



Measuring the Effects of a Cognitive Aid in Deep Space Network Operations

Edward Barraza^{1,2(✉)}, Alexandra Holloway¹, Krys Blackwood¹,
Michael J. Gutensohn³, and Kim-Phuong L. Vu²

¹ The Jet Propulsion Laboratory, Pasadena, CA, USA
edward.barraza@jpl.nasa.gov

² California State University, Long Beach, Long Beach, CA, USA
edward.barraza@student.csulb.edu

³ Rollins College, Winter Park, FL, USA

Abstract. Cognitive aids have long been used by industries such as aviation, nuclear, and healthcare to support operator performance during nominal and off-nominal events. The aim of these aids is to support decision-making by providing users with critical information and procedures in complex environments. The Jet Propulsion Lab (JPL) is exploring the concept of a cognitive aid for future Deep Space Network (DSN) operations to help manage operator workload and increase efficiency. The current study examines the effects of a cognitive aid on expert and novice operators in a simulated DSN environment. We found that task completion times were significantly lower when cognitive aid assistance was available compared to when it was not. Furthermore, results indicate numerical trends that distinguish experts from novices in their system interactions and efficiency. Compared to expert participants, novice operators, on average, had higher acceptance ratings for a DSN cognitive aid, and showed greater agreement in ratings as a group. Lastly, participant feedback identified the need for the development of a reliable, robust, and transparent system.

Keywords: Cognitive aid · Decision support · Technology acceptance
Expertise · Deep Space Network

1 Introduction

Cognitive aids have long been used by industries such as aviation, nuclear, and healthcare to support operator performance during nominal and off-nominal events. The concept of cognitive aids includes checklists, flowcharts, sensory cues, safety systems, decision support systems, and alerts (Levine et al. 2013; Singh 1998). Some examples of cognitive aids are the Quick Reference Handbook for pilots, the National Playbook for air traffic controllers (ATCos), and the Surgical Safety Checklist for surgeons (Playbook 2017; Catalano 2009). They are all designed to guide decision-making by providing users with important information in complex environments. The Jet Propulsion Lab (JPL) is exploring the concept of a cognitive aid for Deep Space Network (DSN) operations.

1.1 The Deep Space Network

The DSN is a global network of telecommunications equipment that provide support for interplanetary missions of the National Aeronautics and Space Administration (NASA) and other space agencies around the world (DSN Functions n.d. 2017). Over the next few decades, the network will experience a significant surge in activity due to an increase in the total number of antennas, an increase in missions, higher data rates, and more complex procedures (Choi et al. 2016). To meet the projected demand, JPL is managing a project called Follow-the-Sun Operations (FtSO). A number of automation improvements designed to help manage the workload and efficiency of operators are in development. One of these improvements is the application of complex event processing (Johnston et al. 2015).

Broadly, complex event processing is a machine learning method of combining data streams from different sources in order to identify meaningful events and patterns (Choi et al. 2016). One complex event processing application is the detection of operational deviations from the norm. The system matches ongoing situations in real-time to previous incidents and then provides procedural resolution advisories. The information output of CEP is intended to guide operator decision making. In doing so, the information aids operator cognition. CEP will be just one of an array of tools that are available for LCOs to perform their job.

1.2 Cognitive Aids

A cognitive aid is a presentation of prompts aimed to encourage recall of information in order to increase the likelihood of desired behaviors, decisions, and outcomes (Fletcher and Bedwell 2014). Cognitive aids include, but are not limited to, checklists, flowcharts, posters, sensory cues, safety systems, alerts, and decision support tools (Levine et al. 2013; Singh 1998). A large body of research has demonstrated the concept is applicable anywhere users operate in stressful working conditions with high information flow and density.

Arriaga et al. (2013) evaluated an aid that assisted operating-room teams during surgical crisis scenarios in a simulated operating room. They were interested in seeing if a crisis checklist intervention would improve adherence to industry best practices. A total of 17 operating-room teams were randomly assigned to manage half the scenarios with a crisis checklist and the other half from memory alone. They found that the use of crisis checklists was associated with a significant improvement in adherence to recommended procedures for the most common intraoperative emergencies, such that 6% of steps were missed when checklists were available as opposed to 23% when they were unavailable. Thus, crisis checklists have the potential to improve surgical care.

There is evidence that decision support tools can reduce workload in addition to improving performance (Van de Merwe et al. 2012). Van de Merwe et al. (2012) evaluated the influence of a tool named Speed and Route Advisor (SARA) on ATCO performance and workload in an air traffic delivery task. In a simulation, the experimenters captured accuracy by measuring the adherence of aircraft to the expected approach time, the controllers' subjective workload through a self-assessment measure, and controllers' objective workload through the total number of radio calls and device

inputs. SARA provided speed and route advisories for every inbound flight, thereby providing controllers with the information to issue a single clearance to each aircraft to manage traffic delivery.

1.3 Current Study

We are not aware of any research that investigates operator performance in DSN operations. Further, there is no research that explores how a cognitive aid affects the workload of LCOs or provides a measure of their acceptance of an aid. The current study examines the impact of a cognitive aid on the performance and workload of expert and novice LCOs in a DSN task. We had the following research questions:

1. What effect will the cognitive aid have on operator performance in the DSN framework?
2. Will experts and novices perform differently using the cognitive aid?
3. What effect will the cognitive aid have on operator subjective workload in the DSN framework?
4. Will novice operators have higher acceptance ratings for the cognitive aid compared to experts?

