Pupil Size and Mental Load

Benjamin Stone (bpstone@psychology.adelaide.edu.au)
Department of Psychology, North Terrace, SA 5005 Australia

Michael Lee (michael.lee@psychology.adelaide.edu.au)
Department of Psychology, North Terrace, SA 5005 Australia

Simon Dennis (simon.dennis@psychology.adelaide.edu.au)
Department of Psychology, North Terrace, SA 5005 Australia

Ted Nettelbeck (ted.nettelbeck@psychology.adelaide.edu.au)
Department of Psychology, North Terrace, SA 5005 Australia

Abstract

This experiment explores whether the size of human pupil dilation is a measure that can discriminate between easier and harder intellectual tasks undertaken by the participant when information is displayed on a computer monitor. While it has been previously found that a participant's pupil dilation will be larger during harder intellectual tasks, these experiments have not generally been conducted under the environmental condition of light radiated from a computer monitor. The findings of this experiment indicate that computer monitor's radiance did not interfere with the researcher's ability to discriminate successfully between the participant's task-related pupil dilation.

Introduction

The call for a measure of cognitive processing load

A common request that spans several research disciplines has been for the identification, development, or evaluation of good measures of cognitive processing load. These disciplines include decision making research (Dennis, Bruza, & McArthur, 2002, p.132; Russo & Dosher, 1983, p.695) and human computer interaction (Christie, 1985, pp177-179). Several methods have been developed and implemented with varying success in response to this need. Russo and Dosher (1983) contend, that in process-oriented decision research that uses eye-tracking devices, eye fixation time can be a suitable measure of cognitive effort (p. 682). In contrast, Payne and colleagues (1988) calculate elementary information processes (EIPs) associated with decision strategies. This measure of effort is derived by demarcating a decision strategy into its theoretically-based component parts; the sum of these parts is assumed to provide an effort score that can be compared between strategies (pp. 534-535). More recently, Dennis et al. (2002) employed a dual task methodology to measure cognitive load during web browser searches of the internet. Participants in their study were required to concurrently browse the internet while also monitoring an auditory signal. Accuracy ratings on the auditory task were used as a measure of cognitive load. While these methods have been employed with mixed success, HCI researchers have continued to discuss the need for physiological measures of mental load (Ward & Marsden, 2003; Christie, 1983).

Pupil dilation as a measure of cognitive load

Historically, the pupil has been considered a window through which researchers may glimpse the cognitive processes of the human brain. Bumke, in his 1911 review of the relevant German literature, proposed that "in general every active intellectual process...produces pupil enlargement" (cited by Hess, 1972, p.492). Bumke's assertion has been supported by the work of many contemporary researchers. More recently, in an experiment that compared pupillary responses during three different mathematical-based tasks (no task, add 1 task, and subtract 7 task), researchers found that pupil diameter and the latency of the pupil constriction response to light stimulation were positively correlated with the level of problem difficulty (Steinhauer, Condray, & Kasparek, 2000); that is, both diameter and latency increased with difficulty. Moreover, mathematical task difficulty has consistently been associated with degree of pupil dilation (Hess & Polt, 1964; Boersma, Wilton, Barham, & Muir, 1970; Ahern and Beatty 1979; Schaefer, Ferguson, Brinton, Klein & Rawson, 1968; Steinhauer, Conray & Kasparek, 2000). In short, these studies found that greater pupillary dilations are observed when task difficulty is increased.

