Assessment of Multiple Systemic Human Cognitive States using Pupillometry

Ayca Aygun (ayca.aygun@tufts.edu)
Department of Computer Science, Tufts University,
420 Joyce Cummings Center 177 College Avenue, Medford, MA, 02155

Thuan Nguyen (tnguyen8@wpi.edu)
Department of Electrical & Computer Engineering, Worcester Polytechnic Institute,
100 Institute Road, WPI - Atwater Kent Laboratories, Worcester, MA 01609

Matthias Scheutz (matthias.scheutz@tufts.edu)
Department of Computer Science, Tufts University,
420 Joyce Cummings Center 177 College Avenue, Medford, MA, 02155

Abstract

How to best and robustly detect human systemic cognitive states like workload, sense of urgency, mind wandering, interference, and others is still an open question as the answer essentially depends both on the employed physiological measurements as well as the trained computational classification models. In this paper, we analyze data from a human driving experiment to explore the validity of eye gaze in assessing different systemic cognitive states and relations among them. Our statistical analyses and classification results indicate that eye gaze, in particular the percentage change in pupil size (PCPS), is a reliable physiological biomarker in assessing multiple systemic human cognitive states including workload, sense of urgency (SoU), and mind wandering (MW) while it does not seem suitable to detect task interference (which can be assessed based on participant's response times).

Keywords: Cognitive workload, mind wandering, sense of urgency, interference, pupillometry, computational models

1 Introduction

Answering the question of how to best detect systemic human cognitive states like workload, sense of urgency, mind wandering, and interference is not only of interest to cognitive psychology, but also to a wide range of application domains where these states can impact and modulate human task performance. Workload to a measure of cognitive effort based on the relationship between task requirements, circumstantial situations under which the task is accomplished, and the abilities, the attitude, and the concentration of the individual who carries out the task (Hart & Staveland, 1988). Sense of urgency refers to an individual's response to urgent task demands that usually have to be completed or attended to within a short period of time. Mind wandering refers to spontaneous thought that is less-intentionally restricted than creative thinking and objective-directed thought but more-intentionally restricted than dreaming (Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016). Interference is another cognitive state which refers to a consequence of accomplishing two or more concurrent tasks which relates to concentration-based disadvantages of human nature (Pashler, 1994). Overly high workload can lead to errors and drop in performance, and interference and sense of urgency can worsen it. At the same time, mind wandering can lead to oversight and missed deadlines. Especially, in high-stress domains like air traffic control or disaster response, it would be important to be able to track such systemic cognitive states to be able to design interventions that prevent escalation and thus task failure.

Yet, despite significant prior work on which physiological measures to use and how to process the information, there is currently still no consensus on the most reliable physiological signals and computational methods for categorizing them. In part the problem is due to the fact that many of the prior studies and results did not aim at understanding a set of systemic human cognitive states; rather, most studies attempt to classify one systemic cognitive state (very often cognitive workload). It is not only important to assess these individual cognitive states but also it is significant to comprehend the association between different pairs of systemic human cognitive states given that systemic cognitive states do not happen in isolation but can impact each other.

The goal of this paper is thus three-fold:

1. We examined the relationship between different pairs of systemic human cognitive states and discussed how two individual human cognitive states interact with each other.
2. We showed that percentage change in pupil size (PCPS) is a reliable physiological marker in determining workload, mind wandering, and sense of urgency.
3. We further investigated that interference between tasks causes a significantly higher response, but without PCPS reflecting it.

2 Related Works

Cognitive workload estimation studies. There have been multiple research efforts which examine eye gaze features as a predictor of *cognitive workload* (Chen & Epps, 2013; Halverson, Estepp, Christensen, & Monnin, 2012; Wang, Li, Wang, & Chen, 2013; Tokuda, Obinata, Palmer, & Chaparro, 2011; Khedher, Jraidi, & Frasson, 2019). Tokuda *et al.* explored the performance of both saccadic intrusion, which is a particular type of fixational eye movement, and pupil diameter in assessing cognitive workload (Tokuda et al., 2011). Although the authors indicated that pupil diameter has better accuracy in determining workload compared to saccadic intrusion, they did not consider the subject-based variations of pupil diameter, while we aimed to remove subject-based discrepancy by using the PCPS rather than the pupil diameter itself. Khedher *et al.* acquired eye gaze and EEG signals from fifteen participant in a virtual learning environment to classify two groups: the students who were capable of completing the tasks successfully and the students who were not (Khedher et al., 2019). They found out that k-Nearest Neighbor is the best predictor for assessing cognitive workload over other classification methods. In contrast, we utilized neural network-based deep learning models in addition to common classification methods to evaluate the efficiency of eye gaze in estimating workload.

