Predicting Rapid Shifts in Cognitive Resource Allocation

Justin Brooks
D-Prime LLC
Washington, D.C., USA
email: justin.brooks@dprime.ai

Philip Dyer
D-Prime LLC
Wading River, NY, USA
email: phil.dyer@dprime.ai

Katherine Cox
Human Research and Engineering Directorate
The U.S. Combat Capabilities Development Command
Army Research Laboratory
Aberdeen Proving Ground, MD, USA
email: katherine.r.cox11.civ@army.mil

Jonroy Canady
Human Research and Engineering Directorate
The U.S. Combat Capabilities Development Command
Army Research Laboratory
Aberdeen Proving Ground, MD, USA
email: jonroy.d.canady.civ@army.mil

Kanika Bansal
Human Research and Engineering Directorate
The U.S. Combat Capabilities Development Command
Army Research Laboratory
Aberdeen Proving Ground, MD, USA
email: phy.kanika@gmail.com

Javier Garcia
Human Research and Engineering Directorate
The U.S. Combat Capabilities Development Command
Army Research Laboratory
Aberdeen Proving Ground, MD, USA
email: javier.o.garcia.civ@army.mil

***Abstract*—Human cognition is the set of abilities humans possess to intelligently interact with the world through goal-directed behaviors and is an essential component of daily life. As with several other processes (e.g., physical exertion), cognitive resources are inherently limited. Moreover, as the systems with which humans interact become more complex (e.g., artificial intelligence), the degree and rate of the associated depletion of these cognitive resources can vary dramatically. A real-time signal of cognitive depletion can be used to moderate task demands imposed on humans to enable high-performing human and complex system interactions; however, previous attempts have identified generalizability across both individuals and a variety of tasks to be a significant challenge in modeling cognitive depletion. Here, we present a model that uses physiological measurements and that generalizes across three different real-world tasks and across heterogeneous samples of participants. Specifically using features that are implicated in autonomic and central nervous system activity, our model detects cognitive resource depletion in multiple tasks, including mental arithmetic, simulated navigation/decision-making, and visuospatial/sensorimotor processing. The model makes second-by-second predictions of cognitive resource depletion, which can be used in real-time human-in-the loop systems.**

***Index Terms*—cognitive resource depletion; humans and complex systems**

I. Introduction

Our perceptual experiences coupled with our intrinsic goals guide our subsequent actions. This requires a series of complex perceptual interpretations, sensorimotor processes, and rapid decisions. While this complex cognitive coordination is often completed with relative ease, there are inherent limitations to each one of these constructs. Notable limitations have been observed in working memory [1] and visual attention (e.g., see [2] but see [3]); importantly, these limitations can have an inherent fluctuations independent of ongoing task demands [3] and a wide range of individual differences [4]. Moreover, fluctuating individual states like changes in stress, fatigue, and other physiological factors can impact these limitations in complex ways and sometimes with dramatic consequences [5].

Real-time monitoring of cognitive resources can facilitate optimal performance by providing actionable insights that can be used to trigger individualized adaptations [6]. These individualized interventions could, for example, take the form of retasking to another cognitive domain or a forced intermission of inactivity if a particular individual's cognitive resources are depleted.

The ways in which cognitive resources have been monitored, however, are somewhat limited. For example, at one extreme, subjective measures like the validated NASA-TLX [7] are widely used; however, they require interruption of the ongoing task and are impossible to use in a real-time state assessment system. At the other end of the spectrum, researchers have developed cognitive workload models using a variety of deep learning and machine learning approaches [8]. Some of the biggest challenges with this latter approach are 1) the lack of generalizability across various tasks and 2) the

Copyright (c) The Government of USA, 2022. Used by permission to IARIA. ISBN: 978-1-61208-982-9

limited application or utility outside of artificially constrained laboratory-based tasks.

Here, we introduce a model of Cognitive Resource Depletion (CRD) which is built on the physiological features extracted during a realistic multi-task dataset and is tested in two independent but similarly realistic task paradigms. The physiological measures are derived from electroencephalography (EEG) and electrocardiography (ECG) and were chosen because of their specificity to central nervous system (CNS) and autonomic nervous system (ANS) functions, both of which have been related to cognitive workload in other studies.

The model accounted for significant sources of variability in performance in all tasks, regardless of individual differences or the variety of sensorimotor actions and complex decisions required for each task. Due to the complex interplay between the ANS, the CNS (measures derived from EEG), and the most predictive measurement features extracted within the model, the model appears to be predictive of cognitive depletion in a subject- and task- agnostic fashion, suggesting the critical role the relationship between the CNS and the ANS plays in cognitive resource depletion.

In Section II, the experimental approaches including the types of data and sensors used are described. In Section III the approach to model development and validation and provided. In Section IV, the modeling results including results for each dataset are presented. In Section V a discussion of the results in the context of CRD are described. In Section VI, we present general conclusions and suggested future research directions.