Participants were asked to monitor, detect, and resolve any issues they encountered during simulated tracks. On some tracks, a cognitive aid detected issues and provided resolution advisories. We measured their performance, acceptance using the perceived usefulness scale of the TAM, and workload using a composite NASA Task Load Index (TLX) score. While the cognitive aid in DSN operations is still a concept, it is important to understand its utility and effects early in its development.

2 Method

2.1 Participants

Eight employees (3 females, 5 males) of the Deep Space Network were recruited for this experiment. To examine different levels of expertise, four participants were expert LCOs from the Goldstone Deep Space Communications Complex; the other four participants were DSN Track Support Specialists from the Space Flight Operations Facility at JPL. The Track Support Specialists were selected as proxy for novice LCOs because they are familiar with the terminology, displays, and data constituents. However, they do not perform the same tasks as LCOs. Participants had 9 months – 40 + years of experience working in the DSN. The expert group had an average of 30.5 years of DSN experience; novice group had an average of 2.75 years of experience. Participants were between 35–70 years of age ($M = 53.71$). The average age for the expert group was 58.75 years; average age for novice group was 46 years (one participant did not disclose). Participants were volunteers and were not compensated for their time.

2.2 Design

The simulation employed a 2 (Group: Experts/Novice) × 2 (Assistance: Cognitive aid/No Cognitive aid) mixed factorial design. Group was a between-subjects variable. Assistance was a within-subjects variable. The dependent measures include performance, subjective workload, and operator acceptance. Performance was assessed with the following variables: system interactions (clicks on the interface display) and efficiency (seconds). Subjective workload was assessed with the NASA TLX. The TAM Perceived Usefulness scale was distributed to assess participant acceptance of the cognitive aid.

2.3 Materials

Participants were tested using the Deep Space Network Track Simulator. The Deep Space Network Track Simulator is a medium fidelity simulator that reproduces the cruising phase of a track. This simulator ingests archival data and presents this information on up to three current DSN displays (see Fig. 1). Three displays were determined by a subject matter expert to be the minimum displays operators required to diagnose and resolve the four simulated issues. The displays are the Signal Flow Performance screen, Current Tracking Performance screen, and the Radio Frequency Signal Path screen. All screens were presented in a single display using a 27-inch Apple Thunderbolt monitor (see Fig. 2). The screen was recorded to capture the interactions enacted by participants. The simulator captured the total task times and tracked the displays that were clicked.

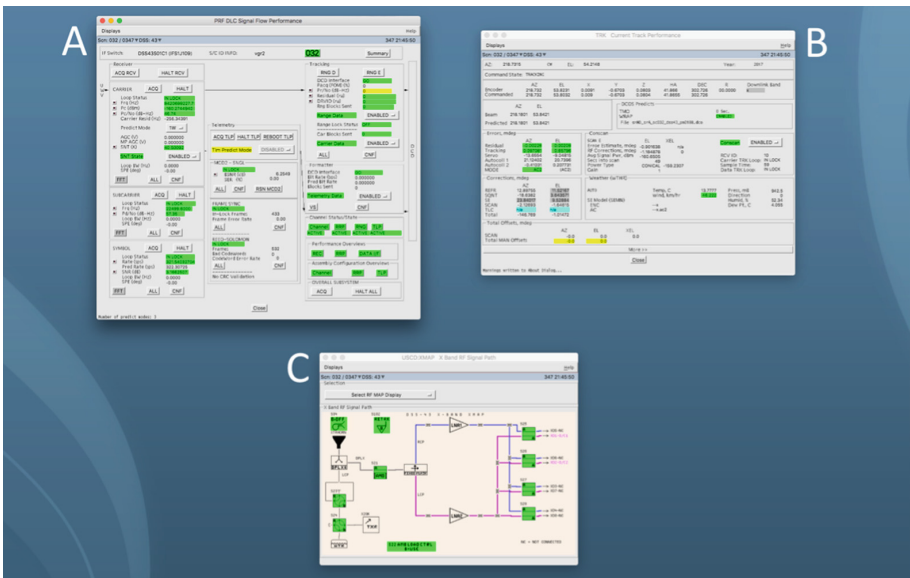


Fig. 1. Deep space network track simulator displaying the (A) signal flow performance display, (B) Current tracking performance display, and the (C) radio frequency signal path display.