Sense of urgency estimation studies. There have been relatively few studies which investigate the performance of different physiological signals to determine *sense of urgency* (Shalom, Dagnino, & Sigman, 2011; Šalkevicius, Damaševičius, Maskeliunas, & Laukienė, 2019; Khalaf et al., 2020). Salkevicius *et al.* studied to predict the anxiety levels of individuals with blood volume pressure, galvanic skin response, and skin temperature (Šalkevicius et al., 2019). Another study proposed a method to assess participants' challenge and threat states while proceeding three mental arithmetic tasks by leveraging multiple physiological signal types (Khalaf et al., 2020). To the best of our knowledge, there is no research which considers human gaze to determine sense of urgency.

Mind wandering estimation studies. There are a few research papers that explore the efficiency of human eye gaze in assessing *mind wandering* (Klesel, Schlechtinger, Oschinsky, Conrad, & Niehaves, 2020; He, Becic, Lee, & McCarley, 2011; Bixler, Blanchard, Garrison, & D'Mello, 2015; Singha, 2021). One study proposed that the participants show few deficits in manipulating vehicles during mind wandering occurrence (He et al., 2011). Although exploring different eye gaze parameters in assessing mind wandering, their dataset has insufficient number of participants which is a potential cause for overfitting. Another study explored the fusion of eye gaze and physiological features to estimate mind wandering while the participants were reading instructional texts. Even though this study investigated mind wandering during a driving task, their simulation platform was too sophisticated for evaluating mind wandering since the participants must perform too many tasks simultaneously such as watching the traffic lights, adjusting the speed, and paying attention to the other vehicles and pedestrians (Bixler et al., 2015). While several studies examine mind wandering using eye gaze in different domains, there is a lack of studies focussing on assessing mind wandering in the context of driving simulation settings.

Interference estimation studies. There are also a few efforts to predict *cognitive interference* using different physiological indicators. One study used EEG and electrooculogram (EOG) to explore the impact of EEG theta power for assessing cognitive interference and demonstrated the performance of the theta power in assessing interference circumstances (Nigbur, Ivanova, & Stürmer, 2011). Another study investigated the relationship between social pertinence of eye gaze direction and spatial interference, and found out an opposite congruency effect between fixation location and eye gaze direction (Cañadas & Lupiáñez, 2012).

3 Experimental Setup

We used the data from a comprehensive interactive multimodal multi-task driving experiment. In particular, the partic-

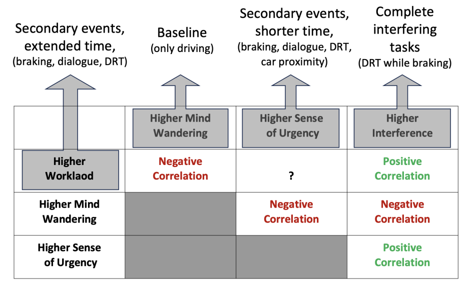


Figure 1: The illustration of 1) the association between different pairs of systemic human cognitive states, and 2) the evaluation criteria of individual cognitive states.

ipants were instructed to drive accident-free while completing multiple secondary tasks. During the first three minutes of the simulation, the only task was driving which helped the participants acclimate the rest of the simulation. Next, participants were instructed to perform three additional tasks. First, participants engaged in brief dialogue communications multiple times during the simulation. There are 40 questions in total (20 questions per session) which included a series of "yes/no" and explanation dialogue occurred every 30-60 seconds during each session. Second, participants completed braking events happened ten times per session. Third, participants had to perform DRT tasks. We attached a cylindrical vibrotactile motor, which was 14 mm in diameter and 4.5 mm thick, to the participants' right collar bone/shoulder. Then, we fixed a response button to their right index fingertip with hook and loop tape. Finally, we instructed the participants to respond to tactile stimuli that took place randomly every 6 to 10 seconds. The experiment included two separate sessions: DRT (included DRT occurrences) and non-DRT. While multiple physiological signals including eye gaze, electroencephalography (EEG), arterial blood pressure, respiration, functional near-infrared spectroscopy (fNIRS), and skin conductance were collected from the experiment, we here focus on eye gaze to demonstrate its effectiveness in assessing different systemic human cognitive states. The details of the driving simulation can be found in (Scheutz et al., 2023).

4 Methods

We start by introducing definitions of the systemic human cognitive states we investigated and how to assess their instantiation and levels in a simulated driving environment:

Workload has been considered as a human-centered paradigm that emerges from the association between task specifications, the environmental circumstances under which the task is completed, and the capabilities, the temperament, and the attention of the individual who proceeds the task (Hart & Staveland, 1988). Cain *et al.* linked the workload with the reflection of mental demands imposed on human operators by the task requirements (Cain, 2007). Specifically, workload was interpreted as a multidimensional concept, encountered by a human operator, which is controlled by the capabilities and the efforts of the operator. Moreover, cognitive workload was defined as a circumstance which must be diminished or regulated at an engaging level to attain a user satisfaction in a

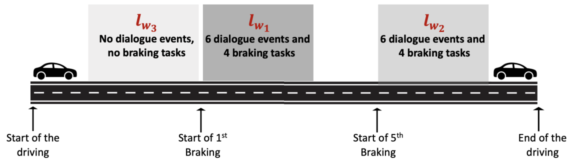


Figure 2: The generation of three workload levels.

human-computer interaction (HCI) setup (Kosch et al., 2023). A higher workload can be generated by integrating secondary task assignments into the primary work (see Figure 1.

