the way the participants chained their actions to solve the problem. The results obtained highlight the difficulty of measuring the impact of time pressure on human behavior because time pressure does not always lead to measurable changes in cognitive strategies: coping processes can balance out the effects of time pressure and maintain the same output [2].

Learning Effect and Training Issues
Performance measures revealed that there were significant order effects, with the second scenario yielding a higher performance. This trend suggests that the training may not have been sufficient to get participants at a reasonably stable level of proficiency, and that the first scenario might have had effects akin to an additional training session. The lack of shifts in strategies between phases of the low time pressure scenario could be a possible consequence of the unsatisfactory level of proficiency achieved by the participants with the interface. At the end of the experiment, multiple participants reported that they realized they should have used the automation after the change of ROE, but that they had been under too much pressure to think straight and actually implement what they recognized to be, a posteriori, the best solution. Some participants were clearly overwhelmed by the additional workload and the time stress engendered by the change in ROE. It is likely that had the experiment been repeated, the subjects would have been more ready to respond to a change in operational parameters simply because they would have had seen one already. It is also worth mentioning that Navy strike planners are trained officers and might therefore exhibit a different type of behavior.

Still, it is difficult to get the right balance between training the user to use an interface and biasing them by overemphasizing a given strategy. Over constraining the user into a pre-determined behavior does not help make an assessment of the cognitive strategies.

(Dis)Trust in Automation and Satisficing
Trust in automation [12, 13] is a vast topic that is barely touched in this experiment. During the training it was specifically mentioned that the automation feature would provide a correct, albeit likely sub-optimal, solution. The MIT population tends to be biased towards technically-inclined and detail-oriented personalities. As a consequence, we had some participants who

refused to use the automation because they didn't like to use an algorithm they were not familiar with and that was described as suboptimal. This behavior was usually linked to the feeling that they could do better by using a fully manual strategy. In essence, this means that such participants were really trying to optimize the solution as much as possible, and would not settle for sub-optimality. Although not addressed by the data analysis, experimental observations tend to suggest that such people were the most affected by the change of ROE because they had to change their optimal search model, and were not trying to rely on simplifying heuristics. On a related note, the majority of people in this situation mentioned that they did not see it worth the effort to get familiar with the automated feature. There was thus a clear cost-benefit analysis that was made regarding the effort needed to understand the automation and the potential advantages it could bring. This conclusion could therefore have impact on the procedure used to train the strike planners.

At the other extreme of the spectrum, one of our participants was a very experienced US Air Force officer used to designing flight plans with the aid of a computer. His experience and training had taught him to trust the automation, and, according to the subject, even though the solution wasn't perfect it was considered to be "good enough". Such users were still trying to optimize the solution based on the automated one. However, because they usually weren't very familiar with the data, such optimization efforts were usually limited. Conversely, their strategy changed little under time pressure since they were able to leverage the automation to instantly create a plan that they accepted as being good enough. The best performance observed on the high time pressure scenario was actually from a participant who explicitly mentioned that he trusted the algorithm and thought it was good enough. After receiving the new ROE, this participant made use of the automation and changed the assignment of a pair of missile, thereby gaining an edge and outperforming all the other participants.

Conclusions

Experimental results supported our main research hypothesis, namely that, under time pressure, subjects tended to use more automation than in a baseline, low temporal stress, situation. Conversely, the experiment did not exhibit statistically significant changes of cognitive strategies between the two conditions. Based on post-experiment verbal reports, however, this result might be attributed to an insufficient amount of training. Further studies should be performed in order to satisfactorily answer this question. Still, the overall conclusions of this experiment highlight the need for a thorough understanding of the nature of the human-automation collaboration, especially in contexts such as command and control where time-critical decisions must be taken contingent on dynamic environments.

REFERENCES

[1]   J. W. Payne, J. R. Bettman, and E. J. Johnson, *The Adaptive Decision Maker*: Cambridge University Press, 1993.
[2]   A. Edland, "Attractiveness judgments of decision alternatives under time stress," *Cognition and Decision Research Unit,* vol. Report No. 21, 1985.
[3]   A. J. Maule and G. R. J. Hockey, "State, Stress and Time Pressure," in *Time Pressure and Stress in Human Judgment and Decision Making*, O. S. a. A. J. Maule, Ed. New York: Plenum Press, 1993.
[4]   P. Wright, "The harassed decision maker: Time pressures, distractions and the use of evidence," *Journal of Applied Psychology,* vol. 59, pp. 555-561, 1974.
[5]   O. Svenson, Edland, A., & Slovic, P., "Choices between incompletely described alternatives under time stress," *Acta Psychologica,* vol. 75, pp. 153-169, 1990.
[6]   I. L. M. Janis, L., *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: The Free Press, 1977.
[7]   A. S. Luchins, "Mechanizations in problem solving: The effect of Einstellung," *Psychological Monographs,* vol. 54, 1942.
[8]   S. Bruni and M. Cummings, "Human Interaction with Mission Planning Search Algorithms," in *Human Systems Integration Symposium*, Arlington, VA, 2005.
[9]   S. Bruni and M. L. Cummings, "Tracking Resources Allocation Cognitive Strategies for Strike Planning," in *COGIS*, 2006.
[10]   S. Bruni, J. Marquez, A. Brzezinski, and M. L. Cummings, "Visualizing Operators' Cognitive Strategies In Multivariate Optimization," in *50th Annual Meeting of the Human Factors and Ergonomic Society*, San Francisco, CA, 2006.
[11]   M. E. Stokes, C. S. Davis, and G. G. Koch, *Categorical data analysis with the SAS System*, 2nd ed. Cary, NC: SAS Institute Inc., 2000.
[12]   J. D. Lee and N. Moray, "Trust, Self-Confidence, and Operators' Adaptation to Automation," *International Journal of Human-Computer Studies,* vol. 40, pp. 153-184, 1994.
[13]   L. J. Skitka, K. L. Mosier, and M. D. Burdick, "Accountability and automation bias," *International Journal of Human-Computer Studies,* vol. 52, pp. 701-717, 2000.