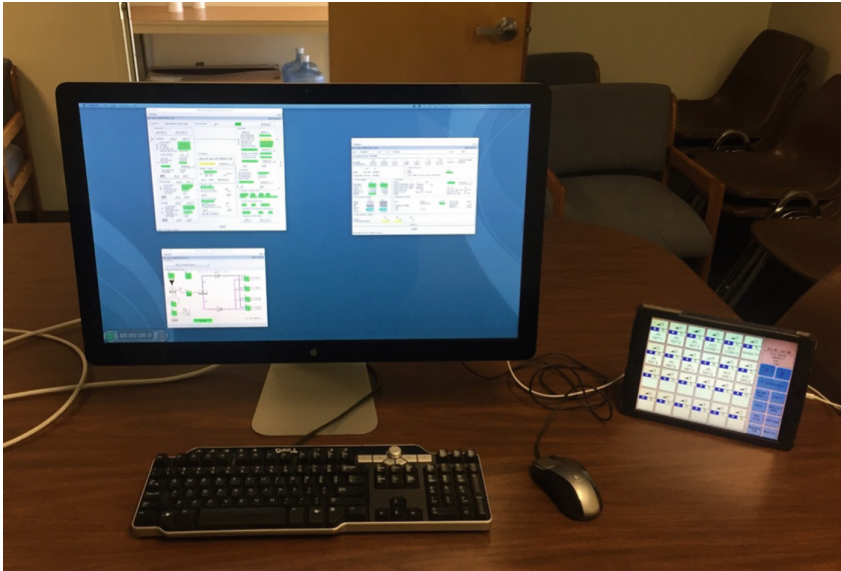


Fig. 2. The link control station setup.

There were four scenarios, each duplicating an issue that LCOs currently manage during operations. Each scenario was approximately 10 min long and with fault injection occurring approximately 3 min in to the track. Archival data was used to simulate a track. Scenarios were counterbalanced between participants.

Other materials used in the JPL study include an informed consent form, a demographics questionnaire, and the NASA TLX (Hart and Staveland 1988) and the Perceived Usefulness Scale (PU scale; Davis 1989). The NASA TLX is a subjective measure of workload that uses a 1 to 100 scale. The NASA TLX assesses workload across six different dimensions: mental demand, physical demand, temporal demand, effort, performance, and frustration. The PU scale uses a 1 to 7 scale that assess various dimensions of perceived usefulness of technology. Participants also filled out a post-questionnaire at the conclusion of experimental trials.

3 Procedure

One participant was run at a time. Upon arrival, participants were instructed to read and sign the informed consent form and fill out the demographics questionnaire. Participants were then briefed for 10 min on the study. They were briefed on the task and the displays that would be available to them.

After the briefing session, participants completed a training scenario which exposed them to the simulation environment through a training track. During this scenario, the participants were able to ask the simulation manager questions about the functionality of the simulator. After the conclusion of the training scenario, participants were asked if they had questions about the procedures.

The participants had two experimental blocks. The two blocks varied in assistance level. The two assistance levels were cognitive aid and no cognitive aid conditions. Participants first ran in one experimental block and then the other. The experimental blocks were counterbalanced across participants. Each experimental block was approximately 30 min long and consisted of two scenarios, each a maximum of 10 min long. A detailed breakdown of a scenario is described in the next section. At the end of each scenario, participants were required to fill out the NASA TLX. At the end of the second experimental block, participants were debriefed and thanked for their participation.

3.1 Cognitive Aid

The variable of Assistance was counterbalanced by either including or omitting a cognitive aid in the second experimental block. For half of the participants, the first two scenarios included no cognitive aid (i.e., participants had to diagnose and resolve issues themselves), and the second experimental block provided participants with cognitive aid assistance, where the tool diagnosed issues and provided resolution advisories three minutes in to the scenario (See Fig. 3). The opposite order was given to the other half of the participants. A total of two resolution advisories were presented in each cognitive aid recommendation, but only one advisory resolved the simulated issue. Clicking the “+” revealed step-by step procedures for the advisory. Whether the correct advisory was listed first or second was also counterbalanced between scenarios.

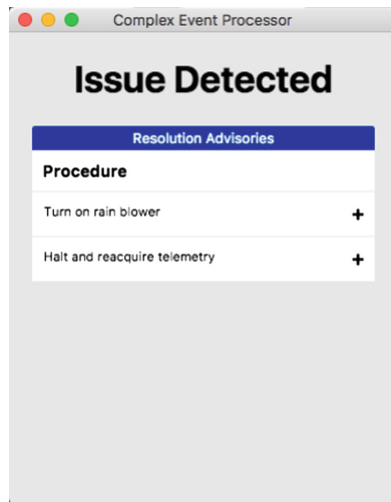


Fig. 3. The cognitive aid (CEP) prompt.

4 Results

Descriptive statistics are reported as well as the effects of separate 2 (Group: Expert/Novice) \times 2 (Assistance: Cognitive Aid/No Cognitive Aid) mixed ANOVAs, with group as the between-subjects variable. Although the sample size was small, due to the specialized qualifications of the participants, ANOVAs were conducted to help identify trends in the data. The dependent variables are system interactions, as indicated by the number of clicks participants performed during each trial, efficiency, as indicated by the total task time for each trial, and workload, as indicated by subjective NASA TLX ratings for each trial. Significance was set at $p < .05$, but we also report trends if $p < .10$.