Sense of Urgency can be described as a perception that an assigned task must be completed immediately, usually within a specific period of time, which impacts decision-making mechanisms in humans. Yau *et al.* interpreted time-varying and evidence-free signals as a sense of urgency, and discovered that it is likely to determine EEG signals which are associated with sense of urgency (Yau et al., 2021). Another study examined the effects of warning signals, including auditory and vibrotactile, in predicting sense of urgency in different cognitive workload conditions (Biondi, Leo, Gastaldi, Rossi, & Mulatti, 2017). Different sense of urgency levels can be generated by taking shorter time frames (compared to workload case) in which the participants need to accomplish secondary events quickly such as completing braking event or responding dialogue interactions immediately.

Mind wandering was considered as a specific type of spontaneous thought, a cognitive state or a sequence of multiple cognitive states that progress relatively freely due to a lack of intense restrictions on the subject matter of any cognitive state or on the shift between any of the two consecutive cognitive states (Christoff et al., 2016). In particular, mind wandering was described as a special type of spontaneous thought that is less-purposely restricted than creative thought and objective-directed thinking but more-purposely restricted than dreaming. Another study ((Smallwood & Schooler, 2015)) linked mind wandering with the self-determination of experiences from impressions and ongoing activities, and emphasized the terms “task unrelated” (Giambra, 1989) and “stimulus independent” (Antrobus, Singer, Greenberg, et al., 1966) which represent the thoughts happen during mind-wandering episodes.

Interference was considered as an outcome of performing two or more simultaneous tasks which associates the concentration-based limitations of human nature (Pashler, 1994). In (Sarason, Pierce, & Sarason, 2014), authors considered interference as referring to unwanted and undesirable thoughts which eventually has detrimental impacts on task performance. In this study, we generated higher interference including secondary tasks in addition to the ongoing task which should be responded timely manner (See Section 4.1).

Moreover, Figure 1 illustrates associations between different pairs of systemic human cognitive states.

Workload vs. Mind Wandering: Zhang *et al.* claimed that there is a negative correlation between cognitive load and mind wandering during driving. In particular, the participants who have decreased cognitive workload have extra mental capacity which may be assigned to other tasks irrelevant to the primary task and cause higher mind wandering (Zhang & Kumada, 2017). Another study claimed that any individual has limited mental capacity, and thus, raised perceptual load would cause to decreased available mental capacity for task-irrelevant activities (Lavie, 2010). This is consistent with our hypothesis that higher mind wandering is likely to diminish with increased cognitive workload.

Workload vs. Sense of Urgency: In contrast to the association between workload and mind wandering, the link between workload and sense of urgency is not entirely comprehensible. Even though there have been research studies which explored the correlation between workload and sense of urgency (Wei, Gong, & Wu, 2022; Keunecke et al., 2019), those research efforts were mainly focused on subjective ratings and did not provide moment-by-moment analyses of the effect of sense of urgency happenings on overall workload. In this study, we aimed to explore the impacts of consecutive immediate task occurrences on overall workload. Intuitively, we may consider three situations to examine how workload is linked to sense of urgency. First, there may be sporadic sense of urgency moments which might not impact overall task-dependent workload. Second, we may consider the circumstances where tasks trigger sense of urgency, *e.g.*, such as braking events in a driving task, occur frequently and would hence potentially lead to increased overall workload as well. Third, even if we monitor higher workload after frequent sense of urgency happenings, we cannot conclusively claim that the increased overall workload is solely an outcome of the task-dependent frequent sense of urgency occurrences as other task-irrelevant activities might have contributed to the overall workload as well.

Workload vs. Interference: Forsyth *et al.* examined how consecutive interference occurrences impact overall workload and found out that both objective and subjective workload measurements are prone to increase with cognitive interference (Forsyth et al., 2018). Another study investigated the effect of social media interference on mental workload while performing six simulated computer tasks and indicated that there is a positive correlation between interference, caused by unpredictable interruption occurrences, and overall cognitive load (Zahmat Doost & Zhang, 2023). These research efforts are consistent with our hypothesis that increased interference is prone to reinforce overall mental load.

Mind Wandering vs. Sense of Urgency: Latinjak *et al.* examined the link between mind wandering and goal-assisted, spontaneous thinking (Latinjak, 2018). Specifically, the author(s) modeled a competition task performed by athletes, to investigate the association between spontaneous thoughts and task-unrelated mental activity, and found out that there is a negative correlation between goal-assisted thinking and mind wandering. Considering that, by nature, mind wandering develops when individual’s attention instantly switches from the ongoing task to inner thoughts (Girardeau et al., 2023), it is acceptable to hypothesize that increased mind wandering is oppositely correlated with higher sense of urgency which is compatible with what was proposed in (Latinjak, 2018).

