Human Factors and Simulation, Vol. 83, 2023, 174–183
https://doi.org/10.54941/ahfe1003575

A Cognitive Model for Guiding Automation

Christopher A. Stevens[1], Christopher R. Fisher[2], and Mary E. Frame[2]

[1]Air Force Research Laboratory, Wright-Patterson, OH 45324, USA
[2]Parallax Advanced Research, Beavercreek, OH 45431, USA

ABSTRACT

A variety of systems exist for managing human-machine team throughput and effectiveness. One example is autonomous managers (AMs), software that dynamically reallocates tasks to individual members of a team based on their workload and performance. Cognitive models can inform these technologies by projecting performance into the future and enabling "what-if" analyses. For example, would removing a task from an individual whose current performance is low cause them to improve? Conversely, can a team member who is currently performing well handle even more work without dropping performance? In the present study, we develop and validate a cognitive model built in the Adaptive Control of Thought – Rational (ACT-R) cognitive architecture in a novel empirical paradigm: The Intelligence, Surveillance, and Reconnaissance Multi-attribute Task Battery (ISR-MATB). In this task, participants engage in a analog ISR task in which they must integrate information from several subtasks to arrive at a decision about a situation. These tasks include searching visual displays, listening for audio chatter, making decisions based on multiple cues, and remaining vigilant for signals. The tasks are based upon analogous laboratory psychology tasks to improve empirical rigor. Eight participants completed the task under two 30-minute conditions: easy and difficult. The difficult task required searching more complex stimuli in the audio and visual domain than in the easy condition. In addition, subjective workload ratings (NASA-TLX) were collected. We describe the preliminary behavioral and self-report results, as well as the ACT-R model's fit to the behavioral data. Further, we describe a new method for workload visualization and task decomposition using model-based analyses.

Keywords: Automation, Cognitive modeling, Act-r, Workload profiling, Dynamic task allocation

INTRODUCTION

Autonomous vehicles and technology are becoming increasingly common in commercial, civilian, and military applications. This creates challenges for human operators as they interact with these systems and maintain situation awareness of their operations (Endsley & Kiris, 1995). Moreover, massive amounts of data can be quickly and autonomously processed and presented for human operators, who must in turn make decisions using that data. As these applications are further developed, it is important to bear in mind the strengths and limitations of human operators and present information in ways that does not overwhelm their moment-to-moment cognitive capacities. In this paper, we discuss the integration of two technologies that help to address these concerns: autonomous managers and cognitive models.

Autonomous managers (AMs) are algorithms that monitor the performance of teams of agents (both human and artificial) and make real time, online tasking recommendations to maximize team performance. In principle, these systems can make these recommendations based on any data that is both relevant and measurable in real-time. For example, if the performance of agents can be measured in real-time, then an AM could monitor the task performance of all the agents in a team. If an agent's performance on a task drops, then the AM could recommend reassigning that task to another agent (Fisher et al., 2022).

Performance data are not always available during online operation. However, this limitation can be overcome using metrics that are correlated with performance or by using models to predict performance. One promising and widely studied metric is workload. Workload can be measured via performance, subjective report (Roscoe & Eillis, 1990; Hart & Staveland, 1988), or by physiological metrics such as EEG, eye movements, and heart rate (Matthews et al., 2015; Christensen et al., 2020). The correspondence between workload and performance is complex and subject to variability (Hancock & Matthews, 2019), but it is nevertheless potentially informative in a variety of contexts, especially in cases where individuals approach a degree of workload beyond which performance drops below an acceptable level (Grier et al., 2008).

A complementary technique that has recently shown promise for projecting workload and performance into the future is cognitive modeling. A cognitive model is a computational or mathematical model that represents cognition and behavior. A commonly used framework for building these models is Adaptive Control of Thought – Rational (ACT-R; Anderson, 2007). ACT-R is a cognitive architecture – a modeling framework based upon a comprehensive theory of cognition in which agents can be developed to complete specific tasks. It assumes that cognition emerges from the interaction of distinct modules each responsible for a specific function (e.g., memory, vision, motor control, etc.). As such, it is situated at a level of analysis that allows it to make meaningful and valid predictions about moment-to-moment cognitive demands.