4.1 System Interactions

System Interactions were captured by the number of clicks participants enacted to troubleshoot scenarios, after controlling for ambiguous clicks (i.e. clicks that had no strategic intent), and for cognitive aid clicks (i.e. clicks that were enacted on the cognitive aid display) (See Table 1). The 2 (Group: Expert/Novice) \times 2 (Assistance: Cognitive aid/No cognitive aid) mixed ANOVA yielded a non-significant main effect of group, $F(1, 6) = 5.56$, $MSE = 2.31$, $p = .06$, $\eta p^2 = .48$. Experts ($M = 3.94$, $SE = .57$) tended to click on their displays more than novice participants ($M = 2.15$, $SE = .54$). The effect of assistance was not significant, $F(1, 6) = .37$, $MSE = 4.77$, $p = .56$, $\eta p^2 = .06$. Participants clicked on their displays about three times, both in conditions with no cognitive aid, ($M = 3.38$, $SE = .75$), and in conditions with a cognitive aid, ($M = 2.71$, $SE = .57$) (See Table 1). No significant interaction was found, $F(1, 6) = .25$, $MSE = 4.77$, $p = .64$, $\eta p^2 = .04$ (See Fig. 4).

Table 1. Descriptive statistics from the 2 (Group: Expert/Novice) \times 2 (Assistance: Cognitive Aid/No Cognitive Aid) mixed ANOVA for system interactions

Assistance	Group	System interactions (clicks)
Cognitive Aid	Expert	$M = 3.88$ $SD = 1.79$
	Novice	$M = 1.54$ $SD = 1.41$
	Total	$M = 1.54$
No Cognitive Aid	Expert	$M = 4.00$ $SD = 1.78$
	Novice	$M = 2.75$ $SD = 2.39$
	Total	$M = 3.38$
Total	Expert	$M = 3.94$
	Novice	$M = 2.15$

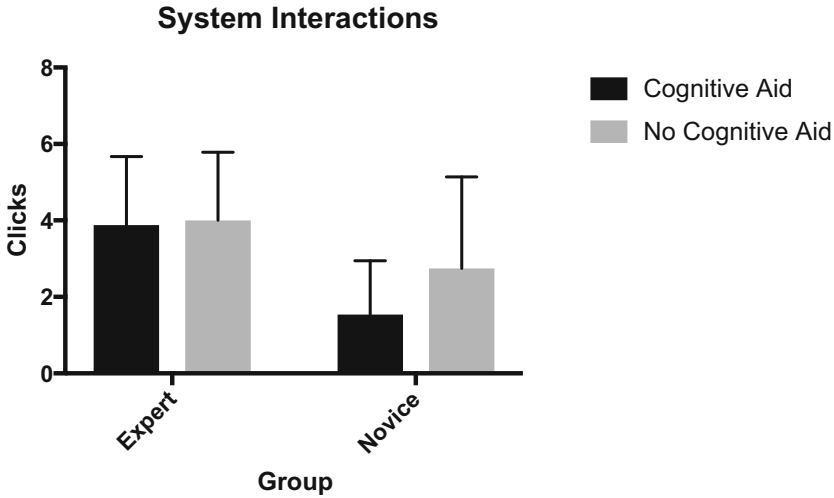


Fig. 4. Mean number of clicks by group and assistance for system interactions.

During cognitive aid trials, participants had the option of adhering to or ignoring the tool. We examined the number of times participants chose the correct resolution as their initial resolution strategy in cognitive aid conditions. Participants chose the correct resolution first in 9 (56%) of 16 cognitive aid trials. Of these 9, experts were responsible for 5 of these trails and novices for 4. All but one trial (performed by a novice) were eventually resolved.

4.2 Efficiency

Efficiency was captured as the total task time, in seconds, for each trial. Each scenario was approximately 600 s in length (10 min). Participants were required to resolve issues within the 600 s scenario, otherwise the simulator would timeout. Timeouts were examined for frequency of occurrence by group. Out of a total of 32 trials, eight (25%) timed out. Of these eight, three timeouts occurred in the expert group and five occurred in the novice group. One novice participant was responsible for three of five timeouts in the novice group. More specifically, seven timeouts occurred in trials where participants received no cognitive aid assistance and one timeout occurred in the novice group trial with cognitive aid assistance.

We performed separate 2 (Group: Expert/Novice) \times 2 (Assistance: Cognitive Aid/No Cognitive Aid) mixed ANOVAs to examine task time measures with and without timeouts, see Table 2 for means. For the analysis with timeout trials, the time on task was set as the limit of 600 s. There was a significant main effect of assistance, $F(1, 6) = 26.97$, $MSE = 4261.12$, $p < .01$, $\eta p^2 = .82$. Participants were faster at resolving issues when they had the assistance of a cognitive aid ($M = 281.06$, $SE = 32.9$) than when they did not ($M = 450.56$, $SE = 35.41$). The effect of group was not significant, $F(1, 6) = .71$, $MSE = 14425.49$, $p = .43$, $\eta p^2 = .12$. Experts were not significantly faster at resolving issues ($M = 340.44$, $SE = 42.46$) compared to novices

($M = 391.19$, $SE = 32.9$) (See Table 2). No significant interaction was found, $F(1, 6) = .03$, $MSE = 4261.12$, $p = .88$, $\eta p^2 = .00$ (See Fig. 5).

