RESEARCH REPORT

Pupillary Stroop effects

**Bruno Laeng · Marte Ørbo · Terje Holmlund ·
Michele Miozzo**

Abstract We recorded the pupil diameters of participants performing the words' color-naming Stroop task (i.e., naming the color of a word that names a color). Non-color words were used as baseline to firmly establish the effects of semantic relatedness induced by color word distractors. We replicated the classic Stroop effects of color congruency and color incongruency with pupillary diameter recordings: relative to non-color words, pupil diameters increased for color distractors that differed from color responses, while they reduced for color distractors that were identical to color responses. Analyses of the time courses of pupil responses revealed further differences between color-congruent and color-incongruent distractors, with the latter inducing a steep increase of pupil size and the former a relatively lower increase. Consistent with previous findings that have demonstrated that pupil size increases as task demands rise, the present results indicate that pupillometry is a robust measure of Stroop interference, and it represents a valuable addition to the cognitive scientist's toolbox.

Keywords Stroop · Pupillometry · Cognitive load · Attention · Semantics

B. Laeng (✉)
Department of Psychology, University of Oslo,
1094 Blindern, 0317 Oslo, Norway
e-mail: bruno.laeng@psykologi.uio.no

M. Ørbo
Department of Psychology,
University of Tromsø, Tromsø, Norway

T. Holmlund
Department of Clinical Medicine,
University of Tromsø, Tromsø, Norway

M. Miozzo
Department of Experimental Psychology,
University of Cambridge, Cambridge, UK

M. Miozzo
Department of Cognitive Science,
Johns Hopkins University, Baltimore, USA

Introduction

As researchers have known for more than a century, variations in human pupil size occur in response to stimuli presented in different modalities or when people are engaged in cognitive tasks, despite constant illumination and no changes in ocular accommodation (Winn et al. 1993; Beatty and Lucero-Wagoner 2000; Steinhauer et al. 2000; Barbur 2004). The first applications of pupillometry (i.e., the measurement of changes in pupillary diameter) to psychology occurred in two seminal experiments by Hess and Polt (1960, 1964; see also for a review: Loewenfeld 1993), which demonstrated that pupil size increased when simply viewing attention-grabbing stimuli but also during mental multiplication, so that pupillary diameters increased as the interest value of stimuli was higher and as the magnitude of the numbers became larger.

Subsequent studies have repeatedly shown that pupil size variations are positively correlated to the intensity of the stimulus (e.g., tones varying in dB; Stelmack and Siddle 1982) or the difficulty of the task (e.g., increasing load on memory; Kahneman and Beatty 1966). One of the first applications of pupillometry to psycholinguistics was conducted by Just and Carpenter (1993) who demonstrated that pupil diameter changed in reading as a function of sentence complexity. The studies mentioned above, together with a dearth of other findings (e.g., Granholm et al. 1996; Kahneman and Beatty 1966; Libby et al. 1973; Pratt 1970;

Peavler 1974; Hoeks and Levelt 1993; Just and Carpenter 1993; Krüger et al. 2001; Karatekin et al. 2004; Verney et al. 2004; Nuthmann and van der Meer 2005; Bijleveld et al. 2009; Piquado et al. 2010; Van der Meer et al. 2010), converge on the proposal that variations of pupil size represent a physiological marker of processing "load", "cognitive effort", or the use of attentional "resources" (Kahneman 1973; Beatty 1982) that are mediated by "autonomic arousal" (Bradley et al. 2007). However, as pointed out by Just and Carpenter (1993), because pupillary responses are only indirect markers or correlates of how intensely a processing system is operating, they are epiphenomenal (not causally linked) to further cognitive processes. In other words, citing an analogy drawn by Beatty and Lucero-Wagoner (2000), just as reporter genes can be used as 'reporter variables' of bio-cellular processes, pupillary responses can only be used as 'reporter variables' of the use of specific cognitive mechanisms and of the engagement of their underlying neural substrates (Koss 1986; Sara 2009).