Mind Wandering vs. Interference: Considering that increased interference, caused by consecutive interruption happenings, is likely to increase overall workload and mind wandering is prone to decrease with increased workload, we can hypothesize that higher mind wandering is negatively correlated with increased interference. Our hypothesis is consistent with the research study which indicated that increased task-unrelated-thoughts (TUT) are linked to decreased cortical processing of task-relevant events and interrupting stimulation (Barron, Riby, Greer, & Smallwood, 2011).

Sense of Urgency vs. Interference: Considering that, in nature, sense of urgency happenings, simultaneously occur with the ongoing task, require the individuals to interfere multiple tasks at the same time, we hypothesize that urgency moments, which include secondary event happenings such as braking tasks or dialogue communications, which are required to be completed timely manner, are likely to reinforce interference. Future efforts will include detailed examination of the association between sense of urgency occurrences and interference.

4.1 Cognitive State Evaluation

Workload Evaluation: Figure 2 illustrates how we distinguished three potential cognitive state levels labelled $l_{w_1}, l_{w_2}, l_{w_3}$ based on the simultaneous activities subjects had to perform. l_{w_1} and l_{w_2} started with the 1st and the 5th braking events, respectively, and included four braking and six consecutive communication events each. We hypothesized that l_{w_1} will be higher than l_{w_2} given that the participants were required to acclimate with different types of secondary events including braking tasks and dialogue communications during l_{w_1} timeline and they were used to proceed the secondary tasks during l_{w_2} timeline. l_{w_3} was the initial period of the driving simulation before the start of the first braking event. We produced a balanced dataset having a total of 276 samples (92 samples per cognitive workload level) taken from both of the sessions of 46 participants.

Sense of Urgency Evaluation: We also considered two potential sense of urgency levels: the first 2.5 seconds of the 1st braking event labelled l_{sou_2}, and the 2.5 seconds time frame 30 seconds before the 1st braking event labelled l_{sou_1} (a period that does not include any events other than the primary driving task). The intuition behind this is that there are circumstantial conditions or task demands which require the participants to respond instantly and hence produce a higher sense of urgency such as responding to a dialogue question in time or suddenly braking during driving.

Mind Wandering Evaluation: Most studies on mind wandering have been designed to assess mind wandering levels of participants subjectively by asking them to press buttons and indicate mind wandering during the experiment (E. Bixler & K. D'Mello, 2021; Beninger, Hamilton-Wright, Walker, & Trick, 2021; Mills, Gregg, Bixler, & D'Mello, 2021; Caruso & D'Mello, 2023; Zhang & Kumada, 2017; Geden, Staicu, & Feng, 2018) or after (Berthié et al., 2015). However, our aim was to design a scenario where we can assess participants' mind wandering levels objectively to eliminate subject-based misjudgments. To this end, we investigated the change in mind wandering levels of the participants during the initial course of driving. Specifically, we captured the first and next 1.5 minutes of driving to define potentially lower and higher mind-wandering levels, respectively. The intuition here is that during the first 1.5 minutes, the participants were attempting to acclimate to the driving environment which requires more concentration, resulting in a lower mind wandering labelled l_{mw_1}. During the next 1.5 minutes, the participants were already familiar with the driving task and only had to drive on a straight road with no task requirement (*i.e.*, no communications, no braking demands) other than simply driving on the right side of a straight highway. Thus, less attention should be required which potentially evokes higher level of mind wandering labelled l_{mw_2}.

Table 1: Collection of different classification tasks including task ID, classification type, and cognitive state levels.

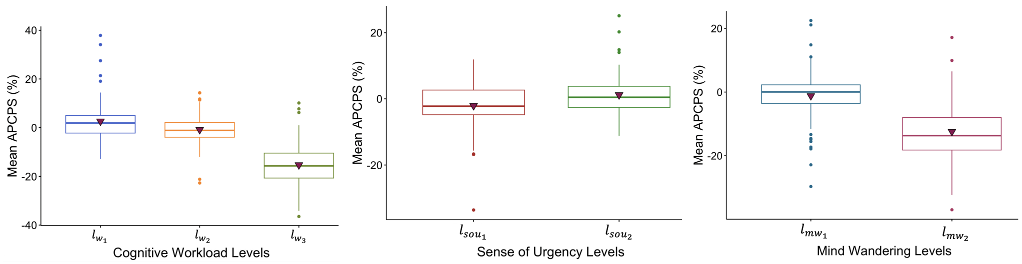
Task ID:	Classification Type:	Levels to be Classified:
W_1	Workload	$l_{w_1}, l_{w_2}, l_{w_3}$
W_2	Workload	l_{w_1}, l_{w_3}
SoU	Sense of Urgency	l_{sou_1}, l_{sou_2}
MW	Mind Wandering	l_{mw_1}, l_{mw_2}
INT	Interference	l_{int_1}, l_{int_2}

Interference Evaluation: To examine the potential change in interference levels, we considered two cases: 1) braking events which include DRT, and 2) braking events which do not include DRT. Specifically, we took the braking occurrences in which the participants were and were not required to push the DRT button to consider potentially higher (l_{int_1}) and lower (l_{int_2}) interference levels, respectively (in particular, we checked the time duration between the onset of the braking event and the moment that the participant changes the angle of the brake pedal to investigate DRT happenings). We identified 66 braking events which included DRT and 314 braking events which did not include DRT collected from 64 participants. Then, we compared the response times for these two cases to determine whether interfering DRT happenings affect the speed of participants' responding to the braking events, or not. Here, response time represents the time duration between the starting point of the braking event and the moment that the participant brakes. Finally, we explored the change in the average PCPS values between l_{int_1} and l_{int_2} by performing t-test, and reported the results in Section 5.