Jo et al., (2012) demonstrated that ACT-R can predict both the performance and subjective workload of subjects performing a suite of laboratory tasks which recruit different combinations of cognitive capacities. Specifically, a weighted sum of activity across the various modules predicts subjective workload scores as measured by the NASA-TLX. Stevens et al. (2022) showed that a similar analysis has convergent validity with physiological workload signals.

In the current paper, we report preliminary empirical and modeling results aimed at applying AM and cognitive modeling technology to a context that represents an analytical workflow in which an operator must monitor multiple channels and make decisions based on ambiguous information. This workflow is represented by a laboratory task we have developed that we call the Intelligence, Surveillance and Reconnaissance Multiattribute Task Battery (ISR-MATB). In a previous paper, we demonstrated that our proposed integration between AMs and cognitive models has the potential to improve team performance in the novel task environment (Fisher et al., 2022).

OVERVIEW

Our work in this paper will be structured as follows. First, we will describe our novel task environment and experiment. Then, we will describe the cognitive model we have developed to explain the empirical data. Then we will discuss ways in which the model can make workload predictions that will be useful to an AM.

ISR-MATB

The Multi-Attribute Task Battery (MATB; Santiago et al., 2011) is commonly used to study workload experienced by pilots while operating aircraft. The ISR-MATB is a modification of the MATB that reflects demands in ISR environments. One feature that distinguishes the ISR-MATB from the original MATB is the inter-dependence between subtasks. In ISR, analysts must obtain intelligence from disparate and multi-modal sources, and integrate it into an actionable decision. Difficulty of subtasks can be configured to achieve a desired level of workload. In the present work we varied difficulty levels in the VST and AST and describe those manipulations in their respective sections.

Psychomotor Vigilance Task (PVT)

During discovery, mission requirements may be updated based on new intelligence or emerging requirements, so vigilance and adaptability toward changing goals is critical. We capture this component of ISR operations with a modified version of the Psychomotor Vigilance Task (PVT; Dinges, 1985). In the PVT, participants respond as quickly as possible to a stimulus presented after a random interval (2-10 seconds). Our modified version of the PVT differed in two regards: (1) the random interval was 0–10 seconds, and (2) on each trial, the stimulus was one of four randomly selected letters with different colors (e.g., black Q), which serves as a target in the VST and AST subtasks.

Visual Search Task (VST)

Visual information, including geospatial intelligence and mobile threat intelligence, are searched by ISR operators for pertinent targets or threats. In the ISR-MATB, we emulate these demands using a conjunctive VST (Treisman, 1980). Participants search for a target from the PVT among an array of scattered distractors, which vary by color and letter. A stimulus is considered a target if it matches on both dimensions (e.g., black Q). On half of the trials the target is present and on the other half of trials the target is absent. Difficulty on the task is manipulated by varying set size and discriminability (Palmer, 1995).

Auditory Search Task (AST)

ISR operations require multimodal intelligence analysis, including searching for targets based on radio communications or audio recordings, which may vary in terms of signal quality. In the ISR-MATB, we emulate these conditions with an audio search sub-task in which a participant scans multiple radio channels with background noise for the search target (e.g., an audio recording of the words "black Q"). The difficulty of this task can be manipulated by changing the number of radio channels and the volume of background noise.

Decision Task (DT)

Following detection of either the presence or absence of the target in the AST and VST, the ISR-MATB contains a multiple-cue DT inspired by similar tasks in the literature (e.g., Sieck and Yates, 2001). Decisions are based on two cues: (1) whether the target state (present or absent) is the same or different between the VST and AST subtasks, and (2) whether confidence in the accuracy of the information is low or high, based on a revealed cue. The decision rule requires all three cues to perform better than chance.