Pupil dilation as a measure of cognitive load in eye-tracking experiments that visually present stimuli on CRT monitors

HCI eye-tracking research has employed the pupil as a measure of cognitive load with mixed success. Lin, Zhang and Watson (2003) failed to find significant differences between the sizes of participants' pupil dilations when operating different computer-based interfaces for thermal-hydraulic processing plant systems. However, their experiment may have been hindered by at least two considerations. First, the experiment was conducted under fluorescent lights, which pulse, and could consequently create noise for the pupil data (Veitch & McColl, 1995). Second, the researchers examined the mean pupil dilation for each participant; however, mean pupil dilation size is affected by extremes cause by participants' blinks, and is therefore unreliable. The median size of pupil dilation is a more appropriate measure. Furthermore, the researchers acknowledged that, in at least some task comparisons, it could be argued that the two tasks were not "more difficult or more demanding" than each other, which is consistent with the pupil-based results that they obtained (p.859). In contrast to these findings, a study conducted by the Netherlands Aerospace laboratory utilized pupil dilation as a measure of user's cognitive load when comparing the effects of the automation of interfaces for Air Traffic Control systems (Jorna, 1997). They found that pupil dilation was a sensitive measure of cognitive load during tasks of increasing difficulty. Also, the pupil dilation results were consistent with other measures of cognitive load used in this study, such as a dual task condition.

It is possible that light radiated from the CRT monitor is interfering with researchers ability to utilize the pupil as a measure of cognitive load. The screen pulsating as it refreshes its display may further confound this situation. The mixed nature of HCI results indicates a need for further experimentation to identify the validity of the pupil as an indicator of cognitive load during tasks that are visual displayed on computer monitors.

Experimental research hypotheses

The aforementioned issues raise several research hypotheses that will be addressed in this thesis. These hypotheses will test under 'adverse' lighting conditions the robust finding, that more difficult mathematical tasks will elicit a greater magnitude of pupil dilation than easier mathematical tasks (Hess & Polt, 1964; Schaefer, et al., 1968; Bradshaw, 1968; Ahern & Beatty, 1979; Steinhauer et al., 2000). Therefore, it is predicted that when a stimulus 'X' is visually presented on a CRT monitor, on average, participants will produce significantly greater pupil width during SUB7 task when compared an easier ADD1 task. Furthermore, it is predicted that this result will be unaffected by the position of the stimulus ' 'X' in the CRT monitor's display. To this end, the CRT screen is delineated into five rows and columns corresponding to a 5x5 matrix. It is predicted that, when the stimulus 'X' is presented in each of the rows and columns in this experimental matrix, on average the participants' pupil width will be greater during an SUB7 task compared to an easer ADD1 task.

Methodology

Participants

Source, age-range, and eye-dominance. 50 participants were recruited from the Adelaide University Psychology Department, 41 females and 9 males. Of these participants, 36 were First Year Psychology Students who were willing to engage in voluntary psychological testing in return for partial course credit. There were no restrictions placed on participants' age, which ranged from 16 to 53 (M = 22 years 1 month, SD = 6 years 5 months). The Rosenbach (1903) sighting test was used to determine each participant's dominant eye, 40 percent of participants were left eye dominant while 60 percent were right eye dominant.

Selection Criteria. Potential participants were excluded if they reported any diagnosed ophthalmologic condition (for example, an astigmatism in the eye). Participants were also excluded if they reported either having a neck/back injury or that they were currently using of prescribed or non-prescribed medicines.

Apparatus & Procedure

Participants were seated in front of a Hitachi CM772 17 inch CRT Monitor set to refresh at 75 hertz. Screen resolution was set at 1024 x 768 pixels. Illuminance at the approximate position of the participant's right eye when facing the monitor (see Item B in Figure 1) was recorded at 24 lux when the monitor switched off and 58 lux when the monitor was displaying the experimental program. The ViewPoint EyeTracker PC-60 QuickClamp System was used to record participants eye-movements and pupil dynamics during the experimental tasks.

Participants consented to two types of testing: eye dominance and cognitive load. All testing was completed within 30 minutes.

First, the Rosenbach (1903) sighting test was used to determine each participant's dominant eye. This is a simple test that requires the participant to point at a distant object. While using their finger as a sight, the participant is then asked to close one eye, then reopen it, and then repeat this process with their other eye. The closing of their dominant eye will cause their finger seem offset from the target (Kommerell et al., 2003).