4.2 Pupillometry

Data Processing: We applied a three-step pre-processing technique to a 400 Hz pupillometry signal to remove any out-of-band sensory noise and the blink artifact explained in (Aygun et al., 2022). We used the left pupillometry signal considering that the left and the right pupil diameters are synchronous.

Percentage Change in Pupil Size (PCPS): Pupil diameter varies across individuals based on different conditions including gender, age, or other physiological reasons such as lens



(a) Workload (b) Sense of Urgency (c) Mind Wandering

Figure 3: Boxplot of the mean APCPS for (a) three workload, (b) two sense of urgency, and (c) two mind wandering levels.

thickness (*e.g.*, a smaller pupil diameter was related to higher age and male gender besides A larger pupil diameter was linked to female gender and larger white-to-white distance (Kiel et al., 2022)). Hence, to eliminate the subject-based fluctuations in pupil diameter, we chose to use the percentage change in pupil size (PCPS) and calculated it with the following equation (Zhao, Gao, Wang, Qu, & Chen, 2020): PCPS = 100·(CM-BM)/BM. where CM and BM represent the current measure of pupil diameter and the baseline measure of diameter, respectively. Here, BM is determined by calculating the mean of a 10-second signal before the stimulus.

4.3 Performance Analyses

Statistical Tests: Firstly, we performed an ANOVA statistical test called Tukey's "honestly significant difference" (HSD) multiple pairwise comparison test (Jaccard, Becker, & Wood, 1984) to examine the change in the PCPS for three workload levels (l_{w_1}, l_{w_2}, and l_{w_3}), two sense of urgency levels (l_{sou_1} and l_{sou_2}) , and two mind wandering levels (l_{mw_1} and l_{mw_2}). In addition, we performed Welch's t-test to examine the differentiation of two interference levels in terms of the average PCPS and participants' responding times to braking events. We used the R programming language to perform ANOVA's Tukey's HSD multiple pairwise and Welch's t-tests using RStudio.

Classification Tasks: Secondly, we aimed to quantify the performance of common machine learning (ML) methodologies in assessing different human cognitive states with the PCPS. In particular, we employed five ML methods: (1) *k*-Nearest Neighbor (k-NN), (2) Naive-Bayes (NB), (3) Random Forest (RF), (4) Support Vector Machines (SVM), and (5) Multiple Layer Perceptron (MLP). We summarized the classification tasks employed in this study in Table 1.

For the given classification task and the learning methodology, we divided the data into training and test sets with a ratio of 80% and 20%, *i.e.*, 37 participants for the training set and 9 participants for the test set, respectively. Next, we applied the cross-validation method (Stone, 1978) to the training set to select the best-learned model to apply to the test set. To mitigate the randomness in the training and test processes as well as to stabilize the testing performance, for a given classification task and learning method, we repeated the whole experiment 10 times and only reported the average accuracy together with its standard deviation.

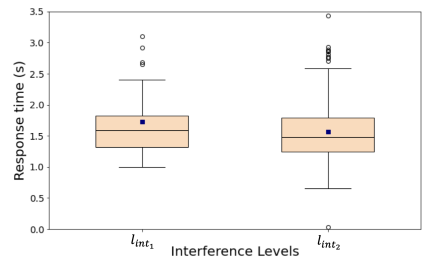


Figure 4: The boxplot of the responding times to braking events for higher (l_{int_1}) and lower interference (l_{int_2}).

5 Results

Figure 3 shows the boxplots of the mean APCPS taken from 46 participants to assess different human cognitive states including (a) workload, (b) sense of urgency, and (c) mind wandering (Here, the APCPS represents the average PCPS across time (Zhao et al., 2020) while the mean APCPS represents the average APCPS across samples obtained from different participants). The results demonstrated that the APCPS is a reliable physiological marker to distinguish multiple levels of different systemic human cognitive states. To investigate the performance of the APCPS in assessing different cognitive states, we further performed the Tukey's HSD multiple pairwise test. Table-3 indicate the results of the p-values corresponding to all pairs of workload, sense of urgency, and mind-wandering levels obtained from Tukey's HSD multiple pairwise comparison test at a significance level of .95. The results demonstrate that there is a statistically significant difference between all pairs of workload levels (p-value <.05). Specifically, the p-values are very small in distinguishing l_{w_1}-l_{w_3} (2.23×10^{-13}) and l_{w_2}-l_{w_3} (2.23×10^{-13}) compared to l_{w_1}-l_{w_2} (7.47×10^{-3}) which indicates that the workload is notably lower during the initial period of driving when there is no secondary event. In addition, the PCPS is a powerful physiological indicator in separating two sense of urgency levels (l_{sou_1}-l_{sou_2}) and two mind wandering levels (l_{mw_1}-l_{mw_2}) with the p-values of 5.25×10^{-4} and 3.60×10^{-14}, respectively.