II. Methods and Materials

In this section, we describe the experimental methods used to collect the data and how the model was derived from those data.

A. Overview

For an overview of the model development and deployment, Figure. 1 displays the training, testing, and validation stages of the model. As shown in 1, a single model was developed with the DualTask data and then applied to the independent but similarly dynamic and realistic TeamTask and VisualTargetTask, which were collected in separate populations and at different experimental sites. Validation of the model was initially performed within the DualTask (tan) using a leave-one-out (LOO) procedure. External validation was performed using the TeamTask and VisualTargetTask (green). Descriptions of the datasets are provided in subsequent sections. Briefly, the DualTask dataset was analyzed and used to develop a real-time model of cognitive resource depletion which was hypothesized to occur during the dual task phase of the experiment (i.e., when participants were performing the simultaneous navigation and mental arithmetic tasks). The model was validated with that dataset using a LOO cross-validation approach. Once trained and tested, the model in DualTask was then applied to the TeamTask and VisualTargetTask datasets.

In the TeamTask, the model output was compared to subjective ratings from the NASA-TLX. To validate the model, the continuous model probabilities were thresholded and then integrated over time, resulting in a cumulative amount of time that the model predicted a participant was cognitively depleted. This cumulative number of seconds was correlated with each NASA-TLX subscore and total score.

In the VisualTargetTask, the percentage of time that the model predicted cognitive resource depletion was compared between high and low difficulty conditions. The length of time that participants experienced the low difficulty condition was longer than the high difficulty condition, so it was necessary to compare this normalization.

All protocols were approved by the U.S. Combat Capabilities Development Command Army Research Laboratory (ARL) Human Research Protection Program.

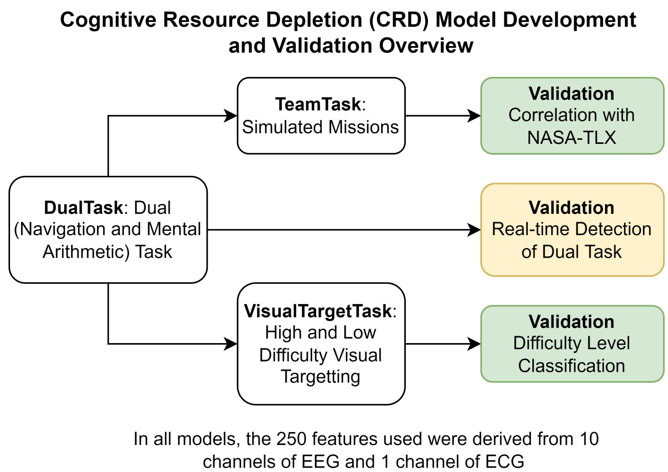

Fig. 1: An overview of the methods presented in the current manuscript.

B. DualTask

Forty-five subjects, recruited from the Los Angeles area, participated in this study [17 females with mean age ± standard deviation (SD) = 36.8 ± 12.3 years; 28 males with mean age ± SD = 41.6 ± 14.4 years]. All subjects were at least 18 years of age or older and able to speak, read, and write English. All subjects signed an Institutional Review Board-approved informed consent form prior to participation and were compensated for their time.

The overarching study aim was to examine physiological markers (i.e., EEG, ECG) of cognitive resource depletion in the context of an experimentally imposed dual-task paradigm. Participants were asked to count the number of target objects in a visual search task. All participants were exposed to all the same target objects, but they were asked to focus on different target objects. Participants were randomly assigned to one out of four target object conditions (1 – Motorcycle, 2 – Humvee, 3 – Furniture, 4 – Aircraft). In addition, the environment also contained non-targets that were used as distractors.

While participants performed the continuous target identification and navigation task, they were simultaneously required

Copyright (c) The Government of USA, 2022. Used by permission to IARIA.

to complete a math task. This math task was given approximately 8 minutes into the experiment and required the addition of a series of 3-4 numbers that were read aloud. Participants were then asked to report the sum aloud while continuing to navigate the virtual environment and count targets. While performing the task, participants wore a 64-lead EEG and 2-lead ECG; the complete description of the data collection has been previously published [9].

C. TeamTask

Participants were seven U.S. Army Soldiers. Six Soldiers were members of the Minnesota U.S. Army National Guard, and one Soldier was assigned to the Army Evaluation Center under the U.S. Army Test and Evaluation Command. The Soldiers were aged 37.71 ± 10.16 (29 – 57 years) and had 19.71 ± 9.52 (12 – 39) years of experience in the Army.