The inherent limitations of pupillometry have not prevented researchers from exploiting this technique for the purposes of understanding specific cognitive processes. An illustrative example of this application of pupillometry comes from perceptual rivalry. Einhäuser et al. (2008) showed that pupillary dilations have a predictive relationship to switches between conscious percepts, a finding suggesting that pupillary changes are the physiological marker of the resolution between competitive representations in awareness (cf. Paulsen and Laeng 2006). Here, we focus on the applicability of pupillometry to the investigation of word meaning (semantics) processing. Recent studies have examined the pupil correlates of the Stroop task (Brown et al. 1999; Siegle et al. 2004, 2008), which has been one of the primary paradigms that cognitive scientists have used to identify how word meaning is represented and accessed in speakers' minds and brains. Participants in the Stroop task name the color with which written words are shown and, although instructed to disregard the written words and concentrate instead on the colors of the displayed words, participants cannot avoid reading the words, with inevitable interfering effects that are particularly noticeable when the words denote color terms different from the word colors (e.g., when the word *red* appears in blue). Color words can generate their interference effect only if their meaning is accessed, a fact explaining the prominence of the Stroop task in research on semantic processing (cf. Laeng et al. 2005). Pupillometric studies replicated the Stroop color effect measuring eye responses: the largest pupil dilations were recorded when word meaning and word color mismatched. As these results show that pupil responses are sensitive to word meaning, they also demonstrate the usefulness of pupillometry for the investigation of semantic processing.

Some of the features of pupillometry make it a promising tool for research on the neurocognition of semantics. To the extent that pupillary changes can be continuously recorded and are phase-locked to the unfolding of cognitive processes, they can provide valuable clues about the real-time structuring of cognitive processing (cf. Richardson et al. 2007). Researchers can also profit from the sensitivity of pupil variations to representations and responses that are only partially activated and that may never pass the threshold for eliciting overt behavior or conscious appraisal (Bijleveld et al. 2009; Laeng and Falkenberg 2007; Laeng et al. 2007). Another advantage is that pupil recording can occur without any disruption of behavioral tasks. Being a non-invasive measure, pupil recording does not require overt responses and is obtainable even without participant knowledge. Finally, the imperviousness of pupil measurements to artifacts induced by head movements or other muscular activity makes pupillometry particularly suited to tasks involving overt verbal responses, including the Stroop color-naming task.

Although prior studies of the pupillometry of Stroop effects revealed impressively robust findings (Brown et al. 1999; Siegle et al. 2004, 2008), they were unfortunately limited in two important respects. The first limitation concerns the baseline (letter strings and color-congruent words) used to establish color-incongruency effects. That is, prior studies reported greater pupil size for the word *blue* than a string of Xs or the word *red*, when the expected color response was "red." Such baselines do not allow a univocal interpretation of color-incongruency effects. Because color words and letter strings differ in meaning as well as lexical status (words vs. non-words), the pupil dilation in Stroop tasks could reflect differences in the processing of meaning, lexical information, or both. Analogous difficulties arise when color-incongruent words are compared to color-congruent words. Differences between the responses to these two types of distractors can also stem from variations in their phonologies and orthographies; that is, while color-congruent words are phonologically and orthographically identical to the word targets, color-incongruent words are not. These problems can be rectified if non-color words are employed as baseline, a procedure commonly adopted in studies of the Stroop paradigm (e.g., Klein 1964; Burt 2002). In this way, it can be established more precisely whether pupil dilations vary as a function of word meaning. The second limitation of prior studies that applied pupillometry to Stroop tasks concerns the time course of pupil responses. Brown et al. (1999) only reported pupil responses averaged across the entire recording interval; although Siegle et al. (2004, 2008) examined the time course of pupillary responses, their analyses were restricted to the comparison of color-congruent versus color-incongruent words. By providing a

continuous recording of a physiological response to cognitive tasks, pupillometry offers the opportunity to determine the time course and thus an additional variable for examining neurocognitive mechanisms.

The Norwegian speakers participating in our study named the colors in which the written stimuli appeared. We recorded the pupil responses—their diameters and time courses—induced by color words and non-color words. Color words appeared either in a congruent color (e.g., the word rød [red] shown in red) or in an incongruent color (e.g., the word rød [red] shown in green). Following the experimental design originally adopted by Stroop (1935), later replications have also included color-congruent words. For sake of comparability with prior studies, a color-congruent condition was also included in our study. Non-color words were used as baseline because, as we discussed above, they provide an adequate reference point for assessing the effects of color congruence/incongruence.

In addition to pupillary responses, we also measured RTs. The reason for collecting response latencies was twofold. On the one hand, we wanted to assure that our design replicated chronometric effects reported in prior studies. On the other hand, we wanted to compare the sensitivity (i.e., effect sizes) to Stroop interference of RTs versus pupillometry. This comparison would help us determine whether pupillometry could represent a valuable alternative or complement to RT measurements.