In addition, we performed Welch's t-test with a significance level of .95 in distinguishing the responding times of the participants for two interference levels with unequal variance condition given that the number of samples are different for two interference levels. The results indicate that the p-value and t-value are .063 (slightly higher than the significance level) and 1.89), respectively, with a degree of freedom

Table 2: Classification accuracies (in %) of the APCPS for different classification tasks using different ML methods.

Signal Modality	Tasks	k-NN	NB	RF	SVM	MLP
Workload	W_1	62.78 ∓ 5.47	**63.35** ∓ 5.63	55.64 ∓ 4.15	**63.35** ∓ 4.43	63.25 ∓ 5.22
	W_2	**90.18** ∓ 3.42	87.91 ∓ 6.04	85.35 ∓ 5.97	88.76 ∓ 4.14	85.71 ∓ 4.41
Sense of Urgency	SoU	57.04 ∓ 6.52	**60.46** ∓ 9.29	54.05 ∓ 5.95	57.04 ∓ 8.41	55.75 ∓ 7.80
Mind Wandering	MW	**82.08** ∓ 7.35	80.51 ∓ 6.54	71.83 ∓ 7.61	77.24 ∓ 5.70	79.94 ∓ 6.98

Table 3: p-values of Tukey HSD multiple pairwise test for the APCPS obtained from different cognitive state levels.

Cognitive State	Levels	p-value
Workload	l_{w_1} - l_{w_2}	7.47×10^{-3}
	l_{w_1} - l_{w_3}	2.23×10^{-13}
	l_{w_2} - l_{w_3}	2.23×10^{-13}
Sense of Urgency	l_{sou_1} - l_{sou_2}	5.25×10^{-4}
Mind Wandering	l_{mw_1} - l_{mw_2}	3.60×10^{-14}

of 78.03. Figure 4 shows the boxplot of the responding times to braking events for DRT (higher interference) and non-DRT (lower interference) cases are taken from 64 participants and demonstrated that participants' responding time is a reliable indicator in assessing different interference levels. Moreover, we performed Student's t-test on the APCPS values to analyze the effect of the APCPS in classifying l_{int_1} and l_{int_2}, and observed that the APCPS is not capable of identifying two interference levels (p-value =.685).

We also run four classification tasks using the APCPS (see Table 1). Table 2 shows the classification accuracies of five algorithms related to four classification tasks W_1, W_2, SoU, and MW. The results indicate that for the task W_1, which aims to distinguish three workload levels (l_{w_1}, l_{w_2}, and l_{w_3}), APCPS has the accuracy of 63.35 ∓ 5.63 for both NB and SVM models. Note that W_1 is three-class classification task, therefore, a random guess will likely provide an accuracy of 33% which is far below the performance achieved by all five tested algorithms on APCPS. On the other hand, for W_2, which aims to classify l_{w_1} and l_{w_3}, APCPS achieves the highest accuracy of 90.18 ∓ 3.42 using the k-NN algorithm which outperforms other ML models. Similarly, we run the five ML algorithms for the classification tasks SoU and MW, which aim to distinguish two sense of urgency levels (l_{sou_1} and l_{sou_2}) and two mind wandering levels (l_{mw_1} and l_{mw_2}), respectively. The results showed that APCPS achieves the highest accuracies of 60.46 ∓ 9.29 using NB and 82.08 ∓ 7.35 using k-NN for the tasks SoU and MW, respectively. The results demonstrate that the APCPS is a reliable indicator in separating multiple levels of different systemic human cognitive states.

6 Discussion

We have studied the utility of eye gaze, in particular the PCPS, in determining different systemic human cognitive states. Our statistical analyses indicated that the PCPS is a reliable physiological biomarker in assessing workload, sense of urgency, and mind wandering. We then validated the efficacy of k-NN, NB, RF, SVM, and MLP in separating different workload, sense of urgency, and mind wandering levels, and demonstrated that the PCPS is capable of achieving three-class workload, two-class workload, two-class sense of urgency, and two-class mind wandering classification.

In contrast to workload, sense of urgency, and mind-wandering estimation, the PCPS is not capable of identifying different interference levels. However, we showed that the response times of the participants can be used to identify two interference levels. Specifically, we observed that interference, which is generated by integrating secondary events to the ongoing task, which must be completed on time, leads to extended responding times. We did not apply ML models to interference cases as we did to observe statistically significant differences (based on t-testw) between two interference levels.