Zephyr BioHarnesses were used to collect ECG data. The BioHarness is a lightweight (50g) belt that is worn around the upper torso, directly against the skin. The 2-lead BioHarness enables the capture and wireless transmission of ECG, torso accelerometry, and respiration data. EEG data were collected using Advanced Brain Monitoring (ABM) X24 EEG Systems. The ABM X24 is a commercial EEG system designed for measuring EEG in an untethered, free-moving manner with minimal impact on the wearer. The system has 20 leads and mounts to the head using an elastic headband.

Subjective measures have been found to be useful in characterizing individual workload and are often used as a validation of other measures. The NASA-TLX [10] has been used extensively to assess individual ratings of overall workload, as well as a breakdown of six categories that are seen as contributing to this overall workload (mental demand, physical demand, temporal demand, performance, effort, and frustration).

In this study, participants completed a set of simulated missions in which vehicle crews navigated to objectives and performed various offensive and defensive maneuvers. Within each mission, participants had to consider multiple factors, including mission completion, enemy engagements, terrain, troops, time, and civilians. The crew was responsible for multiple tasks within each vehicle, including driving, maintaining situational awareness, using the weapon to engage with human-operated opposition forces, and completing mission objectives. Movement through the environment was dynamic and consisted of formations, battle drills, or movements coordinated by the commander and among the team. Participants were fully trained on each of the tasks that they were asked to perform. At the conclusion of each simulated mission, participants completed an electronic version of the NASA-TLX.

D. Visual Target Task

Participants were seventeen students (7 women, 10 men) from a university in the Mid-Atlantic region of the United States. All participants were volunteers recruited through flyers and word-of-mouth on campus, and they provided informed consent in accord with the university's Institutional Review Board. The sample had a mean age of 26.18 years (SD = 3.70).

The dataset used for this analysis came from a neurofeedback study whose methods have been previously published [11]. Briefly, participants wore 64-lead EEG and 2-lead ECG monitors. The task performed was a high difficulty and low difficulty shooting virtual reality simulation in which participants had to shoot enemy targets and refrain from shooting friendly targets. Difficulty was manipulated by target exposure time. In the high difficulty condition, targets were presented very briefly, forcing participants to make a decision on whether to shoot or not very quickly. In the low difficulty condition, the target exposure time was longer, permitting easier discrimination between friendly and enemy targets.

III. Model Development

In the DualTask experiment, the objective was to determine on a second-by-second basis if the participant was in the dual task state or not, which was interpreted as cognitive resource utilization. The model was developed using EEG and ECG features. The EEG features were specifically chosen to characterize frontal/parietal interactions, and the ECG features were chosen to characterize the autonomic nervous system response.

The inputs to the model are features derived from ECG and EEG data. The EEG electrodes used were P3, P4, O1, O2, FP1, FP2, F3, F4, F7, and F8. The output of the model for the DualTask is the predicted probability that the subject was performing the math task. For binary outcomes, the probability was thresholded at 0.5.

Raw ECG data was cleaned using the Neurokit2 package for Python. Raw EEG data was cleaned by low-pass filtering at 30 Hz and then mean-centering across the frontal and parietal channels of interest. ECG features were calculated over 30-second windows, while EEG features were calculated over 3-second windows.

The ECG features calculated were heart rate: min, max, mean, and variability. EEG features calculated were: power of alpha-band frequencies over parietal channels, power of theta-band frequencies over frontal channels, Pearson correlation coefficients between all EEG channel pairs, and functional connectivity features (using the weighted phase lag index) between alpha- and theta-band-passed channels. These frequencies were chosen because of their relationship to mental arithmetic performance.

Subjects were excluded from the dataset if their data suggested either ECG or EEG electrodes were not securely attached, determined by visual inspection of the data. Out of 64 subjects, 19 were excluded. Constraints were applied to heart rate and heart rate variability calculations: heart rates outside the window [25, 175] BPM were deemed invalid; heart rate variabilities greater than 300ms were also deemed invalid. Where invalid values were present, the previous (valid) value was used.

A feed-forward neural network with 2 dense hidden layers and 20 nodes per layer was used. Dropout layers were used

Copyright (c) The Government of USA, 2022. Used by permission to IARIA. ISBN: 978-1-61208-982-9

but replaced in favor of L1 and L2 regularization with factors of 0.003 and 0.01 respectively. ReLU activations were used on the hidden layers, and the final activation was a sigmoid function. Weighted binary cross entropy was used for the loss function, with a weight such that false negatives were penalized twice as much as false positives. Learning curves were used to tune the hyperparameters to prevent overfitting and maximize performance. The final, tuned model was then applied to the TeamTask and VisualTargetTask datasets. In Figure. 2 an example out of the model is provided with a threshold of 0.5. The prediction from the model (blue line) is shown with the Dual Task (mental arithmetic). The model outputs a probability ranging from 0-1 at each second. The probabilities provided by the model were thresholded at different values (dashed lines) to create a binarized time series for subsequent analyses.