Methods

Participants

Forty students (20 females) of the University of Tromsø (Norway) volunteered for the experiment (mean age = 26.8; SD = 5.2). All participants were native Norwegian speakers, had normal or corrected-to-normal visual acuity, and showed no signs of color vision deficiencies, as measured with the Farnsworth-Munsell 100 Hues test (GretagMacbeth©).

Stimuli

Distractors were either Norwegian color words or Norwegian non-color words. There were nine color words, namely: gul [yellow], blå [blue], grønn [green], rød [red], rosa [pink], lilla [violet], brun [brown], oransje [orange], and grå [gray]. The nine non-color words were nouns not associated with a particular color (e.g., bord [table], rye [mat]). Color and non-color word distractors were matched for length (number of letters). Color words were shown either in congruent colors (e.g., rød in red) or incongruent colors (e.g., rød in blue). To reduce the proportion of color

words, we also presented N distractors formed by letter strings (meaningless, pronounceable pseudowords conforming to Norwegian orthography; cf. Kristoffersen 2000), which were used as fillers and excluded from result analyses. Eighty color-congruent words, 80 color-incongruent words, 80 non-color words, and 80 filler trials consisting of pronounceable but meaningless pseudowords were used in the experiment. Each of the color responses occurred with equal frequencies in the whole experiment and in association with the different types of distractors. Care was taken not to pair words and colors that sounded alike, since a phonological overlap facilitates the naming response (see MacLeod 1991, for a review of this effect). Order of presentation was randomized. Words were presented over a white background (RGB: 250, 250, 250) as bitmap images (800 × 600 pixels) in Geneva font 36 and subtended no more than 7 degrees of visual angle.

Pupillometry

A Remote Eye Tracking Device by SensoMotoric Instruments (RED, SMI) was used to record the horizontal and vertical coordinates of each participant's left eye. The eye-tracker had a 2-ms sample rate with a resolution smaller than 0.1 degree; it operated with an infrared-light-sensitive video camera that allowed recording in illuminated rooms. The illumination of the testing room was kept constant throughout testing sessions. A single measure (in pixels) of pupil diameter was obtained for each sample by averaging the horizontal and vertical coordinates of the pupillary diameter.

Procedure

Participants were tested individually in a windowless and soundproof room. Before the experiment proper, participants named palettes of the nine target colors to assure the use of the expected color labels. Participants were seated in front of a computer screen at a distance of 72 cm with their chins and foreheads stabilized in a headrest. They were instructed to ignore the words and name their pixel colors as fast and accurately as possible. A trial began with the presentation of a white screen of the same luminance of the target trials (RGB: 250, 250, 250), which was followed by a 1,000-ms presentation of 3–7 black pound symbols (#), and then, in the same position, by a 2,000-ms presentation of the alphabetic stimuli (word/non-words). The pound symbols were of the same size of the letters used in the words and appeared at the center of the screen. The pound symbols had a black outline and were filled with the same white color of the background. The beginning of pupillometric recording was synchronized with the appearance of the alphabetic stimulus. We always refer to 0 as the onset

of the alphabetic stimulus. E-Prime© software recorded RTs from the onset of the alphabetic stimulus until a vocal emission response was picked up by the microphone (i.e., until the time elapsing between the appearance of the alphabetic stimulus and the onset of the vocal response, which triggered the speech recording device). The experimenter manually recorded the response accuracy. A training block preceded the experiment proper. The word distractors used for training did not appear in the experiment proper. Three non-color words were used at the beginning of the experiment as warm-up stimuli; these stimuli were not part of the experimental word set and were not included in any of the analyses. Responses scored as errors included instances in which participants used incorrect color labels, stuttered, uttered "uhms" or other speech hesitations, or in which microphone failures were recorded. Stimuli were presented in 4 blocks, and the whole experiment lasted approximately 60 min.

Analyses

Response latencies that exceeded a participant's mean by more than 3 SDs or that were shorter than 100 ms were treated as outliers and eliminated from the analyses of RTs and pupil dilations. Errors were also excluded from such analyses. Technical problems made the RT data of three participants unusable, so that the data of only 37 participants were included in the analyses. Pupil responses with diameter equal to zero were eliminated from analyses, as they are likely to result from eye blinking. We also excluded pupil responses with diameters smaller/greater than 3 SDs from the mean of a specific condition. The application of these trimming procedures led to eliminating 1% of the RTs data and 3% of the pupillometry data.