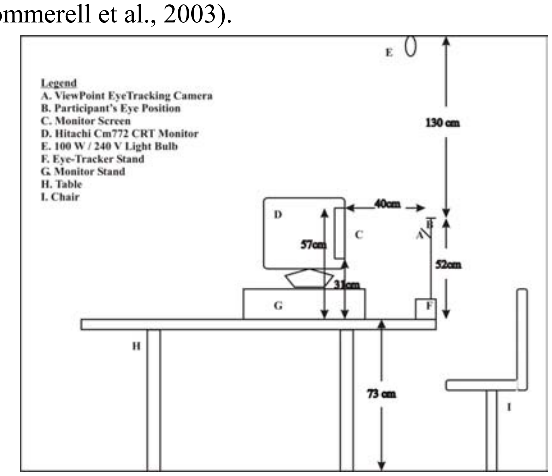

Figure 1: The layout of the experimental room.

Secondly, a View Point Eye-Tracking device was used to record the magnitude of the pupillary response in participants' dominant eye during mathematical tasks selected to that create different levels of cognitive load. There were two mathematical task conditions: ADD 1 and SUB7. The ADD1 condition required participants to continually add one to a randomly generated integer between 1 and 499. In the SUB7 task participants were asked to continually subtract seven from a randomly generated integer between 500-999. The participants were asked to do both tasks silently (in their heads) while visually following a ‘X’ that is presented on a computer monitor (see Figure 2).

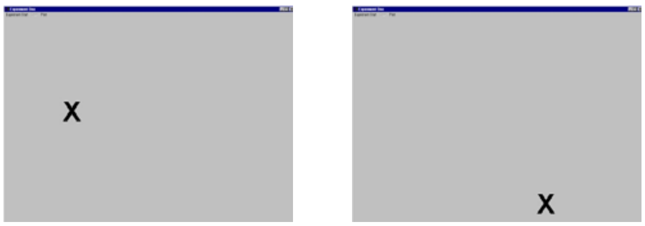

Figure 2: The visual stimulus is randomly moved to another cell in the experimental display every five seconds.

The visual display is divided into 25 invisible cells or regions of interest (ROI) that are arranged into a five by five grid. For each experimental task, the ‘X’ passes through 24 of these cells (all except the top-left cell), stopping in each cell for ten seconds. The sequence through these 24 cells is randomly generated using a C++ function that is ‘seeded’ with an integer from the computer clock, thereby generating a quasi-random sequence and guaranteeing that participants did not receive the stimuli in the same order. An average pupillary response to each task condition for each subject was calculated over the 24 cells. Also, average pupillary responses to each task condition for each subject were calculated over the rows and columns contained in the experimental grid.

The ViewPoint Eye-Tracker measures pupil width in “pixels normalized with respect to the EyeCamera window” (Arlington Research, Inc., 2002, p.25). In the results section, for convenience, this measurement has been multiplied by ten. Participants’ pupil dilation widths were recorded when the eye-tracking software recorded that the co-ordinates of the eye were in the same region of interest as the experimental stimulus ‘X’. This method generated several thousand pupil width measurements for each participant during each task. The median of these measurements (referred hereafter as ‘Median Pupil Width’) was used to combat confounding effects of extreme scores generated from participant’s blinking and moving

Results

This results section is delineated into three parts. The first sub-section reports overall findings when participants’ pupil width was compared for both the easy (ADD1) and harder (SUB7) mathematical tasks. The second and third sections report findings when participants’ pupil widths were compared during these tasks across the rows and columns of the experimental grid, respectively.

Descriptive statistics derived from the participants’ pupillary responses during the mathematical tasks are presented in Table 1. In all comparisons, participants’ Median Pupil Width was found to be larger during the SUB7 task when compared to the easier ADD1 task. For each of these comparisons, a related sample t-test has been used to compare means. These findings were statistically significant in nearly all of the experimental locations used for these comparisons.

