Industrial Cognitive Ergonomics and Engineering Psychology, Vol. 35, 2022, 30–37
https://doi.org/10.54941/ahfe1001598

Pupil Dilation Variability as an Indicator of Arousal Regulation Over Time: Towards Understanding Operator Functional State

Brittany N. Neilson[1], Shannon P. Devlin[2], Sabrina M. Drollinger[3], Noelle L. Brown[2], Cíara Sibley[2], Cyrus K. Foroughi[2], and Joseph T. Coyne[2]

[1]Strategic Analysis, Inc. Arlington, VA 22203, USA
[2]Naval Research Lab Washington, D.C. 20375, USA
[3]Naval Aerospace Medical Institute Pensacola, FL 32508, USA

ABSTRACT

The pupil dilation of fifteen air traffic control students was recorded as they completed the mental counters working memory task. Standard deviation of pupil dilation for each trial (32 total trials) was computed for each individual and modeled as a growth curve. Pupil dilation variability fluctuated over time in a nonmonotonic manner. Interestingly, the magnitude and direction of pupil dilation variability differed across individuals, suggesting individual differences in arousal regulation. Performance measures of mental counters (i.e., accuracy and response time) were added as predictors to the growth curve model. Higher accuracy was associated with lower pupil dilation variability in general, suggesting better arousal regulation. Longer response times were associated with a greater fluctuation in pupil dilation variability, suggesting longer responses are associated with larger dysregulation of arousal. These findings are important to consider when developing real-time indicators of an operator's functional state.

Keywords: Aerospace systems, Psychophysiology, Neuroergonomics

INTRODUCTION

Air traffic control (ATC) operators play a critical role in aviation safety. Despite the strict selection and training processes, skilled and highly motivated operators still experience performance breakdowns that lead to endangering situations. Of interest among human factors professionals within aviation is to better identify the functional state of the operator, either in real time or offline (Hockey 2003). Hockey (1997, 2011) proposed a compensatory control model in which an individual can compensate for higher task loads (e.g., an unexpected event) by increasing their regulatory control and effort to maintain performance. Thus, the ability to self-regulate may be an important aspect of maintaining performance and preventing errors.

Recent research has suggested that variability in pupil dilation can indicate general arousal regulation, such that higher variability reflects arousal dysregulation (Robison & Brewer 2021). Further, this same line of research presented evidence that individual differences in working memory capacity were related to arousal regulation, as indicated by pupil dilation variability (Robison & Brewer 2020). The researchers hypothesize that differences in working memory capacity may be due to breakdowns in the locus coeruleus-norepinephrine (LC-NE) system, which is thought to be responsible for arousal regulation and attention control (Unsworth & Robison 2017). Such a breakdown would result in attentional lapses, which is a prominent feature of those with lower working memory capacity. Pupil diameter has served as an indicator of locus coeruleus activity (representing arousal) and has high temporal resolution. Thus, fluctuations in pupil dilation are thought to represent dysregulation in the LC-NE system, thereby producing arousal dysregulation. Importantly, there is a gap in this aforementioned research that we aim to address: Arousal regulation has not been studied at a trial-by-trial level as a means to understand changes in regulation during a task requiring working memory.

The present research explored if variation in pupil dilation during the working memory assessment task called mental counters was informative of overall performance differences, which may ultimately aid in better understanding operator functional state. Mental counters was used for its practical application. It was developed for military selection as a measure of working memory capacity and evaluated as part of the Enhanced Computer Administered Test (ECAT) project (Alderton et al. 1997) to improve the selection test used by the US Department of Defense to select enlisted personnel, the Armed Services Vocational Aptitude Battery (ASVAB). During the ECAT project, mental counters added incremental validity for ATC students specifically (Held & Wolf 1997). Further, mental counters is under evaluation by the US Navy to determine whether it can improve selection and classification across a range of specialties. It should be noted that mental counters was not used as an individual differences metric like previous research, but rather as a tool to impose task load that required attention control.

METHOD

Fifteen air traffic control students at the Naval Air Station in Pensacola, FL participated in the study. All were men with ages ranging from 18 to 29 years ($M = 22.64$ years, $SD = 3.82$ years). Participants completed 32 trials of the mental counters task on the computer. Each trial involved participants updating three values retained in memory. Specifically, participants were instructed to add or subtract these values by 1 based upon corresponding stimulus presentation. After the stimulus presentation, participants responded by entering the 3 updated values using the keyboard. Overall performance across the 32 trials was quantified as total proportion correct (i.e., accuracy) and average response time in seconds (i.e., speed). Eye tracking was recorded during the task using the Gazepoint HD GP3 high resolution eye tracking system. Poor quality data was identified by Gazepoint's algorithm and a difference greater than 25% between the right and left pupil diameter; specifically, 27.3% of trials met these criteria and were excluded from the analysis. Since the recordings were sampled at either 60Hz or 150Hz, all data was down-sampled to 60Hz sampling rate. Down-sampling was performed in R using the *pupillometry* package (Tsukahara 2022). Pupil dilation variability was operationally defined as each participant's standard deviation (SD) of the right pupil's dilation during the stimulus presentation of each trial.