A. Exploratory Analysis: Thresholding

As part of our exploratory analyses in the external validation datasets (TeamTask and VisualTargetTask), we examined the effect of various thresholds for the model predictions. This was motivated by the observations that 1) the tasks in each dataset were markedly different and therefore likely had different cognitive demands, and 2) the subjects used in each dataset were different. Thus, we wanted to examine whether different thresholds might result in different model predictions based on the differences in tasks and the individuals performing them. For these analyses, we varied the threshold for the model prediction from 0.1 to 0.9 in increments of 0.1. At each threshold level, we binarized the models' probability output and computed a prediction. In TeamTask, these varied threshold predictions were compared with the NASA-TLX subscores and total score, and in VisualTargetTask these thresholds were used to compare the differences across high and low difficulty conditions. In Figure. 3, the results of this analysis are shown across NASA-TLX subscores and various thereshold is provided. Mental demand and Temporal Demand were highly correlated with the model output across a wide range of thresholds. This statistically significant correlation over a range of thresholds was not evident for the other subscales of the NASA-TLX nor the total score.

IV. Results

In this section, results of the model output are discussed separately and concludes with a discussion of the dynamic thresholding results.

A. Dual Task

The model presented in this manuscript was developed and fit to DualTask using the Matthews Correlation Coefficient (MCC) as the objective function. This was chosen because of its ability to better assess the performance of the model on unbalanced data.

The performance of the model using a LOO cross validation was evaluated with accuracy, precision, recall, F1 score, and the Matthews Correlation Coefficient. These values (mean +/-

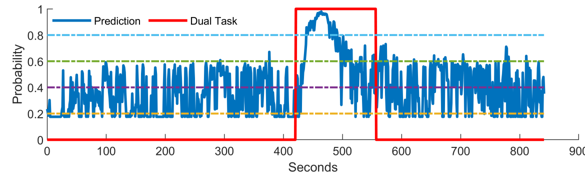

Fig. 2: Example dynamics in prediction of the DualTask and thresholding.

standard deviation) were: 0.72 +/- 0.06; 0.30 +/- 0.12; 0.40 +/- 0.18; 0.336 +/- 0.12; 0.18 +/- 0.15, respectively.

Feature importances were determined to understand what physiological parameters contributed to the prediction of being in the DualTask. Of the 250 features used in the full model, we report the 10 most important features in order of most-to-least important: mean heart rate, minimum heart rate, maximum heart rate, heart rate variability, FP1T, FP2-F4 correlation, O1-F4 correlation, F7-FP2 correlation, O2-FP2 correlation, theta band P3-FP2 WPLI.

B. Team Task

In TeamTask, participants performed realistic military missions (e.g., area defense) in a simulated environment. At the conclusion of each mission (75 minutes), participants completed a NASA-TLX survey. Our model, trained on DualTask, was used to create second-by-second predictions. These binarized predictions were integrated over time and used to correlate with NASA-TLX subscores and total score, as we expected a survey response to be based on perceived lasting cognitive burden throughout the task.

Using a threshold of 0.5 (that was used to train the model in the DualTask), significant correlations were observed for the mental demand ($r = 0.42$, $p = 0.0197$) and temporal demand ($r = 0.42$, $p = 0.0201$) components of the NASA-TLX. This was an expected finding, considering known difficulties in multitasking and validated estimates of workload determined by NASA-TLX. However, we also explored additional thresholds in deploying the model in TeamTask, as the demand and burden could have been considerably different from the DualTask procedures. It is likely that different tasks would have different levels of workload and thus, thresholds should be adapted. We show this exploratory analysis below.

Over a range of thresholds, mental demand and temporal demand showed statistically significant correlations (r = 0.38-0.50; all p ¡ .05). At the very highest thresholds, additional components of the NASA-TLX were correlated, including the NASA-TLX total score with the highest noted correlation ($r = 0.5536$, $p = 0.01$).

C. Visual Target Task

In the VisualTargetTask experiment, participants performed a simulated shooting task in high difficulty and low difficulty conditions. The model output was compared between the difficulty conditions. At a threshold of 0.5, there was a significant difference between the model predictions for

Copyright (c) The Government of USA, 2022. Used by permission to IARIA.
ISBN: 978-1-61208-982-9

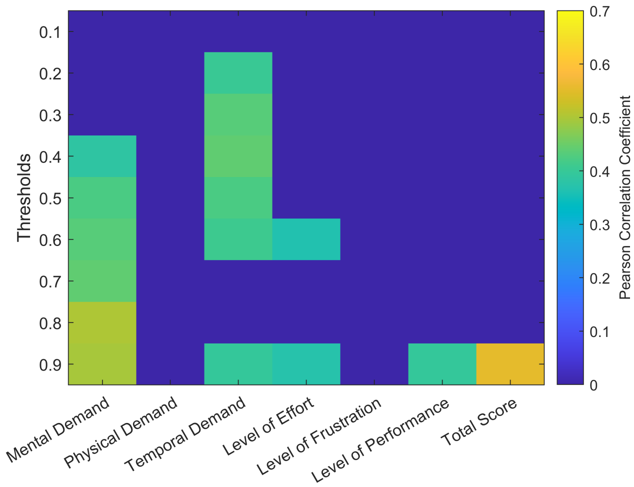

Fig. 3: Correlations between the amount of time cognitively depleted and the various subscores of the NASA-TLX (x-axis) across a number of thresholds (y-axis).