Additionally, we explored the association between different pairs of systemic human cognitive states. To successfully comprehend the impacts of them on human performance, it is essential to understand the corporation of individual human cognitive states considering that systemic cognitive states do not take place in solitude but can affect each other. We further investigate the link between different systemic human cognitive state pairs in detail in our future work.

7 Conclusion

The goal of this study was to explore the association between different pairs of systemic human cognitive states and validate the efficiency of eye gaze, in particular the PCPS, in assessing different human cognitive states including workload, sense of urgency, mind wandering, and interference in a multimodal driving simulation environment. Our statistical analyses along with the classification results demonstrated that the "percent change in pupil size" is a reliable physiological marker for predicting different levels of workload, sense of urgency, and mind wandering. However, we observed that the PCPS is not an efficient physiological indicator in detecting interference, instead, the responding times of the individuals can be used to identify interference occurrences. While we used data from the driving dataset, we believe that this dataset has the potential to generalize to other tasks, and hence, to enhance human performance in many applications.

8 Acknowledgments

This work was in part funded by AFOSR grant FA9550-18-1-0465.

References

Antrobus, J. S., Singer, J. L., Greenberg, S., et al. (1966). Studies in the stream of consciousness: Experimental enhancement and suppression of spontaneous cognitive processes. *Perceptual and Motor Skills*, *23*(2), 399–417.

Aygun, A., Lyu, B., Nguyen, T., Haga, Z., Aeron, S., & Scheutz, M. (2022). Cognitive workload assessment via eye gaze and eeg in an interactive multi-modal driving task. In *Proceedings of the 2022 international conference on multimodal interaction* (pp. 337–348).

Barron, E., Riby, L. M., Greer, J., & Smallwood, J. (2011). Absorbed in thought: The effect of mind wandering on the processing of relevant and irrelevant events. *Psychological science*, *22*(5), 596–601.

Beninger, J., Hamilton-Wright, A., Walker, H. E., & Trick, L. M. (2021). Machine learning techniques to identify mind-wandering and predict hazard response time in fully immersive driving simulation. *Soft Computing*, *25*(2), 1239–1247.

Berthié, G., Lemercier, C., Paubel, P.-V., Cour, M., Fort, A., Galéra, C., … Maury, B. (2015). The restless mind while driving: drivers' thoughts behind the wheel. *Accident Analysis & Prevention*, *76*, 159–165.

Biondi, F., Leo, M., Gastaldi, M., Rossi, R., & Mulatti, C. (2017). How to drive drivers nuts: effect of auditory, vibrotactile, and multimodal warnings on perceived urgency, annoyance, and acceptability. *Transportation research record*, *2663*(1), 34–39.

Bixler, R., Blanchard, N., Garrison, L., & D'Mello, S. (2015). Automatic detection of mind wandering during reading using gaze and physiology. In *Proceedings of the 2015 acm on international conference on multimodal interaction* (pp. 299–306).

Cain, B. (2007). A review of the mental workload literature. *DTIC Document*.

Cañadas, E., & Lupiáñez, J. (2012). Spatial interference between gaze direction and gaze location: A study on the eye contact effect. *Quarterly Journal of Experimental Psychology*, *65*(8), 1586–1598.

Caruso, M., & D'Mello, S. (2023). Do associations between mind wandering and learning from complex texts vary by assessment depth and time? In *Lak23: 13th international learning analytics and knowledge conference* (pp. 230–239).

Chen, S., & Epps, J. (2013). Automatic classification of eye activity for cognitive load measurement with emotion interference. *Computer methods and programs in biomedicine*, *110*(2), 111–124.

Christoff, K., Irving, Z. C., Fox, K. C., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nature reviews neuroscience*, *17*(11), 718–731.

E. Bixler, R., & K. D'Mello, S. (2021). Crossed eyes: Domain adaptation for gaze-based mind wandering models. In *Acm symposium on eye tracking research and applications* (pp. 1–12).

Forsyth, K. L., Hawthorne, H. J., El-Sherif, N., Varghese, R. S., Ernste, V. K., Koenig, J., & Blocker, R. C. (2018). Interruptions experienced by emergency nurses: implications for subjective and objective measures of workload. *Journal of emergency nursing*, *44*(6), 614–623.

Geden, M., Staicu, A.-M., & Feng, J. (2018). The impacts of perceptual load and driving duration on mind wandering in driving. *Transportation research part F: traffic psychology and behaviour*, *57*, 75–83.

Giambra, L. M. (1989). Task-unrelated thought frequency as a function of age: a laboratory study. *Psychology and aging*, *4*(2), 136.

Girardeau, J., Ledru, R., Gaston-Bellegarde, A., Blondé, P., Sperduti, M., & Piolino, P. (2023). The benefits of mind wandering on a naturalistic prospective memory task. *Scientific Reports*, *13*(1), 11432.

Halverson, T., Estepp, J., Christensen, J., & Monnin, J. (2012). Classifying workload with eye movements in a complex task. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 56, pp. 168–172).

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). Elsevier.

He, J., Becic, E., Lee, Y.-C., & McCarley, J. S. (2011). Mind wandering behind the wheel: Performance and oculomotor correlates. *Human factors*, *53*(1), 13–21.