Results

Naming responses

Errors (incorrect color names) occurred rarely (1%) and with similar rates across the various types of distractors. Given this pattern, errors were not further analyzed. Naming latencies were entered in a within-subject ANOVAs with type of distractor as variable (color-congruent vs. color-incongruent vs. non-color words). The ANOVA revealed significant differences across the response latencies of the three distractor types, $F(2, 78) = 15.2$, $P < 0.0001$. Responses were faster for congruent colors (mean RT = 771 ms; SD = 89) than for incongruent colors (mean RT = 914 ms; SD = 95), whereas responses to non-color words (mean RT = 820 ms; SD = 68) showed intermediate latencies. Post hoc Scheffé tests on these

means showed that the incongruent condition caused significantly slower RTs than the congruent condition (mean difference = 142 ms; critical difference = 39 ms; $P < 0.0001$; effect size $d = 0.6$) and the non-color words (mean difference = 99 ms; critical difference = 39 ms; $P < 0.0001$; effect size $d = 0.5$); the difference in overall mean change in pupillary diameter for the congruent condition and the non-color words conditions was also significant (mean difference = 50 ms; critical difference = 39 ms; $P < 0.01$; effect size $d = 0.3$).

Pupillometric changes

For each participant, we determined the average pupillary diameter recorded for fixation symbols during the 200-ms preceding word onset. This average was subtracted from the pupillary diameters recorded for correct responses at each 20-ms sample point within the 0- to 2,000-ms interval. The obtained baseline-corrected data on mean change in pupillary diameters for the whole 2-s period were then entered in a repeated-measures ANOVA with condition (color-congruent vs. color-incongruent vs. non-color words) as the within-subject factor. This analysis revealed a main effect of condition, $F(2, 78) = 15.8$, $P < 0.0001$. In essence, pupillary responses replicated the response pattern obtained with RTs. That is, color-incongruent words caused the largest change in pupil diameter (mean = 0.072 mm; SD = 0.11). In contrasts, color-congruent words produced the smallest change (mean = 0.044 mm; SD = 0.91). Non-color words (mean = 0.053 mm; SD = 0.97) resulted in intermediate changes (see Fig. 1). Post hoc Scheffé tests on these mean changes in pupillary diameter over the whole recording epoch confirmed that the incongruent condition caused small but significantly larger pupillary dilations than the congruent condition (mean difference = 0.028; critical difference = 0.05; $P < 0.0001$; effect size $d = 0.14$) and the non-color words (mean difference = 0.019; critical difference = 0.05; $P < 0.0001$; effect size $d = 0.09$); the difference in overall mean change in pupillary diameter for the congruent condition and the non-color words conditions was also significant (mean difference = 0.008; critical difference = 0.05; $P < 0.004$; effect size $d = 0.05$).

In addition, we identified the maximum peak of each of the three conditions based on the maximum mean value for each condition. We then computed the mean (1,380 ms) and standard deviation (60 ms) of the latencies of these three peaks, and, based on these obtained values, we identified an epoch around the mean latency that was 2 SDs wide (120 ms); that is, an epoch that included samples from 1,320 to 1,440 ms. We then computed, for each subject, the means of the three conditions within this epoch, and we entered these data into a final ANOVA, with condition (color-congruent vs. color-incongruent vs.
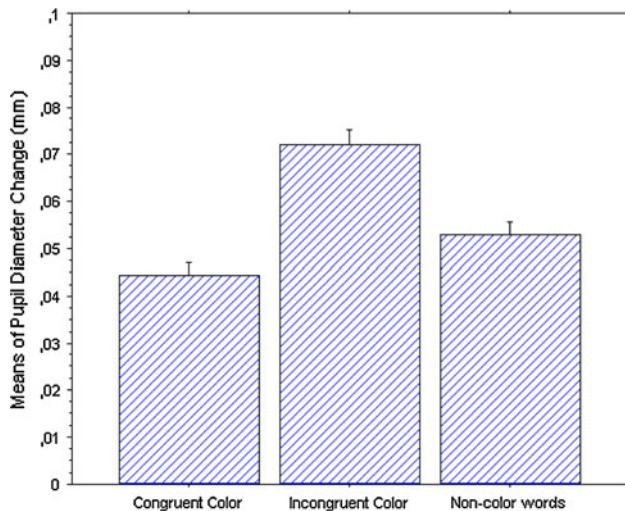
Fig. 1 Change in mean pupil diameters (in mm) averaged over a 2-s epoch from onset of each distractor stimulus. *Bars* indicate 95% confidence intervals for within-subject designs (Loftus and Masson 1994)