Table 2. Descriptive statistics from the 2 (Group: Expert/Novice) × 2 (Assistance: Cognitive Aid/No Cognitive Aid) mixed ANOVA for efficiency

Assistance	Group	Efficiency with timeouts (seconds)	Efficiency without timeouts (seconds)
Cognitive Aid	Expert	$M = 253.13$	$M = 253.13$
		$SD = 86.23$	$SD = 86.23$
	Novice	$M = 309.00$	$M = 263.83$
$SD = 99.39$		$SD = 50.76$	
	Total	$M = 281.06$	$M = 257.71$
No Cognitive Aid	Expert	$M = 427.75$	$M = 323.38$
		$SD = 82.65$	$SD = 99.13$
	Novice	$M = 473.38$	$M = 340.33$
$SD = 115.01$		$SD = 156.09$	
	Total	$M = 450.56$	$M = 330.64$
Total	Expert	$M = 340.44$	$M = 288.25$
	Novice	$M = 391.19$	$M = 302.08$

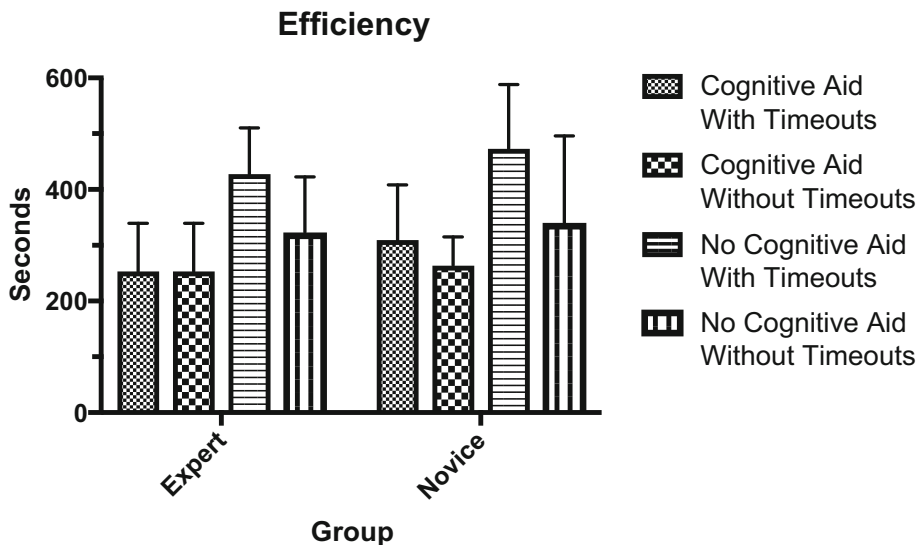


Fig. 5. Mean task completion time by group and assistance for efficiency, with and without timeouts.

A second 2 (Group: Expert/Novice) × 2 (Assistance: Cognitive Aid/No Cognitive Aid) mixed ANOVA was performed with the timeout data coded as missing values. For this analysis, there was no main effect of group, $F(1, 5) = .04$, $MSE = 15225.67$,

$p = .84$, $\eta p^2 = .01$. However, the numerical pattern remained the same as the previous analysis: experts were numerically faster ($M = 288.25$, $SE = 43.63$) at resolving issues compared to novices ($M = 302.09$, $SE = 50.38$). The effect of assistance was also not significant, $F(1, 5) = 3.13$, $MSE = 5907.78$, $p = .14$, $\eta p^2 = .39$. Participants were not significantly faster at resolving issues when they had the assistance of a cognitive aid ($M = 258.48$, $SE = 28.30$) than when they did not ($M = 331.85$, $SE = 47.76$), see Table 2. No significant interaction was found, $F(1, 5) = .006$, $MSE = 5907.78$, $p = .94$, $\eta p^2 = .00$ (See Fig. 5).

4.3 Subjective Workload

Subjective workload was captured by the NASA Task Load Index at the end of each trial. All the ratings for the seven dimensions were combined to provide one composite score for each participant. The higher the TLX rating, the higher the workload. A 2 (Group: Expert/Novice) \times 2 (Assistance: Cognitive aid/No Cognitive Aid) mixed ANOVA did not yield a significant main effect of group, $F(1, 6) = .01$, $MSE = 766.6$, $p = .93$, $\eta p^2 = .00$, nor assistance, $F(1, 6) = .27$, $MSE = 198.51$, $p = .62$, $\eta p^2 = .04$. Experts reported a mean workload rating of 41.26 and novices 39.93. Conditions with a cognitive aid were reported to have a mean workload of 38.77. Mean workload in conditions with no cognitive aid was 42.42 (See Table 3). No significant interaction was found, $F(1, 6) = .03$, $MSE = 198.51$, $p = .87$, $\eta p^2 = .01$ (See Fig. 6).

Table 3. Descriptive statistics from the 2 (Group: Expert/Novice) \times 2 (Assistance: Cognitive Aid/No Cognitive Aid) mixed ANOVA for subjective workload

Assistance	Group	Workload (NASA TLX)
Cognitive Aid	Expert	$M = 40.05$ $SD = 29.53$
	Novice	$M = 37.50$ $SD = 18.74$
	Total	$M = 38.77$
No Cognitive Aid	Expert	$M = 42.48$ $SD = 18.91$
	Novice	$M = 42.36$ $SD = 18.69$
	Total	$M = 42.42$
Total	Expert	$M = 41.26$
	Novice	$M = 39.93$

4.4 Technology Acceptance

Acceptance was measured by the Perceived Usefulness (PU) scale taken from the Technology Acceptance Model (Davis 1989). The acceptability items were all rated on a scale from 1 “Extremely Likely” to 7 “Extremely Unlikely” and a 4 as “Neither”. Lower numbers indicate a greater acceptance of technology. One-way ANOVAs with