Table 1: Descriptive statistics of the participants’ Median Pupil Width

	Task	N	Range	Min.	Max.	M	SD	Skew	SE	Kurtosis	SE
Overall	ADD1	50	116.00	112.00	228.00	170.82	26.49	-0.19	0.34	-0.36	0.66
	SUB7	50	138.00	118.00	256.00	186.02	30.94	-0.21	0.34	-0.33	0.66
Rows:											
R0	ADD1	50	128.00	118.00	246.00	182.09	30.59	-0.101	0.34	-0.58	0.66
	SUB7	50	144.00	124.00	268.00	197.61	34.07	-0.17	0.34	-0.43	0.66
R1	ADD1	50	128.00	112.00	240.00	175.12	28.30	-0.14	0.34	-0.25	0.66
	SUB7	50	150.00	118.00	268.00	191.55	32.37	-0.28	0.34	-0.20	0.66
R2	ADD1	50	112.00	109.00	221.00	167.90	25.60	-0.28	0.34	-0.48	0.66
	SUB7	50	141.00	118.00	259.00	182.20	30.00	-0.05	0.34	-0.07	0.66
R3	ADD1	50	112.00	106.00	218.00	165.16	25.65	-0.23	0.34	-0.39	0.66
	SUB7	50	125.00	115.00	240.00	179.14	28.92	-0.39	0.34	-0.44	0.66
R4	ADD1	50	109.00	112.00	221.00	168.00	25.54	-0.27	0.34	-0.40	0.66
	SUB7	50	129.00	121.00	250.00	183.10	32.16	0.15	0.34	-0.57	0.66
Columns:											
C0	ADD1	50	328.00	115.00	443.00	185.90	54.56	3.18	0.34	12.93	0.66
	SUB7	50	156.00	118.00	274.00	196.33	34.19	-0.09	0.34	-0.10	0.66
C1	ADD1	50	112.00	112.00	224.00	174.50	26.13	-0.36	0.34	-0.40	0.66
	SUB7	50	138.00	121.00	259.00	189.56	32.24	-0.18	0.34	-0.63	0.66
C2	ADD1	50	116.00	112.00	228.00	171.79	26.95	-0.10	0.34	-0.44	0.66
	SUB7	50	129.00	121.00	250.00	186.20	31.21	-0.27	0.34	-0.43	0.66
C3	ADD1	50	112.00	109.00	221.00	168.17	27.04	-0.24	0.34	-0.64	0.66
	SUB7	50	135.00	118.00	253.00	181.65	30.22	-0.23	0.34	-0.41	0.66
C4	ADD1	50	100.00	115.00	215.00	163.24	25.18	-0.14	0.34	-0.69	0.66
	SUB7	50	128.00	115.00	243.00	177.16	30.09	-0.08	0.34	-0.40	0.66

Overall

Increased task difficulty (ADD1 to SUB 7) was paralleled by an amplification of mean Median Pupil Width (170.82, 186.02, respectively). Overall, the distributions of the Median Pupil Width were normal in both mathematical tasks. The significant differences between average SUB7 and ADD1 Median Pupil Width ($t(49) = -12.31$, $p < .001$) indicated a medium effect size ($d = -0.53$). The differences between these means are graphically displayed in Figure 3.

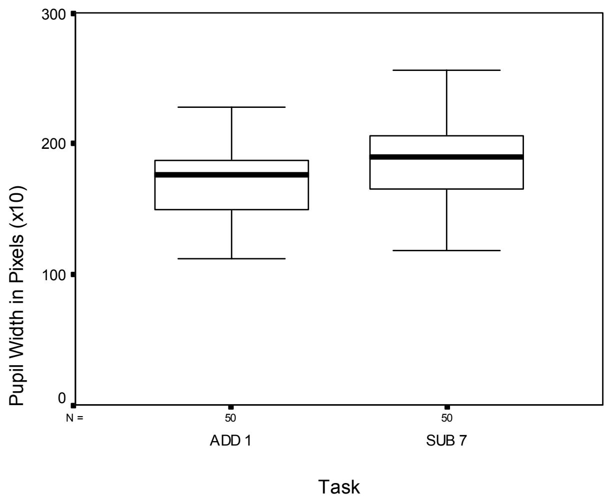

Figure 3: A Boxplot illustration of the differences between participants' overall average SUB7 and ADD1 Median Pupil Widths.