ACT-R Architecture

We developed a model of the ISR-MATB based on the Adaptive Control of Thought-Rational (ACT-R) cognitive architecture (Anderson, et al., 2004). A diagram of ACT-R's architecture is illustrated in Figure 1. ACT-R's architecture is composed of specialized information processing modules for functions such as memory, action, and perception. Each module is connected to a capacity-limited buffer which can process only a single request from the procedural module at a time and return and hold a single chunk of information per request. The resulting information processing bottleneck is responsible, in part, for producing realistic human-like errors and response times.

ACT-R operates as a production system in which cognition unfolds over a series of selection-action cycles called a *production cycle*. In a production cycle, a production-rule (i.e., an IF-THEN statement) is selected based on its match to the state of the architecture, and then its actions are executed. A production rule is the basic unit of procedural knowledge consisting of two components: a set of conditions, and a set of actions that are executed when the conditions are satisfied. For example, a production rule for the VST might

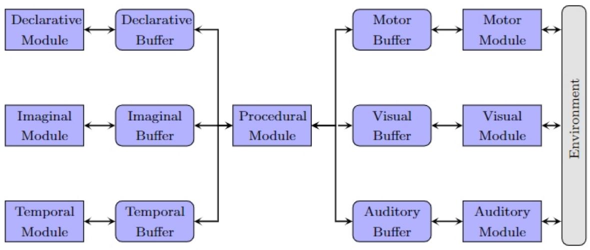

Figure 1: An illustration of the ACT-R cognitive architecture.

specify in natural language, "If the visual object in the visual buffer matches the target, then respond *present*."

ISR-MATB Model

A diagram of the model's strategy can be found in Figure 2. The model completes the subtasks in a clockwise fashion: PVT, VST, AST, and the DT. The model begins by waiting for the target to appear during the PVT. Once the target appears, the model automatically orients to it, encodes the target, and responds. Next, the model performs the VST using the Pre-attentive Attentive Vision (PAAV) extension of ACT-R's visual system (Nyamsuren & Taatgen, 2013) to provide more realistic search behavior. In PAAV, objects have less acuity based on distance from the point of visual fixation. The model fixates on the visual object with the highest visual activation value (a combination of bottom-up and top-down influences) exceeding a dynamic termination threshold (Moran et al., 2013). If no object's activation exceeds the threshold, the model responds "absent". If the model finds the target, it responds "present". Otherwise, it attempts to find a new visual stimulus. After attending to a distractor, the dynamic threshold increases, which increases the probability of responding "absent" on the next visual fixation.

Next the model scans through the radio channels in the AST. If the message matches the target value, the model responds "present". Otherwise, it attempts to find and listen to a new channel. If the model runs out of channels to inspect, it will respond "absent". Finally, in the DT, the model attempts to retrieve a decision rule matching the responses from the VST, AST and confidence cue. If a rule cannot be retrieved from memory, the model guesses randomly.

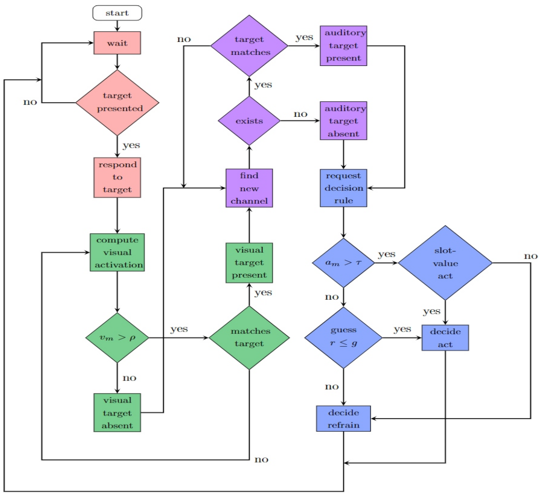

Figure 2: A flow chart of the strategy used in the ISR-MATB model. v_m is the visual object with the highest activation; ρ is the visual termination threshold; a_m is the chunk with the highest memory activation; τ is the retrieval threshold. Nodes are color-coded according to subtask. Pink: PVT, green: VST, purple: AST, blue: DT.