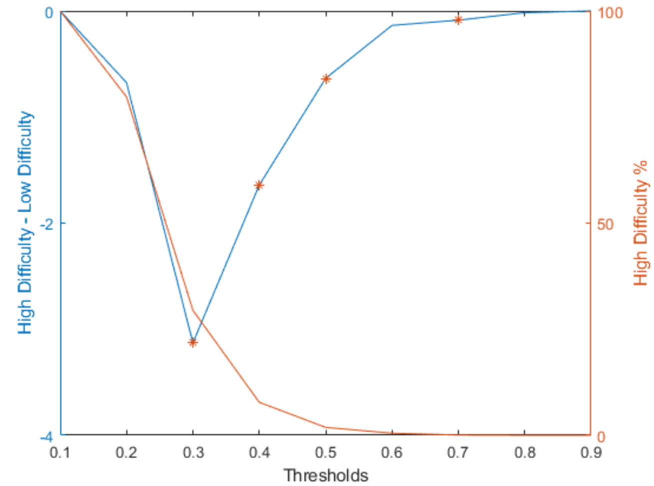

Fig. 4: Exploratory analysis was performed to investigate relative thresholding of probabilities and its effect on discriminating high and low difficulty conditions.

the high and low difficulty conditions. In the low difficulty condition, the model estimated that participants experienced cognitive resource depletion 2.5% of the time compared to 1.86% in the high difficulty condition (difference = 0.64%, p = 0.0013, via t-test). This was somewhat unexpected, as we predicted that there would be greater model-determined CRD for the high difficulty compared to the low difficulty condition; however, the unexpected result may be due to peculiarities in the design of this task. Specifically, accurately shooting the targets was made more difficult by less exposure time to the target, which then could have impacted the way in which resources are deployed (e.g., more reflexive responses). Despite this peculiarity, the CRD model can successfully classify difficulty across task conditions.

We also performed an exploratory analysis over thresholds as was done for the TeamTask (Figure 4). In Figure. 4, the mean differences between high and low difficulty are plotted for various thresholds (blue line). The blue line shows the difference between the model predictions, while the orange shows just the high difficulty predictions. Starting at a threshold probability of 0.3, there is a statistically significant difference between the model predictions for the two conditions (asterisks) which were determined by t-test. For all comparisons, the low difficulty condition had higher levels of cognitive resource depletion predictions. This number represents the difference in time (expressed as a percentage) between the time spent in a CRD state for both difficulty conditions. For thresholds of 0.3, 0.4, 0.5, and 0.6, a significant difference was observed wherein the model predicted more CRD in the low difficulty condition than in the high difficulty condition (asterisks). As a reference, the percentage of time in spent in a CRD state is shown for the high difficulty task alone without the low task difficulty percentages subtracted (orange).

V. Discussion

We have developed a continuous and generalizable Cognitive Resource Depletion (CRD) model that successfully estimates resource depletion in a variety of realistic tasks and that is robust to individual differences from non-invasively collected physiological signals. Physiological signals collected during three different realistic tasks were used as predictors to train and test the model, and the reliable and validated NASA-TLX survey [7], [10] or experimental manipulations were used to validate the cognitive burden of each individual subject. The CRD model, trained on a dataset where subjects were continuously monitoring an environment and occasionally asked to perform an additional task, the DualTask dataset, was able to be deployed on two independent datasets: a team-based navigation and decision making task (TeamTask) and an individual visual target detection task (VisualTargetTask) with different levels of difficulty. This model, and the underlying physiological signals used as predictors, appears to successfully predict a state that underlies each one of these tasks and is behaviorally relevant.