Rows

In comparisons of row data, increases in task difficulty corresponded with enlarged Median Pupil Widths (see Table 1). The related sample t-tests shown in Table 2 indicate that the differences between means associated with these increases were statistically significant in all comparisons.

Table 2: Results of related samples t-tests used to compare average Median Pupil Widths recorded while participants were viewing the 'X' Stimulus in rows 0 to 4.

Comparison ADD1-SUB7 Median Pupil Width	t	d.f.	p
R0	-9.60	49	<0.001
R1	-10.51	49	<0.001
R2	-9.81	49	<0.001
R3	-11.00	49	<0.001
R4	-8.16	49	<0.001

Moderate effect sizes were found when comparing the ADD1 and SUB7 row data ($d = -0.48$, $d = -0.54$, $d = -0.51$, $d = -0.51$ and $d = -0.52$) in rows 0 to 4, respectively. All row data were normally distributed and are graphically display in Figure 4.

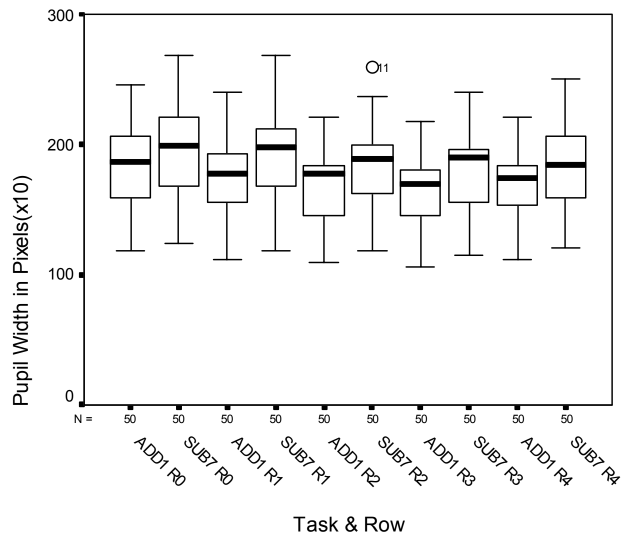

Figure 4: A Boxplot illustration of the differences between participants' average SUB7 and ADD1 Median Pupil Widths in each of the 5 rows in the experimental grid.

Columns

In all column comparisons increases in task difficulty corresponded with enlarged Median Pupil Widths (see Table 1). Related sample t-tests found that the differences between means associated with these increases were statistically significant in 4 of 5 comparisons (see Table 3).

Table 3: Results of related samples t-tests used to compare average Median Pupil Widths recorded while participants were viewing the 'X' Stimulus in columns 0 to 4.

Comparison ADD1-SUB7 Median Pupil Width	t	d.f.	p
C0	-1.78	49	n.s.
C1	-9.07	49	<0.001
C2	-8.59	49	<0.001
C3	-9.76	49	<0.001
C4	-8.90	49	<0.001

The data gathered when participants were viewing stimuli in the leftmost column (C0) was not associated with significant statistical differences in the Median Pupil Width during the ADD1 and SUB7 tasks. Moderate effect sizes were found when comparing the ADD1 and SUB7 row data ($d = -0.51$, $d = -0.49$, $d = -0.47$ and $d = -0.50$) in columns 1 to 4, respectively.

With the exception of C0, all column data were normally distributed, and can be viewed in Figure 5. The data in C0 is positive skewed by of two outliers that are 3.74 and 4.71 standard deviations from the mean (see Figure 6). When these outliers are removed the difference between ADD1 and SUB7means in C0 is statistically significant ($t(47) = -8.91$, $p < .001$), revealing a medium effect size ($d = 0.60$).