METHOD

Data for this experiment was collected at the University of Dayton Research Institute (UDRI). All data was collected in compliance with the Air Force Research Laboratory's Institutional Review Board standards and protocols. This is an ongoing study, and we present the first 8 subjects of collected data (N = 8). Behavioral data was collected using a task developed in the Unity game engine Prior to full data collection, participants were given a 5-minute practice trial to work with the ISR-MATB and become familiar with the interface and making appropriate responses for each of the inter-dependent tasks. Task difficulty was manipulated within-subjects to induce varying levels of workload. There were two conditions of task difficulty (easy and difficult) that were presented to each participant, where the order of the tasks counterbalanced between subjects. Each condition lasted for 30 minutes. Participants made behavioral response inputs using a standard computer keyboard.

RESULTS

A 4×2 within-subjects ANOVA was conducted with subtask (PVT, VST, AST, and DMT) and difficulty condition (easy, hard) as factors on both accuracy and response times. Greenhouse-Geisser corrections were applied when violations of sphericity were significant. Overall, we observed an effect of subtask on accuracy ($F(1.28, 8.97) = 6.745$, $p = .024$ $\eta_p{}^2 = .491$). There were no significant effects of difficulty condition on accuracy and no interactions. For response times we observed statistically significant effects of subtask ($F(1.26, 8.82) = 459$, $p < .0001$; $\eta_p{}^2 = .98$), difficulty condition ($F(1, 7) = 18.8$, $p = .003$, $\eta_p{}^2 = .728$), and an interaction between difficulty condition and subtask ($F(1.45, 10.18) = 9.411$, $p = .007$, $\eta_p{}^2 = .573$).

Overall, the results suggest participant performance varies reliably on each of the four subtasks in terms of accuracy and response. But the workload manipulation appears to affect only the response times. Moreover, the manipulation does not appear to have an effect on PVT response times. This makes sense because the PVT portion of the task is identical in the two conditions.

A global response to the NASA-TLX was computed for each participant. There was a very slight numerical trend such that participants reported higher global workload in the hard condition ($M = 59.0$, $SD = 12.3$) than in the easy condition ($M = 60.1$, $SD = 16.1$). But this trend did not approach significance.

Model Predictions

We fit the model to the data aggregated across participants. We set the retrieval threshold to -1, activation noise to.20, and base levels to 20 for key mappings and 2 for decision rules. These base level values were selected because we anticipated that the decision rules were more difficult to learn than the key mappings for responding. All other parameters were set to default values.

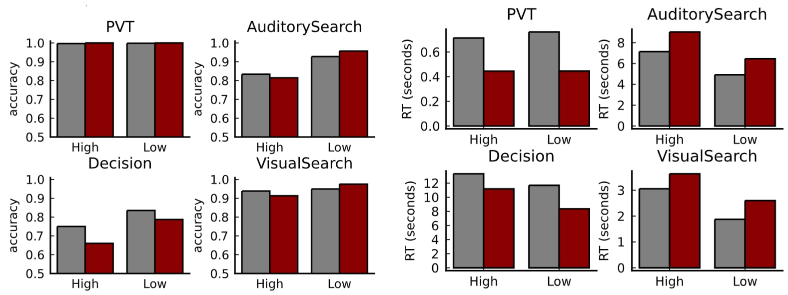

Figure 3: Left: accuracy panelled by subtask. Right: RT panelled by subtask. Observe data in grey and model predictions in red. RTs measured relative to target

The model captured two important qualitative trends for accuracy (see Figure 3). First, the model captured the rank order of accuracy for subtask: PVT > VST > AST > DT. In addition, the model accurately captured lower accuracy in the high difficulty conditions for the AST and the DT.