A. Generalizability of the CRD Model

Generalizability of models in human state assessment is often quite challenging, as there are many contributing factors that deter these models from being generalizable, including individual differences in cognitive strategy, measurement noise, and task-specific physiological targets. Our CRD model was trained on a single dataset that involved monitoring an environment and performing mental arithmetic and that was purposely designed to engage multiple systems and burden subjects. In this model, we attempted to predict whether there was a single task being completed or subjects were engaged in the dual task. Model output was then validated with NASA-TLX subscales when the model was deployed on

Copyright (c) The Government of USA, 2022. Used by permission to IARIA. ISBN: 978-1-61208-982-9

the TeamTask, a challenging and complex realistic navigation and decision making task that required coordination between individuals. Interestingly, the estimated total time of cognitive depletion was significantly correlated with multiple subscales of the NASA-TLX and most robustly (across many thresholds) with the mental and temporal demand subscales. Moreover, in the VisualTargetTask, this time of cognitive depletion was also associated with task difficulty, however in a way that was opposite of what was predicted. Perhaps due to the implementation of "difficulty" in this task and different cognitive resource deployment compared to the other tasks, the CRD model did not relate to task difficulty as predicted, but it still predicted a difference in these conditions. Thus, it appears as though the CRD model may be tracking a cognitive construct that changes continuously and is sensitive to multi-tasking, perceived mental and temporal demand, and task difficulty, which we call cognitive depletion.

The training set, a dual visual monitoring and math task, appears to sufficiently engage cognitive depletion such that the CRD Model does not need additional training for robustness. This dual task robustness is unique in its use of mathematics and visual processing and general multi-tasking. Performing arithmetic calculations in one's mind involves a complex coordination of different brain regions and networks [12] and a level of abstraction that separates itself from both the object and social world in which we live and interact [13] but paradoxically may be applied to both [14]. Multi-tasking is also a specialized state that recruits more neural resources than a single task alone [15] and can potentially lead to distraction, overload, and information loss. These properties, coupled with current notions of general intelligence within the brain [16], could suggest that the training task (DualTask) is uniquely suited for a generalized model of cognitive depletion, as it engages resources within the mind (through the mathematics component) that can be applied to a variety of experiences and to multitasking. Future studies may further test the generalizability of the model, perhaps using more abstract tasks and parametrically modulating the number of tasks.

Interestingly, the CRD model is also robust to individual differences, as it does not need to be calibrated to each subject. This result of robustness to individual differences underlying the construct of cognitive depletion is quite surprising. Intuitively, one might understand that cognitive depletion would require individual calibration, especially when considering the model is based on physiological signals that often suffer from nuisance signals, especially in a realistic non-laboratory setting [17], [18]. It may be, however, that special properties of the features extracted and used in the model may have substantially contributed to the generalizability across subjects.

B. Covert physiological events and the CRD Model

Physiological monitoring of covert mental events is commonly used to understand the mechanistic properties of thought, including a variety of cognitive phenomena (e.g., perception, emotions, etc.) and fleeting mental states (e.g., fatigue, stress). Within the CRD Model, heart rate information and specific EEG features (i.e., connectivity) were the most diagnostic to this cognitive depletion.

The most informative features of the ANS were those associated mostly with fluctuations in sympathetic and parasympathetic nervous system activity; specifically, modulation of heart rate and heart rate variability [19], [20]. Interestingly, these features were among the most important among the 250 features tested. While the relationships between cognitive workload and autonomic metrics have been studied extensively [21], cardiovascular metrics such as those used here vary significantly with age, and our participants varied quite dramatically in age. This may suggest that the CRD model learns complex relationships between these features which transcend this inherent limitation in cardiovascular metrics.

The most informative EEG features involved not the commonly used power fluctuations in EEG, but rather the statistical dependence between different sensors, especially involving the frontal cortex (see Results). This interesting finding is perhaps not surprising, as power fluctuations in EEG are often susceptible to many different sources of noise [22], especially in a realistic environment in which subjects may move more freely or perform continuous tasks in which eyeblinks and other movements are difficult to control. We have recently shown that in another realistic tasks (i.e., driving), a synchronization metric between EEG sensors was also more sensitive to task demands than fluctuations in power [23] to the extent that if power alone were used, no neural differences would have been observed. Moreover, fluctuations in neural networks underlying a large variety of tasks has been shown to be predictive of a large and growing list of behaviors including multitasking [24] and adaptations to new stimuli or stress [25].

Thus, in addition to the idiosyncrasies of evoked task demands, underlying physiological features of the CRD Model appear to have a substantial contribution to its generalizability. Future studies may be able to deploy this model in different settings and contexts, using additional features that may capture the dynamic physiological processes more broadly and perhaps extending the simple statistical dependencies between sensors used here to more complex systems or network science approaches known to capture physiological behavior at a variety of scales.