RESULTS

We predicted that poorer task performance would be related to greater standard deviation of pupil dilation, given the assumption that higher variability serves as an indicator of arousal dysregulation associated with task disengagement. The way in which this relationship presents itself at the trial-by-trial level is exploratory; prior research has indicated an increase in variability with time-on-task, but it is unclear if this is generalizable to the mental counters task. Growth curve modeling was used to assess the relationship between mental counters performance metrics (total proportion correct and response time) and within-person fluctuations in pupil dilation (i.e., SD of pupil dilation) across each trial of mental counters. We assessed this using two steps: 1) Modeled the *unconditional growth curve* for pupil dilation variability, which is defined as the best fitting growth curve for which trial is the only predictor in the model, and 2) Added the performance metrics as time-invariant predictors to the model to get *conditional growth curves* for pupil dilation variability. For both steps, multilevel modeling was used to build the growth curve models and consisted of two types of effects: *Fixed effects* are parameter estimates that are constant across individuals. Specifically, the fixed effects estimated the SD of pupil dilation over the course of mental counters. *Random effects* captured the variance associated with a given fixed effect. If the variance is significant (i.e., statistically greater than zero), it would estimate a specific value of the fixed effect for each individual participant. In this work, random effects were implemented to assess the presence of individual differences in the trajectory of pupil dilation variability over time (Mirman 2014).

Unconditional Growth Model: Trial as the Sole Predictor

The general model fitting process for the unconditional growth curve model followed the guidance of Barr et al. (2013) and Matuscheck et al. (2017), in that the most maximal model feasible was prioritized but compared against models with simpler random effects structures in order to balance power and Type I error rates. Pupil dilation variability was modeled as a general linear multilevel model and built with the *lme4* and *lmerTest* package in R (Bates et al. 2015; Pinheiro et al. 2021). In the interest of brevity, the likelihood ratio test results from the best fitting unconditional growth curve model are discussed. The most maximal, nonsingular model for the SD of pupil dilation was a fixed quartic, random quadratic time model

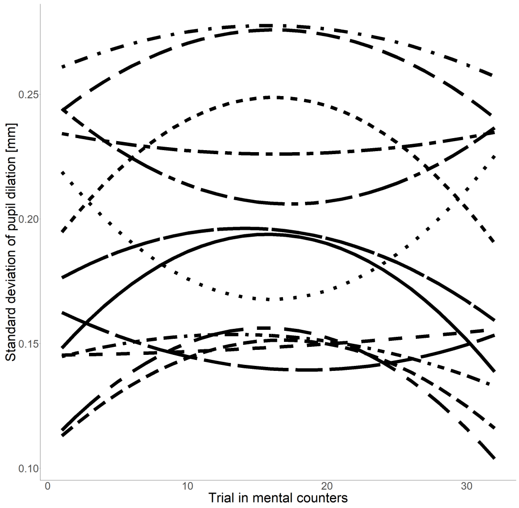

Figure 1: The estimated unconditional growth curve model of pupil dilation variability during the 32 trials of mental counters. Each line represents individual participants.

(BIC = -961.3). However, the best fitting unconditional growth curve model for response time was a random quadratic time model with no random correlations. This generally suggests the trajectory of pupil dilation variability varied in magnitude and shape across all participants, but it did not monotonically change over the course of mental counters. The intercept was a significant random effect ($\tau^2_{U_0}$=0.002), meaning pupil dilation variability was dependent on the participant and ranged from [-0.048-0.086]. The fixed linear time slope (γ_{10}=-0.009, SE=0.018, p=0.61) was not significant, suggesting SD of pupil dilation was not estimated to significantly change across trials. However, it was a significant random effect (τ^2_1=0.006), suggesting the change in SD of pupil dilation significantly depended on the participant and ranged from [-0.020-0.027]. Additionally, the fixed quadratic time slope (γ_{20}= -0.022, SE=0.027, p=0.43) was not significant, suggesting the trajectory of pupil variability was not estimated to significantly change directions across trials. However, there was a significant random effect (τ^2_1=0.003), suggesting the steepness of change in SD of pupil dilation significantly depended on the participant and ranged from [-0.079-0.120]. Taken together, the unconditional growth curve model estimated the magnitude and trajectory of pupil dilation variability to depend on the participant. Specifically, the magnitude of variability differed greatly amongst participants, as some had minimal changes in their pupil size across trials while others changed largely across all trials. Interestingly, the fluctuation in pupil dilation variability was not monotonic, as prior literature had suggested variability increased with time-on-task. Some participants variability initially decreased, while others initially increased, and all at differing rates. However, the model estimated all participants would return to their initial variability levels by the end of mental counters. Figure 1 illustrates these trajectories.