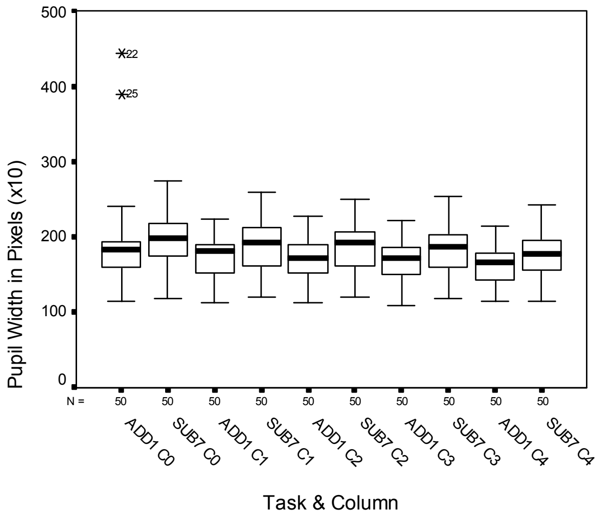

Figure 5: A Boxplot illustration of the differences between participants' average SUB7 and ADD1 Median Pupil Widths in each of the 5 columns in the experimental grid.

Discussion

This report has focused on one broad question. This question asks, is it possible to distinguish between tasks of different cognitive difficulty based on participants' pupil widths when stimuli are visually presented on a CRT monitor? It was argued that presence of the CRT monitors illumination, and the pulsating nature of the screen refreshing, may have confounded the use of participants' pupil dynamics as a measure of cognitive load. Supporting the first research hypothesis, overall the results of this research suggest that CRT monitors can successfully be used to visually present information to participants in experiments that utilize pupil dynamics as a measure of cognitive load.

This finding is consistent when data is recorded from the information display area as a whole and when this display is divided into rows. However, in contrast to the second research hypothesis, the results were not always statistically significant when the display area was delineated into columns. While it was possible to statistically correct or remove the erroneous column data, this result suggests that calibration errors occurred during the experiment. These errors may have been introduced by participant's moving or fidgeting during the experiment. That is, movement of the eye-ball away from its calibrated line of sight may have increased the area defined as pupil beyond its normal bounds when the participant was viewing some areas of the screen. One way of addressing this problem would be to incorporate monitoring software into these experiments that alerts the experimenter when eye-tracking calibration has been compromised.

There are several other limitations present in this research. This article has not examined the results gathered for the individual cells in the experimental matrix. To extend the

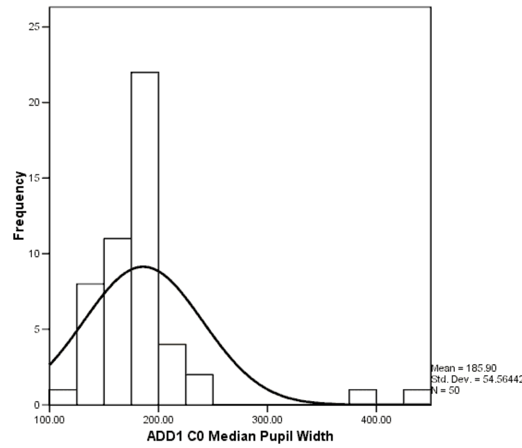

Figure 6: A histogram displaying the relative position of the outliers in the distribution of ADD1 in C0.

reporting of this research to level of the individual cells was considered to be beyond the scope of this article. Furthermore, the pupil's discriminatory power needs to be more fully explored. While pupil dynamics can be used to discern between SUB7 and easier ADD1 tasks, can they also be used in finer discriminations between cognitive tasks? This is a question that should be addressed in future research.

The robust effect studied in this experiment indicates that the processes subserving pupil dilation, whether they be the sympathetic innervation (Hess, 1972) or the vasculature of the iris (Appenzeller & Oribe, 1997), may be affected, either directly or indirectly, by the information processing demands placed on the human brain during mathematical problem solving. Several researchers have suggested that pupil dilation is part of the orienting reflex (or response) (Sokolov, 1963; Kahneman, 1973), and others have speculated as to its functional significance. For example, Spinks and Siddle (1983) have tentatively suggested that increased pupil dilation during arousal will allow more light into the eye and may facilitate clearer perception of focused objects of interest, whilst simultaneously blurring distracting peripheral visual cues.