The model also captured the rank order of reaction across subtasks, starting with the PVT, followed by VST, AST, and ending with the DT. For each subtask, the model also captured the effect of task difficulty in which responses where faster for low compared to high task difficulty. Finally, the model predicts higher workload in the high difficulty condition (M = 1.06) than in the low difficulty condition (M = .87). Note that these workload estimates include the wait time during which workload is estimated to be zero.

Workload Profiles

One role of automation in human-machine teaming is to reassign agents to subtasks in response to increased workload. Automation can leverage Cognitive Metrics Profiling (CMP; Gray et al., 2005; Jo et al., 2012) to infer and predict workload levels across time. CMP uses a cognitive model to estimate workload from the latent cognitive activity across the various modules within the architecture. Thus, increased cognitive activity is interpreted as increased workload. Automation can use CMP in three ways: (1) to estimate overall workload across a period of time, (2) to estimate capacity specific workload (e.g., vision vs. memory), and (3) to identify aspects of a task associated with high workload.

As an example of the third application, Figure 4 shows workload plotted across time and is color-coded according to the probability of working on a subtask. A few points of interest can be gleaned quickly through visual inspection. First, workload peaks acutely during VST, and the DT. Second, increasing the difficulty of the AST leads to the largest sustained period of workload between conditions (in green). This suggests that an intervention targeting the AST would have the most potential to mitigate total workload, but an intervention targeting VST or DT would reduce workload peaks.

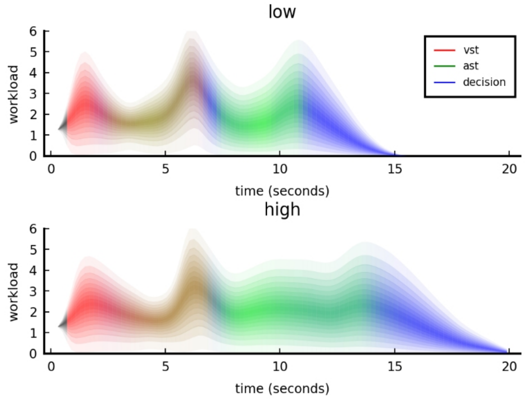

Figure 4: Workload as a function of time for low and high difficulty conditions. Darker regions represent higher workload probability. Workload is color-coded according to the probability of working on a subtask. See legend for subtask colors.

DISCUSSION

In the present work, we have presented a computational model of multitasking performance, validated the model against performance data collected in an empirical task, and demonstrated how the model can make projections of workload that would be useful for dynamic tasking applications.

These preliminary data suggest that the cognitive model provides a good description of participants completing this task. The model accounts for the relative difficulty of the tasks for average participants both in terms of accuracy and response time. It also explains the effects of task difficulty on response time using theories of visual search and human-computer interaction derived from laboratory studies. Finally, the model captures the numerical ordering of the difficulty conditions in terms of subjective load, but the observed differences in these conditions is currently far too small to draw definitive conclusions. If additional data does not reveal a difference between difficulty conditions, it may suggest need for revision of individual module contributions to overall subjective load in the model.

Due to the small sample size, these empirical findings are preliminary. We are in the process of collecting additional data to further inform the model and ensure the generalizability of the findings. Moreover, we are collecting physiological data that we believe will be informative both to the predictions of the model and to the tasking decisions of an Autonomous Manager.

We envision cognitive models informing automation in both offline and online roles. Offline, cognitive models can be used to analyze workflows, interfaces, schedules and task demands and provide projections of expected workload and performance under these scenarios. This input can used prospectively to inform the design of autonomous systems (see Fisher et al., 2022, for an example). Models can also be fit to individual humans or modified based on known individual differences to represent the diversity of performance profiles possible in a task environment.