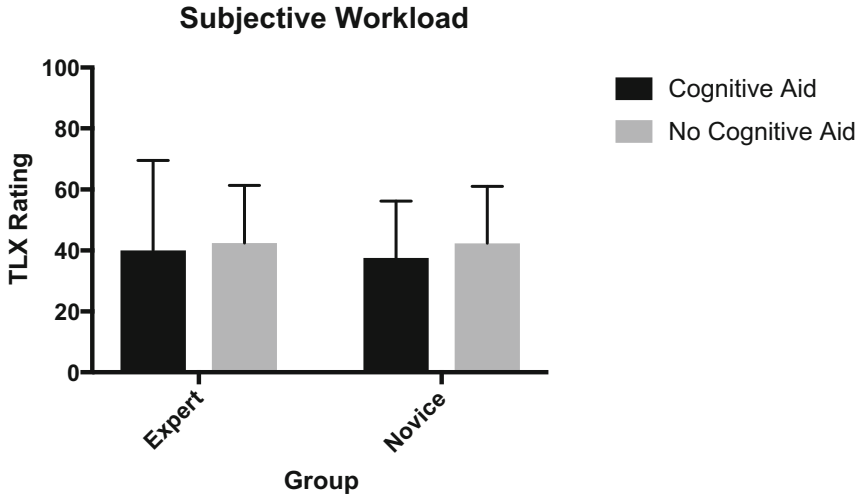


Fig. 6. Mean workload ratings by group and assistance for subjective workload.

(Group: Expert/Novice) as a between subject variable did not yield any significant effects of group, see Table 4 for means, F-ratios, and p-values. Overall, expert rated questions one through five slightly above a 4, indicating indifference to Complex Event Processing (CEP or cognitive aid). Question six was the only question that experts rated the CEP has more acceptable ($M = 2.75$) and with the lowest variability ($SD = .98$). Although novices rated all questions below a 4, indicating greater acceptance of a cognitive aid when compared to expert ratings, these differences were not significant.

Table 4. Descriptive statistics from the perceived usefulness scale (lower numbers indicate greater acceptance to technology). CEP refers to complex event processing (cognitive aid).

Technology acceptance scale: perceived usefulness	Expert	Novice	F-ratio; P-value
1. Using CEP in operations would enable me to accomplish tasks more quickly	$M = 4.25$ $SD = 2.75$	$M = 2$ $SD = 0$	$F(1, 6) = 2.67$ $p = .15$
2. Using CEP would improve my operational performance	$M = 4.25$ $SD = 2.75$	$M = 2.25$ $SD = .5$	$F(1, 6) = 2.04$ $p = .20$
3. Using CEP in operations would increase my productivity	$M = 4.5$ $SD = 2.38$	$M = 3$ $SD = 1.41$	$F(1, 6) = 1.17$ $p = .32$
4. Using CEP would enhance my effectiveness in operations.	$M = 4.25$ $SD = 2.75$	$M = 2.5$ $SD = 1$	$F(1, 6) = 1.42$ $p = .28$
5. Using CEP would make it easier to conduct operations	$M = 4.5$ $SD = 2.38$	$M = 2.5$ $SD = .58$	$F(1, 6) = 2.67$ $p = .15$
6. I would find CEP as a useful tool in operations	$M = 2.75$ $SD = .98$	$M = 2.75$ $SD = .98$	$F(1, 6) = .65$ $p = .45$

A frequency distribution of participant ratings for each question of the Perceived Usefulness scale revealed two patterns. First, expert participants tended to be split in their ratings such that half of expert participants clustered on one end of the scale and the other end of the scale. This pattern is consistent for all questions. On the other hand, novice participants tended to cluster together in their responses in agreement to accept CEP.

4.5 Post Questionnaire

Participants were asked to complete a six-question survey at the conclusion of the simulation. The first five questions used a five-point Likert scale from 1 “Strongly Disagree” to 5 “Strongly Agree” and a 3 as “Neither”. The sixth question was open ended. One-way ANOVAs with (Group: Expert/Novice) as a between subject variable did not yield a significant main effect for any questions, see Table 5 for means, F-ratios, and p-values. Both groups gave mostly neutral responses to four of five questions. The fifth question, which asked if a tool like CEP would be useful in Follow-the-Sun operations, had the highest agreement among the two groups ($M = 3.88$).

Table 5. Descriptive statistics from the post questionnaire (higher numbers indicate greater agreement). CEP refers to complex event processing (cognitive aid).

Post questionnaire	Expert	Novice	F-ratio; P-value
1. I trust the CEP system	$M = 2.5$	$M = 3.5$	$F(1, 6) = 1$
	$SD = 1.91$	$SD = .58$	$p = .36$
2. The CEP system lowered my workload	$M = 3$	$M = 3.25$	$F(1, 6) = .06$
	$SD = 1.83$	$SD = .96$	$p = .82$
3. The CEP system increased my situation awareness	$M = 2.5$	$M = 3.25$	$F(1, 6) = .87$
	$SD = 1.29$	$SD = .96$	$p = .39$
4. A tool like CEP will be useful in Follow-the-Sun operations	$M = 3.5$	$M = 4.25$	$F(1, 6) = .7$
	$SD = 1.73$	$SD = .5$	$p = .44$
5. The simulated environment was believable	$M = 3$	$M = 3.75$	$F(1, 6) = 1.42$
	$SD = 1.15$	$SD = .5$	$p = .28$

A frequency distribution of participants’ ratings for the first three questions and question 5, in the post questionnaire revealed that experts tended to have greater variability in their ratings, with two experts giving ratings of 2 or lower, and 2 experts giving ratings of 3–5. Novice participants’ ratings clustered together closer to the middle of the scale. For question 4, seven of eight participants gave ratings that agreed or strongly agreed when asked if they though CEP would be a useful tool in Follow-the-Sun Operations.