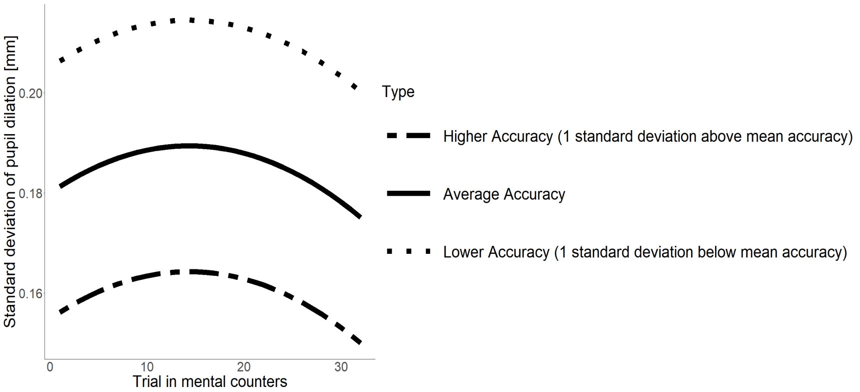

Figure 2: The estimated conditional growth curve model of pupil dilation variability during the 32 trials of mental counters at average accuracy and $\pm$ 1 SD of average accuracy. The model indicates mean SD of pupil dilation was significantly related to accuracy.

Conditional Growth Models: Mental Counters Performance and Trial as Predictors

After establishing the unconditional growth curve model, the significance of time-invariant predictors was explored (i.e., various *conditional growth curve models*). A time-invariant predictor is a measure of the individual that is not expected to change over time or is only reliably measured once throughout the study (Hoffman 2015). In the present research, the mental counters performance metrics were modeled as time-invariant predictors because they were only calculated once per person (i.e., total proportion correct and average response time across all 32 trials). Exploring the predictive ability of these performance measures consisted of using the method presented by Hoffman (2021) and focused on determining the bivariate relationship between each performance metric for each parameter in the unconditional growth curve model (e.g., intercept, linear time slope, etc.). All results are briefly mentioned, but the details from the final conditional growth curve model for each performance metric is presented in detail and interpreted.

Accuracy was a significant predictor of the growth curve model's intercept ($\chi^2(1) = 4.33$, $p < 0.037$, BIC = -983.9) but not a significant predictor of the linear ($\chi^2(1) = 0.33$, $p = 0.85$, BIC = -978.1) or quadratic time slope ($\chi^2(2) = 0.54$, $p = 0.76$, BIC = -972.8). Pragmatically, this means accuracy during mental counters was significantly related to the SD of pupil dilation averaged across all trials, such that higher than average accuracy was associated with smaller mean pupil dilation variability, and lower than average accuracy was associated with larger mean variability. Figure 2 shows the conditional growth curve model when accuracy is included as a time-invariant predictor.

Average response time during mental counters was not a significant predictor of the intercept ($\chi^2(1) = 3.00$, $p = 0.084$, BIC = -982.6) nor was it a significant predictor of the linear time slope ($\chi^2(2) = 3.52$, $p = 0.17$, BIC = -977.3). However, it was a significant predictor of the quadratic time slope ($\chi^2(3) = 8.49$, $p = 0.037$, BIC = -976.4). This suggests response time

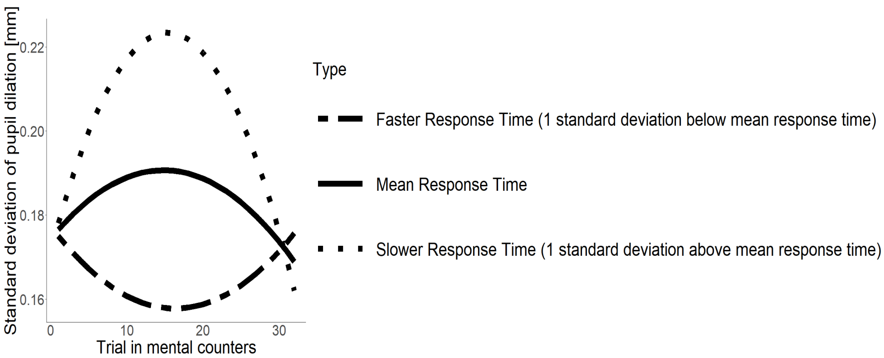

Figure 3: The estimated conditional growth curve model of pupil dilation variability during the 32 trials of mental counters at average response time and ± 1 SD of average response time. The model indicates that over the trials, the SD of pupil dilation significantly changed in magnitude and direction as a function of response time.

during mental counters was indicative of the rate and direction at which pupil dilation variability fluctuated over time. Specifically, faster than average response time was associated with a shallow decline in variability that then increased to initial variability. Slower than average response time was associated with the opposite, more dramatic effect, such that there was a steep increase in variability that then declined to initial variability. Figure 3 shows the conditional growth curve when response time is included as a time-invariant predictor.