non-color words) as the within-subject factor. This analysis revealed a highly significant effect of condition, $F(2, 78) = 19.2$, $P < 0.0001$. Color-incongruent words caused the largest change in pupil diameter (mean = 0.157 mm; SD = 0.13), color-congruent words the smallest change (mean = 0.091 mm; SD = 0.10), and pupillary responses to non-color words were intermediate (mean = 0.124 mm; SD = 0.11). Post hoc Scheffé tests showed that the incongruent condition caused significantly larger pupillary dilations than the congruent condition during this selected

time window (mean difference = 0.066; critical difference = 0.024; $P < 0.0001$; effect size $d = 0.27$) and the non-color words (mean difference = 0.033; critical difference = 0.024; $P < 0.004$; effect size $d = 0.14$); the difference in overall mean change in pupillary diameter for the congruent condition and the non-color words conditions was also significant (mean difference = 0.033; critical difference = 0.024; $P < 0.004$; effect size $d = 0.15$).

As Fig. 2 clearly shows, pupillary changes in diameter showed two general peaks. The first, smaller, peak reached its asymptote early at about 400 ms and showed no evident differences among conditions. The second, larger, peak reached its maximal value at about 1,400 ms; thus later than the first peak but also later than the average onset of vocal responses. Most importantly, this second peak showed diverging distractor effects in the distribution of pupil diameters that remained separate up to the last sample of pupillary diameter recording (i.e., from 1,400 to 2,000 ms).

Finally, we performed a simple regression analysis based on each participant's mean of pupillary change (as the regressor) and the respective mean RTs as the dependent variable. Pupillary change had no significant predictive value on RTs (y = 848–101*x; slope coefficient: $t(35) = 0.74$, $P = 0.40$). More interestingly, in a second simple regression in which we analyzed measures of the Stroop effect (i.e., incongruent condition minus congruent condition) corresponding to RTs and pupillary change, respectively, we found that the "pupillary Stroop" significantly predicted the "RT Stroop" (y = 126–259*x; slope coefficient: $t(35) = 2.2$, $P < 0.05$; $R = 0.4$).

Fig. 2 Mean pupil diameters (in mm) at each 20-ms sample and for each distractor condition. Time 0 represents the onset of each stimulus. The *colored vertical lines* represent the point in time of each condition's mean RT
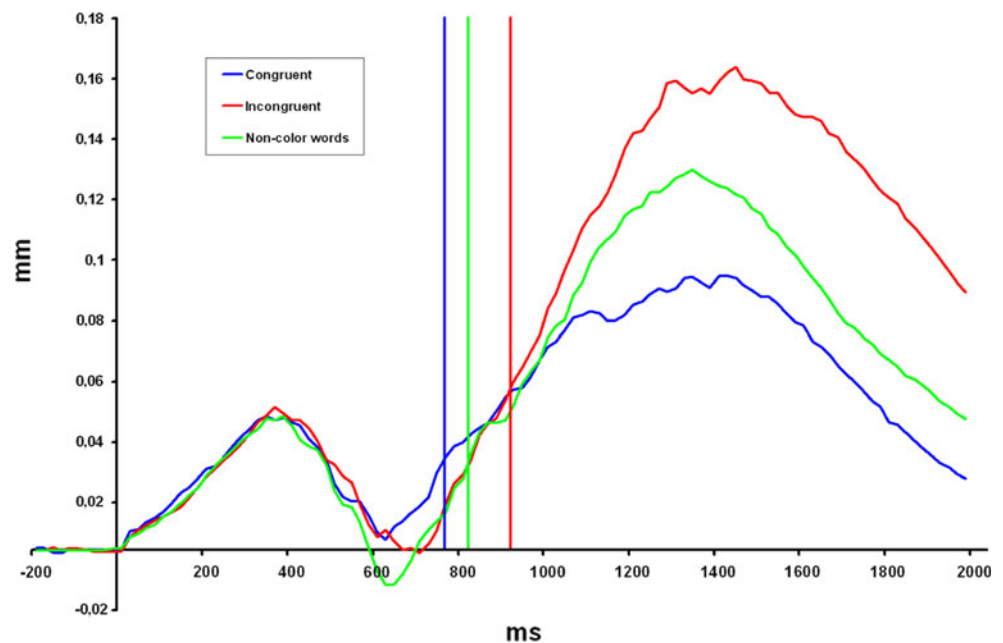
Discussion