Jaccard, J., Becker, M. A., & Wood, G. (1984). Pairwise multiple comparison procedures: A review. *Psychological Bulletin*, *96*(3), 589.

Keunecke, J. G., Gall, C., Birkholz, T., Moritz, A., Eiche, C., & Prottengeier, J. (2019). Workload and influencing factors in non-emergency medical transfers: a multiple linear regression analysis of a cross-sectional questionnaire study. *BMC health services research*, *19*(1), 1–11.

Khalaf, A., Nabian, M., Fan, M., Yin, Y., Wormwood, J., Siegel, E., … others (2020). Analysis of multimodal physiological signals within and between individuals to predict psychological challenge vs. threat. *Expert Systems with Applications*, *140*, 112890.

Khedher, A. B., Jraidi, I., & Frasson, C. (2019). Predicting learners' performance using eeg and eye tracking features. In *The thirty-second international flairs conference.*

Kiel, M., Grabitz, S. D., Hopf, S., Koeck, T., Wild, P. S., Schmidtmann, I., … others (2022). Distribution of pupil size and associated factors: Results from the population-based gutenberg health study. *Journal of Ophthalmology*, *2022*.

Klesel, M., Schlechtinger, M., Oschinsky, F. M., Conrad, C., & Niehaves, B. (2020). Detecting mind wandering episodes in virtual realities using eye tracking. In *Information systems and neuroscience: Neurois retreat 2020* (pp. 163–171).

Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., & Woźniak, P. W. (2023). A survey on measuring cognitive workload in human-computer interaction. *ACM Computing Surveys*.

Latinjak, A. T. (2018). Goal-directed, spontaneous, and stimulus-independent thoughts and mindwandering in a competitive context. *The Sport Psychologist*, *32*(1), 51–59.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current directions in psychological science*, *19*(3), 143–148.

Mills, C., Gregg, J., Bixler, R., & D'Mello, S. K. (2021). Eye-mind reader: An intelligent reading interface that promotes long-term comprehension by detecting and responding to mind wandering. *Human–Computer Interaction*, *36*(4), 306–332.

Nigbur, R., Ivanova, G., & Stürmer, B. (2011). Theta power as a marker for cognitive interference. *Clinical Neurophysiology*, *122*(11), 2185–2194.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, *116*(2), 220.

Šalkevicius, J., Damaševičius, R., Maskeliunas, R., & Laukienė, I. (2019). Anxiety level recognition for virtual reality therapy system using physiological signals. *Electronics*, *8*(9), 1039.

Sarason, I. G., Pierce, G. R., & Sarason, B. R. (2014). Domains of cognitive interference. In *Cognitive interference* (pp. 139–152). Routledge.

Scheutz, M., Aeron, S., Aygun, A., de Ruiter, J., Fantini, S., Fernandez, C., ... Lyu, B. (2023). Estimating systemic cognitive states from a mixture of physiological and brain signals. *Topics in Cognitive Science*.

Shalom, D. E., Dagnino, B., & Sigman, M. (2011). Looking at breakout: Urgency and predictability direct eye events. *Vision research*, *51*(11), 1262–1272.

Singha, S. (2021). *Gaze based mind wandering detection using deep learning*. Unpublished doctoral dissertation, Texas A&M University-Commerce.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual review of psychology*, *66*, 487–518.

Stone, M. (1978). Cross-validation: A review. *Statistics: A Journal of Theoretical and Applied Statistics*, *9*(1), 127–139.

Tokuda, S., Obinata, G., Palmer, E., & Chaparro, A. (2011). Estimation of mental workload using saccadic eye movements in a free-viewing task. In *2011 annual international conference of the ieee engineering in medicine and biology society* (pp. 4523–4529).

Wang, W., Li, Z., Wang, Y., & Chen, F. (2013). Indexing cognitive workload based on pupillary response under luminance and emotional changes. In *Proceedings of the 2013 international conference on intelligent user interfaces* (pp. 247–256).

Wei, D., Gong, H., & Wu, X. (2022). Residents' subjective mental workload during computerized prescription entry. *Informatics for Health and Social Care*, *47*(3), 283–294.

Yau, Y., Hinault, T., Taylor, M., Cisek, P., Fellows, L. K., & Dagher, A. (2021). Evidence and urgency related eeg signals during dynamic decision-making in humans. *Journal of Neuroscience*, *41*(26), 5711–5722.

Zahmat Doost, E., & Zhang, W. (2023). Mental workload variations during different cognitive office tasks with social media interruptions. *Ergonomics*, *66*(5), 592–608.

Zhang, Y., & Kumada, T. (2017). Relationship between workload and mind-wandering in simulated driving. *PloS one*, *12*(5), e0176962.

Zhao, M., Gao, H., Wang, W., Qu, J., & Chen, L. (2020). Study on the identification of irritability emotion based on the percentage change in pupil size. In *Proceedings of the 2020 2nd international conference on image, video and signal processing* (pp. 20–24).