CONCLUSION

The unconditional growth curve model highlighted individual differences in the way pupil dilation fluctuates over time. Differences among individuals would not have been identified if measured at the aggregate level, which further supports modeling variability in pupil dilation longitudinally. The only commonality estimated across participants was their return to their original arousal regulatory level by the end of the task. This finding was unexpected, as we assumed variability may increase monotonically over time as a function of time-on-task. One explanation may be that a trial ledger present during mental counters informed participants what trial they were on (e.g., it would show “trial 3 out of 32 trials”), and this feedback may have impacted effort for some participants, thereby impacting their arousal regulation.

The conditional growth curve models suggested that poor accuracy on mental counters was associated with arousal dysregulation, as indicated by greater fluctuations in pupil dilation. Further, slower response times (i.e., poorer performance) was also associated with bouts of arousal dysregulation, as indicated by large increases in pupil dilation variability across trials. Taken together, performance during mental counters is related to pupil dilation variability over time and thus thought to be related to arousal regulation.

The first implication from this work is that assessing variability in pupil dilation longitudinally provides greater insight into the individual’s process during cognitive tasks, such as performing a working memory task. Future research could further explore its ability to predict performance breakdowns in real-time. The second implication is that individuals differ in their self-regulation of arousal in the moment, and this appears to impact performance. Related to Hockey's compensatory control model, ATC students must be able to regulate energetic resources in order to perform accurately and quickly to a myriad of events over time. Akin to athletes receiving real-time feedback on their physical state during exercise, pupil dilation variability could be used to provide arousal regulation feedback during ATC training scenarios. Such feedback could aim to train one's self-regulation ability that could serve as an operator trait (as opposed to an operator's state) and be embedded in current training.

ACKNOWLEDGMENT

The authors would like to acknowledge Audrey Moore for assistance with data collection and Jason Tsukahara for guidance with his pupillometry package in R.

REFERENCES

Alderton, D.L., Wolfe, J.H. and Larson, G.E., 1997. The ECAT battery. *Military psychology*, 9(1), pp. 5–37.

Barr, D.J., Levy, R., Scheepers, C. and Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), pp. 255–278.

Bates, D., Kliegl, R., Vasishth, S. and Baayen, H., 2015. *Parsimonious mixed models*. arXiv. 1506.

Held, J.D. and Wolfe, J.H., 1997. Validities of unit-weighted composites of the ASVAB and the ECAT battery. *Military Psychology*, 9(1), pp. 77–84.

Hockey, G.R.J., 1997. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological psychology*, 45(1-3), pp. 73–93.

Hockey, G.R.J., 2003. *Operator functional state: the assessment and prediction of human performance degradation in complex tasks* (Vol. 355). IOS Press.

Hockey, G. R. J. (2011). A motivational control theory of cognitive fatigue. In P.L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 167–188). Washington, DC: American Psychological Association

Hoffman, L. (2021). *Time-invariant predictors in longitudinal models* [online]. Available from: https://www.lesahoffman.com/ [accessed 28 February 2022].

Honour, Eric C. (2006) "A Practical Program of Research to Measure Systems Engineering Return on Investment (SE-ROI)", proceedings of the Sixteenth Annual Symposium of the International Council on Systems Engineering, Orlando, FL.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of memory and language,* 94, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mirman, D., 2017. *Growth curve analysis and visualization using R*. Chapman and Hall/CRC.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B. and Maintainer, R., 2021. Package 'nlme'. *Linear and nonlinear mixed effects models, version 3.1*. Available from: http://cran.rapporter.net/web/packages/nlme/nlme.pdf [accessed 28 February 2022].

Robison, M.K. and Brewer, G.A., 2020. Individual differences in working memory capacity and the regulation of arousal. *Attention, Perception, & Psychophysics*, 82(7), pp. 3273–3290.

Robison, M. and Brewer, G., 2021. Individual differences in working memory capacity, attention control, fluid intelligence, and pupillary measures of arousal. PsyArXiv Preprints arXiv: https://doi.org/10.31234/osf.io/n9avp.

Tsukahara, J.S. (2022). *pupillometry: An R package to preprocess pupil data* (v0.7.0). http://doi.org/10.5281/zenodo.6216399

Unsworth, N. and Robison, M.K., 2017. The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of experimental psychology: Learning, memory, and cognition*, 43(12), p.1962.