Alternatively, models can be used in an online capacity using techniques such as model-tracing (Hefferman et al., 2008) and lookup tables (Fisher et al., 2016) to find appropriate model variants that most accurately reflect the present situation. These models can be used to update expected performance projections allowing for online adjustments to task configurations and assignments. For example, a model may project a performance decline due to fatigue after a certain amount of time on task. This projection could be delivered to an AM, allowing the AM to re-task the agent before performance declines. Similarly, a model could determine that a particular task combination will result in effective performance for a shorter period than another combination, allowing the AM to avoid that combination.

CONCLUSION

The integration of cognitive models and autonomous managers is a promising approach for addressing workload, performance, and tasking concerns faced by human machine teams in industrial, commercial, and military settings.

ACKNOWLEDGMENT

The opinions expressed herein are solely those of the authors and do not necessarily represent the opinions of the United States Government, the U.S. Department of Defense, the Department of the Air Force, or any of their subsidiaries or employees. Distribution A: Approved for public release. Case number: AFRL-2023-0647.

REFERENCES

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. Psychological Review, 111(4), 1036.

Christensen, J. C., Estepp, J. R., Wilson, G. F., & Russell, C. A. (2012). The effects of day-to-day variability of physiological data on operator functional state classification. NeuroImage, 59(1), 57–63.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behavior research methods, instruments, & computers, 17(6), 652–655.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. Human factors, 37(2), 381–394.

Fisher, C. R., Frame, M. E., & Stevens, C. (2022). Using cognitive models to design dynamic task allocation systems. Journal of Defense Modeling and Simulation: Applications, Methodology, Technology.

Fisher, C. R., Walsh, M. M., Blaha, L. M., Gunzelmann, G., & Veksler, B. (2016). Efficient parameter estimation of cognitive models for real-time performance monitoring and adaptive interfaces. In Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016). University Park, PA.

Gray, W. D., Schoelles, M. J., & Sims, C. (2005). Cognitive metrics profiling. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 49, No. 12, pp. 1144–1148). Los Angeles, CA: SAGE Publications.

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & St. John, M. (2008, September). The red-line of workload: Theory, research, and design. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 52, No. 18, pp. 1204–1208). Sage CA: Los Angeles, CA: Sage Publications.

Hancock, P. A., & Matthews, G. (2019). Workload and performance: Associations, insensitivities, and dissociations. Human factors, 61(3), 374–392.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in psychology (Vol. 52, pp. 139–183). North-Holland.

Heffernan, N. T., Koedinger, K. R., & Razzaq, L. (2008). Expanding the model-tracing architecture: A 3rd generation intelligent tutor for algebra symbolization. International Journal of Artificial Intelligence in Education, 18(2), 153–178.

Jo, S., Myung, R., & Yoon, D. (2012). Quantitative prediction of mental workload with the ACT-R cognitive architecture. International Journal of Industrial Ergonomics, 42(4), 359–370. doi: 10.1016/j.ergon.2012.03.004.

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich IV, J. (2015). The psychometrics of mental workload: Multiple measures are sensitive but divergent. Human factors, 57(1), 125–143.

Moran, R., Zehetleitner, M., Muller, H. J., Usher, M., 2013. Competitive guided search: Meeting the challenge of benchmark rt distributions. Journal of Vision. 13, 24–24.

Newell, A. (1989). Unified theories of cognition. Harvard University Press.

Nyamsuren, E., Taatgen, N. A., 2013. Pre-attentive and attentive vision module. Cognitive systems research 24, 62–71.

Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. Current directions in psychological science, 4(4), 118–123.

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock Jr, J. R. (2011). The multi-attribute task battery ii (matb-ii) software for human performance and workload research: A user's guide (No. L-20031).

Sieck, W. R., & Yates, J. F. (2001). Overconfidence effects in category learning: A comparison of connectionist and exemplar memory models. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(4), 1003.

Roscoe, A. H., & Ellis, G. A. (1990). A subjective rating scale for assessing pilot workload in flight: A decade of practical use. Royal Aerospace Establishment Farnborough (United Kingdom).

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive psychology, 12(1), 97–136.