Using non-color words as baseline, we found that pupil size increased for color-incongruent distractors, but decreased for color-congruent distractors. We thus essentially replicated the result pattern observed in prior studies of the pupillometry of Stroop effects (Brow et al. 1999; Siegle et al. 2004, 2008). If we take pupil dilation as an index of distractor interference, with greater dilation corresponding to larger interference, recordings of pupil sizes and naming latencies converged impressively in the present study. Both measures replicated the classic Stroop effect of color-incongruence—larger dilations and longer naming latencies for color-incongruent words than non-color words. There was an opposite pattern for color-congruent words, which induced smaller pupil size and shorter naming latencies relative to non-color words. Moreover, a Stroop incongruent condition was compared to either (a) the traditional baseline of comparing incongruent distractors to congruent distractors or (b) an alternative baseline based on comparing incongruent distractors to non-color words. Hence, we conclude that, although semantics is a determining factor for the occurrence of Stroop interference, a significant part of the interference effect is contributed by variations in the phonologies and orthographies of the stimulus words.

Interestingly, the physiological and performance measurements differed in terms of effect sizes, which were greater for RTs than pupillary responses also when compared with peak pupillary responses. However, the pupillary measures were also clearly sensitive to the cognitive conflict of Stroop interference; in fact, they were positively (albeit modestly) related to the RTs. Thus, an implication of these findings is that the physiological measure of Stroop interference as changes in pupillary diameter could be a valid substitute to the more traditional performance measure (i.e., RT) of Stroop interference in cases where these are difficult or impossible to obtain (e.g., in aphasic patients). Future research will clarify whether, in other cognitive paradigms, pupillometry may also be able to identify effects that would remain undetected with RT measures.

Altogether, the pupillary changes we observed with the different distractors allow us to better characterize the correlates of pupil responses. As apparent from Fig. 2, the pupil responses started to clearly separate off at about 1,200-ms post-stimulus onset. This point in time occurs after the onset of the naming responses, which may raise the question of whether pupil responses reflected the onset of the behavioral responses (Simpson 1969). Such a question would have received an affirmative answer, if the pupil responses induced by the different distractors had the same pattern. But pupil responses increased, reduced, or

remained stable across distractors, a finding that is difficult to reconcile with the proposal that pupil changes were simply reflections of the onsets of naming responses. In contrast, a 400-ms peak or relatively lower amplitude showed no evident differences among conditions. This early peak was most likely unrelated to the meaning of the stimuli and could be best interpreted as mirroring the pupillary response to the attentional changes associated with stimulus appearance (cf. Richer and Beatty 1985).

Our finding that distractor meaning modulates pupillary sizes suggests that pupil responses are not indicators of generalized and indistinct conflicts in the input. On the contrary, pupil responses are also sensitive to semantic variables. This does not necessarily mean that the mechanisms specifically involved in semantic word processing could modulate pupil responses directly. The responsiveness that pupils exhibited for semantics could have a different source—for example, pupil responses could have been modulated by attentional mechanisms, which in turn were affected by word semantic properties. But the point is that even as indirect markers of semantic processing, pupil changes can provide a tool for investigating the neurocognitive correlates of word meaning processing. The pupillometry of Stroop effects offers a primary example of such approach.

Our analyses of the time courses of pupillary responses to word distractors have revealed two time signatures: an early peak at about 400 ms and a later peak at about 1,400 ms. In this respect, pupillometry shows a close resemblance with the ERPs correlates of the Stroop task, which also revealed two time-locked effects associated with the presentation of color-incongruent words, one at about 400 ms after stimulus onset, the other after the behavioral response (Liotti et al. 2000; Rebai et al. 1997; West 2003; West and Alain, 1999). The 'earlier' ERPs effect exhibits greater negativity over the frontal region of the scalp and appears to reflect the activity of neural generators in the cingulate cortex, particularly in the anterior cingulate cortex (ACC). Given the time accuracy of the ERPs recording, the 'later' ERPs effect unequivocally occurs after the onset of the naming responses and has been proposed to reflect post-response attentional control mechanisms (e.g., monitoring of performance and error detection; Carter et al. 1998; van Veen and Carter 2002). Do these time coincidences between pupillometry and ERPs arise because the two measures index similar processes? The differences observed between the two measures of Stroop effects seem to provide critical clues to this question. Pupillary responses were not modulated by distractors at 400 ms, unlike ERPs responses that demonstrated clear distractor effects at this point in time. This major discrepancy makes it unlikely that the early correlates of pupillary responses and ERPs index similar