Question six asked participants to share any other thoughts on complex event processing. Of eight participants, seven submitted responses. Overall, there were three major themes. The first theme encapsulates the responses that were entirely positive in

regard to CEP. The second theme encapsulates positive responses but also captures some dependencies. Lastly, the third theme captures the desire for validation of the concept.

5 Discussion

The Deep Space Network is the system that provides support for all interplanetary missions of space agencies around the world. As DSN demands continue to grow, there will be an increased need to understand how the technologies designed to address those demands affect operations. The Follow-the-Sun paradigm shift will surely see a spike in complexity and in the number of tracks that are attended to by a human operator (Choi et al. 2016). To help operators, automated systems will likely be employed. Therefore, there should be research aimed at understanding how human operators interact with those technologies and what factors contribute to overall system success.

The purpose of this study was to better understand the impact of a cognitive aid, operationally known as Complex Event Processing (CEP), on LCO workload, performance, and acceptance. We focused on two groups of LCOs. The first group were the expert LCOs who have many years of experience with configuring, monitoring, operating, and troubleshooting DSN tracks. The second group were novice proxies, composed of DSN TSS who specialize in monitoring, but do not configure, operate, nor troubleshoot tracks. Each participant was asked to interact with a DSN simulator where we presented them with four scenarios that they had to resolve. Half of the conditions provided them with the assistance of a cognitive aid and half of them did not. The sample size was small, so there were few significant effects. The results show numerical trends that distinguish experts from novices in their system interactions and efficiency. However, most of these differences were not statistically significant. This study was limited in that the population of DSN operators in the US are small, and this in turn limited the sample size of participants in the study. Therefore, trends evident from this study need to be investigated further in future studies.

The first and second research questions were: What effect will the cognitive aid have on operator performance in the DSN framework? Will experts and novices perform differently using the cognitive aid? Results indicate that neither the effects of group nor assistance were statistically significant for the total number of clicks. Experts tended to perform more clicks compared to novices, but they tended to engage in the same number of clicks for cognitive aid than no cognitive aid conditions. Novices had slightly less clicks in conditions with a cognitive aid than in conditions without. Rather than limiting their strategies to cognitive aid advisories, experts tended to rely more on their internal troubleshooting schemas. The observed pattern is consistent with the idea of expert-based intuition (Salas et al. 2010). This theory proposes that in the later stages of experience, the decision-maker draws on a deep and rich knowledge base from extensive experience within a domain such that decisions become intuitive. Intuition, by definition, occurs without outside assistance. It is the product of “affectively charged judgments that arise through rapid, non-conscious, and holistic associations” (Dane and Pratt 2007). Thus, experts may rely on intuition more than cognitive aids.

There was a significant effect of assistance for total time on task, when timeouts were taken into consideration. Task completion times were significantly lower when cognitive aid assistance was available when compared to when it was not. This suggests that regardless of group, participants tended to resolve issues faster when a cognitive aid was available. The cognitive aid therefore facilitated detection and resolution of issues by narrowing down the causes and resolutions to those issues. This finding is consistent with the idea that a cognitive aid is associated with significantly improved operational performance (Arriaga et al. 2013). The effect of group was not found to be significant. Although experts were numerically faster at resolving issues than novices, they also tended to engage in more clicks, indicative of trying to solve the problem on their own. It could be the case that if the expert group trusted the cognitive aid more, they would also see a larger benefit in the time to detect and resolve the problem.

The second analysis did not include the 600 s ceiling value for timeouts and did not yield any significant effects. This suggests that the effect of assistance found in the previous analysis was driven by observed timeout times. However, the numerical trend in the data remained the same: experts were more efficient than novices at resolving issues. A descriptive analysis of timeouts revealed that most of them occurred in the novice group during conditions where cognitive aid assistance was not available. This is not surprising, as the novice group are not as familiar with the procedural nature of troubleshooting tracks. Recall that the novice group was composed of DSN Track Support Specialists, a role that is largely responsible for supervising configurations and not to directly manipulate the displays to solve issues. It is no surprise that 62.5% of timeouts (5 of 8) occurred with this group as their schemata for LCO tasks is limited. The expert group was responsible for 37.5% of timeouts (3 of 8). With more time, the experts would have likely been able to solve the problem. Thus, even for the expert group, a cognitive aid has the potential of improving the efficiency of a solution, if it is used.