C. Temporal Resolution and Thresholding

An overarching concept that emerges between the tasks, model fit, physiological sensors, and underlying construct which we suggest we are predicting is that of temporal resolution. The convergence of these items into a successful and robust model suggests a waxing and waning of cognitive depletion across time, and while our thresholding and correlational analyses suggest a pathway in which it may be calibrated for individual tasks, this model could also provide a framework in which to detect other potential states of interest. Currently, we suggest that thresholds will need to be determined on a per task basis which is congruent with the notion that various tasks impose varying levels of cognitive demand. In order to support generalizability across tasks we suggest two future

Copyright (c) The Government of USA, 2022. Used by permission to IARIA. ISBN: 978-1-61208-982-9

research directions. First, the model output can be thresholded at varying levels and an integration across those thresholds performed to identify robust periods of CRD detection. This would generate additional parameterization (e.g., continuity of significant differences across thresholds). This type of analysis is similar to cluster-based statistical testing [26]. A second approach could involve empirical derivation of appropriate thresholds for a dictionary of task types (e.g., visuospatial demanding task, mental arithmetic, etc.). This second approach has some traction in the cognitive ergonomics literature that examine different types of naturalistic tasks [27]. With respect to timescales, the EEG and ECG features were extracted in 3 sec and 30 sec windows respectively, thus limiting the fluctuations which may be diagnostic for other states or limiting the generalizability of the model. Rather than a limitation, these two different timescales could also have contributed to the model's robustness. Future studies may explore the complex dynamics between the length and dynamics underlying cognitive constructs, the measurement resolution, and the need for temporally flexible or scale-free feature extractions for covert physiological recordings.

VI. Conclusions and Future Directions

The continuous, generalizable, and robust CRD Model successfully estimates resource depletion in a variety of realistic tasks and is resilient to individual differences from non-invasively collected physiological signals; however, the deployment of this model is still limited. Future investigations of the CRD Model robustness may inspect its success in non-visual tasks and expand the context to additional real-world applications. Additionally, future research may include additional physiological sensor types and features including fNIRS and pupillometry (although the latter is difficult without measurement of ambient luminence). Finally, more psychological assessments will be useful to understand its uses and limitations. Future applications may use this model in a closed-loop fashion where real-time assessments are used to drive adaptations between humans and complex systems.

Acknowledgements

Research was sponsored by the U.S. Army Research Laboratory's (ARL) Human Research and Engineering Directorate (HRED) and was accomplished under Cooperative Agreement Number W911NF-21-2-0108. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory's (ARL) Human Research and Engineering Directorate. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological Review*, vol. 63, no. 2, pp. 81–97, Mar. 1956. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0043158

[2] J. M. Wolfe, "The binding problem lives on: comment on Di Lollo," *Trends in Cognitive Sciences*, vol. 16, no. 6, pp. 307–308; author reply 308–309, Jun. 2012.

[3] V. Di Lollo, "The feature-binding problem is an ill-posed problem," *Trends in Cognitive Sciences*, vol. 16, no. 6, pp. 317–321, Jun. 2012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364661312000988

[4] E. Hunt, J. W. Pellegrino, and P. L. Yee, "Individual Differences in Attention," in *Psychology of Learning and Motivation*. Elsevier, 1989, vol. 24, pp. 285–310. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S007974210860540X

[5] A. H. Taylor and L. Dorn, "Stress, Fatigue, Health, and Risk of Road Traffic Accident among Professional Drivers: The Contribution of Physical Inactivity," *Annual Review of Public Health*, vol. 27, no. 1, pp. 371–391, Apr. 2006. [Online]. Available: https://www.annualreviews.org/doi/10.1146/annurev.publhealth.27.021405.102117

[6] X. Gu, Z. Cao, A. Jolfaei, P. Xu, D. Wu, T.-P. Jung, and C.-T. Lin, "EEG-Based Brain-Computer Interfaces (BCIs): A Survey of Recent Studies on Signal Sensing Technologies and Computational Intelligence Approaches and Their Applications," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 5, pp. 1645–1666, Sep. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9328561/

[7] S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006. [Online]. Available: http://journals.sagepub.com/doi/10.1177/154193120605000909

[8] Y. Zhou, S. Huang, Z. Xu, P. Wang, X. Wu, and D. Zhang, "Cognitive Workload Recognition Using EEG Signals and Machine Learning: A Review," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–1, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9458953/

[9] L. R. Enders, R. J. Smith, S. M. Gordon, A. J. Ries, and J. Touryan, "Gaze Behavior During Navigation and Visual Search of an Open-World Virtual Environment," *Frontiers in Psychology*, vol. 12, p. 681042, 2021.

[10] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Advances in Psychology*. Elsevier, 1988, vol. 52, pp. 139–183. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0166411508623869

[11] D. P. Spangler, S. Alam, S. Rahman, J. Crone, R. Robucci, N. Banerjee, S. E. Kerick, and J. R. Brooks, "Multilevel Longitudinal Analysis of Shooting Performance as a Function of Stress and Cardiovascular Responses," *IEEE Transactions on Affective Computing*, vol. 12, no. 3, pp. 648–665, Jul. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9097281/

[12] M. Delazer, F. Domahs, L. Bartha, C. Brenneis, A. Lochy, T. Trieb, and T. Benke, "Learning complex arithmetic—an fMRI study," *Cognitive Brain Research*, vol. 18, no. 1, pp. 76–88, Dec. 2003. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0926641003002155

[13] M. Mitchelmore and P. White, "Abstraction in Mathematics and Mathematics Learning," *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, vol. 3, pp. 329–336, 2004.