processes. Distractor effects appeared later in pupillometry—at about 1,400 ms, thus after naming was initiated. We can easily reconcile these discrepancies, if we consider that ERPs are associated with faster physiological responses than pupil dilation, a difference reflecting the fact that it takes time for activation modulating autonomic arousal and originating within cortical areas (e.g., prefrontal and cingulate cortex; Matthews et al. 2004; Nagai et al. 2004) to spread through the brainstem and induce measurable changes in pupil size. Thus, the 400-ms ERPs correlate could be shifted in time in pupillometry, so that distractors generate noticeable changes in pupil size not earlier than about 1,400 ms. According to this hypothesis, pupil changes appearing after the naming responses most likely corresponded to brain processes that did occur at least half a second earlier, and therefore before the behavioral responses were generated. Moreover, the late, differential, pupillary responses to distractor meaning reflected processes implicated in the successful resolution of the cognitive conflict engendered by the distractors, rather than processes associated with post-response attentional control. Naturally, future results need to confirm that the late pupillary responses observed in Stroop tasks are the signature of semantic processing triggered by conflicting stimuli. Nevertheless, our findings clearly demonstrate the potential of pupillometry for charting the time course of neurocognitive processes. Even if pupillometry does not supply an almost instantaneous recording of brain correlates (unlike ERPs), it can still provide important insights on the temporal unfolding of neurocognitive mechanisms when proper delays are factored in.

It is significant to note here that the cingulate area responding to color-incongruent stimuli in neuroimaging studies of the Stroop task (e.g., Banich et al. 2000; Bench et al. 1993; Carter et al. 1998; MacDonald et al. 2000; Pardo et al. 1990) is also related to autonomic arousal (e.g., Brown et al. 2002; Matthews et al. 2004; Nagai et al. 2004), which in turn is implicated in pupil dilation. The involvement of identical or adjacent areas in the processing of incongruent stimuli and autonomic responses gives strong plausibility to the hypothesis that pupil size may vary as an effect of the congruency of the Stroop-like stimuli. More direct evidence pointing to this conclusion comes from an fMRI study in which a numerical variant of the Stroop task was used (Critchley et al. 2005). Pairs of numbers were presented side-by-side on a computer screen, and participants indicated as quickly as possible which number was larger (size of font being the interfering parameter; e.g., the small number being printed with a large font). MRI scans and pupil measurements were both recorded, while participants performed their size judgments. Findings implicated the cingulate gyrus in autonomic arousal, since activity in this area was correlated to

pupillary changes, especially in those trials where errors were produced. These data led to propose that the cognitive conflict engendered by incongruent stimuli engages considerable attentional resources and cognitive control in prefrontal and cingulate areas, which in turn would activate the autonomic arousal system thus determining pupillary dilations of size directly correlated to cognitive load (Critchley et al. 2005; Siegle et al. 2004, 2008).

To conclude, we found that changes in pupillary size are able to index word meaning processing. These data naturally raise the question as to whether pupillary effects can substitute verbal responses in studies where subjects cannot or should not produce a verbal response. It would provide unique opportunities to test word processing with adult speakers affected by word production deficits (e.g., aphasia) or in circumstances such as fMRI where overt naming can produce motion artifacts. But perhaps the most attractive application would involve infants and monkeys. Indeed, one of the most exciting developments of pupillometry in cognitive sciences relates to the finding that attention-induced changes in pupil diameter are reliably measurable with infants and children (Munsinger and Banks 1974; Karatekin et al. 2007; Chatham et al. 2009;) as well as with non-human primates (Iriki et al. 1996). Recently, researchers created ingenious paradigms to generate response uncertainty and cognitive conflict in the minds of infants and to record it with pupillometry (e.g., Gredebäck and Melinder 2010; Jackson and Sirois 2009). Given that our results indicate that pupillary measures of Stroop interference are positively related to the traditional measure of Stroop interference as revealed by RT, then it is hopeful that, in future studies, researchers will be equally resourceful in applying pupillometry to analogs of Stroop effects in pre/non-verbal participants.