The third research question was: What effect will the cognitive aid have on operator subjective workload in the DSN framework? No significant effects of group nor assistance were found for workload. Participants in both groups rated their workload about the same, regardless of assistance condition. The mean workload ratings of expert and novice participants across both assistance conditions was 40.6. In Grier's (2015) cumulative frequency distribution of global TLX scores, a rating of 40.6 is greater than 30% of all scores. Furthermore, if only scores for monitoring tasks are considered, as were implemented in this study, a rating of 40.6 is only above 25% of observed scores. According to this analysis, both groups were experiencing low workload levels. The biggest numerical difference was observed with the novice group: participants experienced lower levels of workload in conditions where they had cognitive aid assistance. This suggests that a cognitive aid may potentially benefit novice workload levels to a greater degree than experts. However, this conclusion cannot be made based solely on the results of the present study.

The final research question was: Will novice operators have higher acceptance ratings for the cognitive aid compared to experts? Responses to post questionnaires showed that expert participants had mixed feelings about the usefulness of a cognitive aid in operations. A frequency distribution showed that expert ratings tended to be split,

such that half felt they were likely to accept CEP, and the other half felt unlikely to accept CEP. One expert participant consistently rated all questions a six and another expert participant rated all questions a seven. Experts' mean ratings resulted in neutral responses to five of six questions pertaining to the Perceived Usefulness scale. This makes sense – experts take pride in their ability to perform their jobs. A tool that diagnoses and provides resolutions advisories on their behalf, in essence, performs an aspect of their job for them. However, experts generally agreed that a cognitive aid could be a useful tool in Follow-the-Sun operations. This indicates that experts acknowledge that CEP can be a useful tool in future operations.

Novices had higher mean acceptance ratings for complex event processing, and their responses were more clustered together. All novices had the greatest agreement when asked if they believed using CEP in operations would enable them to accomplish tasks more quickly. In the post questionnaire, both groups utilized the halfway point more than they did on the Perceived Usefulness scale. This may be due to the fact that the PU scale is a seven-point scale while the post questionnaires used a five-point scale.

Participant open-ended feedback provided insight to opinions about complex event processing. The majority of responses captured positive dispositions toward CEP. Overall, participants indicated there was utility for CEP in Follow-the-Sun operations, training, and learning. However, areas for improvement were identified. CEP should be robust in its diagnosis and recommendations to facilitate operator trust. Additionally, the CEP needs to provide system transparency to keep LCOs in the loop. One expert participant, who rated all PU scale questions a 7, expressed that CEP will be useful as it matures over time. A second expert participant, who rated all PU scale questions a 6, expressed the need for real world validation of CEP. These opinions on CEP highlight important system attributes for trust in human-machine interactions (Sheridan 1988). Operators communicated the need for a reliable, robust, and transparent system, which coincide with Sheridan's (1988) list of attributes that facilitate trust in the human-machine environment.

References

- Levine, A.I., DeMaria, J.S., Schwartz, A.D., Sim, A.J.: The Comprehensive Textbook of Healthcare Simulation. Springer, New York (2013). <https://doi.org/10.1007/978-1-4614-5993-4>
- Singh, T.D.: Incorporating cognitive aids into decision support systems: the case of the strategy execution process. *Decis. Support Syst.* **24**(2), 145–163 (1998)
- Playbook: Federal Aviation Administration (2017). <https://www.fly.faa.gov/Operations/playbook/current/current.pdf>. Accessed 2 June 2017
- Catalano, K.: The world health organization's surgical safety checklist. *Plast. Surg. Nurs.* **29**(2), 124 (2009)
- DSN Functions: Jet Propulsion Laboratory (n.d.). <https://deepspace.jpl.nasa.gov/about/DSNFunctions/#>. Accessed 17 Apr 2017
- Choi, J.S., Verma, R., Malhotra, S.: Achieving fast operational intelligence in NASA's deep space network through complex event processing. In: *SpaceOps 2016 Conference*, Daejeon, Korea (2016)

- Johnston, M.D., Levesque, M., Malhotra, S., Tran, D., Verma, R., Zendejas, S.: NASA deep space network: automation improvements in the follow-the-Sun Era. In: 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina (2015)
- Fletcher, K.A., Bedwell, W.B.: Cognitive Aids: design suggestions for the medical field. In: Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care, vol. 3, no. 1, pp. 148–152 (2014)
- Arriaga, A., Bader, A., Wong, J., Lipsitz, S., Berry, W., et al.: Simulation-based trial of surgical-crisis checklists. *New Engl. J. Med.* **368**(3), 246–253 (2013)
- Van de Merwe, K., Oprins, E., Eriksson, F., Van der Plaat, A.: The influence of automation support on performance, workload, and situation awareness of air traffic controllers. *Int. J. Aviat. Psychol.* **22**(2), 120–143 (2012)
- Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* **52**, 139–183 (1988)
- Davis, F.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* **13**(3), 319–340 (1989)
- Salas, E., Diazgranados, D., Rosen, M.A.: Expertise-based intuition and decision making in organizations. *J. Manag. JOM* **36**(4), 941–973 (2010)
- Dane, E., Pratt, M.G.: Exploring intuition and its role in managerial decision making. *Acad. Manag. Rev.* **32**, 33–64 (2007)
- Grier, R.: How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. *Proc. Hum. Factors Ergon. Soc. Ann. Meet.* **59**(1), 1727–1731 (2015)
- Sheridan, T.: Trustworthiness of command and control systems. *IFAC Proc. Vol.* **21**(5), 427–431 (1988)