[14] P. L. Ferrari, "Abstraction in mathematics," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 358, no. 1435, pp. 1225–1230, Jul. 2003. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rstb.2003.1316

[15] K. P. Madore and A. D. Wagner, "Multicosts of Multitasking," *Cerebrum: The Dana Forum on Brain Science*, vol. 2019, pp. cer–04–19, Apr. 2019.

[16] J. Hawkins, M. Lewis, M. Klukas, S. Purdy, and S. Ahmad, "A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex," *Frontiers in Neural Circuits*, vol. 12, p. 121, Jan. 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fncir.2018.00121/full

[17] Lun-De Liao, Chin-Teng Lin, K. McDowell, A. E. Wickenden, K. Gramann, Tzyy-Ping Jung, Li-Wei Ko, and Jyh-Yeong Chang, "Biosensor Technologies for Augmented Brain–Computer Interfaces in the Next Decades," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1553–1566, May 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6168212/

Copyright (c) The Government of USA, 2022. Used by permission to IARIA. ISBN: 978-1-61208-982-9

[18] A. Ortiz-Rosario and H. Adeli, “Brain-computer interface technologies: from signal to action,” *Reviews in the Neurosciences*, vol. 24, no. 5, pp. 537–552, Jan. 2013. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/revneuro-2013-0032/html

[19] R. D. Dias, M. C. Ngo-Howard, M. T. Boskovski, M. A. Zenati, and S. J. Yule, “Systematic review of measurement tools to assess surgeons’ intraoperative cognitive workload,” *British Journal of Surgery*, vol. 105, no. 5, pp. 491–501, Mar. 2018. [Online]. Available: https://academic.oup.com/bjs/article/105/5/491-501/6123151

[20] A. M. Hughes, G. M. Hancock, S. L. Marlow, K. Stowers, and E. Salas, “Cardiac Measures of Cognitive Workload: A Meta-Analysis,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 61, no. 3, pp. 393–414, May 2019. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0018720819830553

[21] A. R. Hidalgo-Muñoz, D. Mouratille, N. Matton, M. Causse, Y. Rouillard, and R. El-Yagoubi, “Cardiovascular correlates of emotional state, cognitive workload and time-on-task effect during a realistic flight simulation,” *International Journal of Psychophysiology*, vol. 128, pp. 62–69, Jun. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167876017306542

[22] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, “The PREP pipeline: standardized preprocessing for large-scale EEG analysis,” *Frontiers in Neuroinformatics*, vol. 9, Jun. 2015. [Online]. Available: http://journal.frontiersin.org/Article/10.3389/fninf.2015.00016/abstract

[23] E. E. Flynn-Evans, L. R. Wong, Y. Kuriyagawa, N. Gowda, P. F. Cravalho, S. Pradhan, N. H. Feick, N. G. Bathurst, Z. L. Glaros, T. Wilaiprasitporn, K. Bansal, J. O. Garcia, and C. J. Hilditch, “Supervision of a self-driving vehicle unmasks latent sleepiness relative to manually controlled driving,” *Scientific Reports*, vol. 11, no. 1, p. 18530, Dec. 2021. [Online]. Available: https://www.nature.com/articles/s41598-021-92914-5

[24] M. Alavash, C. C. Hilgetag, C. M. Thiel, and C. Gießing, “Persistency and flexibility of complex brain networks underlie dual-task interference,” *Human Brain Mapping*, vol. 36, no. 9, pp. 3542–3562, Sep. 2015.

[25] V. Paban, J. Modolo, A. Mheich, and M. Hassan, “Psychological resilience correlates with EEG source-space brain network flexibility,” *Network Neuroscience*, vol. 3, no. 2, pp. 539–550, Jan. 2019. [Online]. Available: https://direct.mit.edu/netn/article/3/2/539-550/2214

[26] S. H. Kwon, “Threshold selection based on cluster analysis,” *Pattern Recognition Letters*, vol. 25, no. 9, pp. 1045–1050, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865504000674

[27] P. Braarud, “Subjective Task Complexity and Subjective Workload: Criterion Validity for Complex Team Tasks,” *International Journal of Cognitive Ergonomics*, vol. 5, no. 3, pp. 261–273, Sep. 2001. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1207/S15327566IJCE0503_7

Copyright (c) The Government of USA, 2022. Used by permission to IARIA. ISBN: 978-1-61208-982-9