In summary, while the level of information processing required by humans during mathematical tasks appears to produce a robust effect, displayed in the magnitude of pupillary dilation, the functional significance of this phenomenon, if any, remains a matter for speculation. In the current experiment, this effect was not noticeably confounded by light emitted from the CRT monitor. However, pupil data accuracy appeared to be affected by either stimulus placement and/or participant movement. Even though both the physiological and functional basis of this phenomenon remain controversial, it appears that participants' pupil dilation can be exploited by researchers as a measure of cognitive load.

Acknowledgments

We would like to thank the University of Adelaide Psychology department for their ongoing support of this research.

References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing varying with scholastic test scores. *Science, 205,* 1289-1292.

Arlington Research, Inc. (2002). ViewPoint EyeTracker ™ PC-60 Software Users Guide.

Appenzeller, O., & Oribe, E. (1997). *The Autonomic Nervous System: An introduction to the basic and clinical concepts.* Elsevier: Amsterdam.

Boersma, F., Wilton, K., Barham, R., & Muir, W. (1970). Effects of arithmetic problem difficulty on pupillary dilation in normals and educable retardates. *Journal of Experimental Child Psychology, 9,* 142-155.

Christie, B. (1985). *Human Factors of the User-System Interface: A Report on an ESPRIT Preparatory Study*. Elsevier: Amsterdam.

Dennis, S., Bruza, P., & McArthur, R. (2002). Web Searching: A process-oriented experimental study of three interactive search paradigms. *Journal of the American Society for Information Science and Technology, 53,* 120-133.

Hess, E. H. (1972). Pupillometrics: A method of studying mental, emotional, and sensory processes. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of Psychophysiology*. New York: Holt, Rinehart and Winston.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 140,* 1190-1192.

Jorna, P. G. A. M. (1997). Human machine interfaces for ATM: objective and subjective measurements on human interactions with future flight deck and air traffic control systems. url: *http://atm-seminer-97.eurocontrol.fr/jorna.htm*

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kommerell, G., Schmitt, C., Kromeier, M., & Bach, M. (2003). Ocular prevalence versus ocular dominance. *Vision Research, 43,* 1397-1403.

Lin, Y., Zhang, W. J., & Watson, L. G. (2003). Using eye movement parameters for evaluating human-machine interface frameworks under normal control operation and fault detection situations. *International Journal of Human-Computer Studies, 59,* 837-873.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology, 14,* 534-552.

Russo, J. E., & Dosher B. A. (1983). Strategies for multiattribute binary choice. *Journal of experimental Psychology: Learning, Memory and Cognition, 9,* 676-696.

Schaeffer, T, Jr., Ferguson, J. B., Klein, J. A., & Rawson, E. B. (1968). Pupillary responses during mental activities. *Psychonomic Science, 12,* 4, 137-138.

Simpson, H. M., & Hale, S. M. (1969). Pupillary changes during a decision-making task. *Perceptual and Motor Skills, 29,* 495-498.

Sokolov, Y. N. (1963). *Perception and the conditioned reflex.* Pergamon Press: Oxford.

Spinks, J. A., & Siddle, D. (1983). The functional significance of the orienting response. In D. Siddle (Ed.), *Orienting and habituation: Perspectives in human research*. John Wiley & Sons: Chichester.

Steinhauer, S. R., Condray, R., & Kasparek, A. (2000). Cognitive modulation of midbrain function: task-induced reduction of the pupillary light reflex. *International Journal of Psychophysiology, 39,* 21-30.

Veitch, J.A, & McColl, S.L. (1995). Modulation of Fluorescent Light: Flicker Rate and Light Source Effects on Visual Performance and Visual Comfort. *Lighting Research and Technology, 27,* 243-256

Ward, R. D., & Marsden P. H. (2003). Physiological responses to different WEB page designs. *International Journal of Human-Computer Studies, 59,* 199-212.