References

Banich MT, Milham MP, Atchley R, Cohen NJ, Wszalek T et al (2000) fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. J Cogn Neurosci 12:988–1000

Barbur JL (2004) Learning from the pupil–studies of basic mechanisms and clinical applications. In: Chalupa LM, Werner JS (eds) The Visual Neurosciences. MIT Press, Cambridge

Beatty J (1982) Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychol Bull 91:276–292

Beatty J, Lucero-Wagoner B (2000) The pupillary system. In: Cacioppo JT, Tassinary LG, Berntson GC (eds) Handbook of Psychophysiology. Cambridge University Press, Cambridge

Bench C, Frith C, Grasby P, Friston K, Paulesu E, Frackowiak R et al (1993) Investigation of the functional anatomy of attention using the Stroop test. Neuropsychologia 31:907–922

Bijleveld E, Custers R, Aarts H (2009) The unconscious eye opener: pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. Psychol Sci 20:1313–1315

Bradley MM, Miccoli L, Escrig MA, Lang PJ (2007) The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 44:128–144

Brown GG, Kinderman SS, Siegle GJ, Granholm E, Wong EC, Buxton RB (1999) Brain activation and pupil response during covert performance of the Stroop color word task. J Int Neuropsychol Soc 5:308–319

Brown TL, Gore CL, Carr TH (2002) Visual attention and word recognition in Stroop color naming: is word recognition "automatic"? J Exp Psychol Gen 131:220–240

Burt JS (2002) Why do non-color words interfere with color naming. J Exp Psychol Hum Percept Perform 28:1019–1038

Carter CS, Beaver TS, Barch DM, Botvinick MM, Noll D, Cohen JD (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280:747–749

Chatham CH, Frank MJ, Munakata Y (2009) Pupillometric and behavioral markers of a development shift in the temporal dynamics of cognitive control. Proc Natl Acad Sci USA 106:5529–5533

Critchley HD, Tang J, Glaser D, Butterworth B, Dolan RJ (2005) Anterior cingulate activity during error and autonomic response. NeuroImage 27:885–895

Granholm E, Asarnow RF, Sarkin AJ, Dykes KL (1996) Pupillary responses index cognitive resource limitations. Psychophysiology 33:457–461

Gredebäck G, Melinder AMD (2010) Infants understanding of everyday social interactions: a dual process account. Cognition 114:197–206

Hess EH, Polt JM (1960) Pupil size as related to interest value of visual stimuli. Science 132:349–350

Hess EH, Polt JM (1964) Pupil size in relation to mental activity during simple problem solving. Science 140:1190–1192

Hoeks B, Levelt WJM (1993) Pupillary dilation as a measure of attention: a quantitative system analysis. Behav Res Methods Instrum Comput 25:16–26

Iriki A, Tanaka M, Iwamura Y (1996) Attention-induced neuronal activity in the monkey somatosensory cortex revealed by pupillometrics. Neurosci Res 25:173–181

Jackson I, Sirois S (2009) Infant cognition: going full factorial with pupil dilation. Dev Sci 12:670–679

Just MA, Carpenter PA (1993) The intensity dimension of thought: pupillometric indices of sentence processing. Can J Exp Psychol 47:310–339

Kahneman D (1973) Attention and Effort. Prentice-Hall, Engelwood Cliffs

Kahneman D, Beatty J (1966) Pupil diameter and load on memory. Science 154:1583–1585

Karatekin C, Couperous JW, Marcus DJ (2004) Attention allocation in the dual-task paradigm as measured through behavioural and psychophysiological responses. Psychophysiology 41:1–11

Karatekin C, Marcus DJ, Couperous JW (2007) Regulation of cognitive resources during sustained attention and working memory in 10-year-olds and adults. Psychophysiology 44:128–144

Klein GS (1964) Semantic power measured through the interference of words with color-naming. Am J Psychol 77:576–588

Koss M (1986) Pupillary dilation as an index of central nervous system α2-adrenoceptor activation. J Pharmacol Methods 15:1–19

Kristoffersen G (2000) The phonology of Norwegian. Oxford University Press, Oxford

Krüger F, Nuthmann A, van der Meer E (2001) Pupillometric indices of temporal order representation in semantic memory. Zeitschrift für Psycologie 209:402–415

Laeng B, Falkenberg L (2007) Women's pupillary responses to sexually significant others during the hormonal cycle. Horm Behav 52:520–530

Laeng B, Låg T, Brennen T (2005) Reduced Stroop interference for opponent colors may be due to input factors: evidence from individual differences and a neural network simulation. J Exp Psychol Hum Percept Perform 31:438–452

Laeng B, Waterloo K, Johnsen SH, Bakke SJ, Låg T, Simonsen SS, Høgsæt J (2007) The eyes remember it: oculography and pupillometry during recollection in three amnesic patients. J Cogn Neurosci 19(11):1888–1904

Libby WL, Lacey BC, Lacey JI (1973) Pupillary and cardiac activity during visual attention. Psychophysiology 10(3):270–294

Liotti M, Woldorff MG, Perez R, Mayberg HS (2000) An ERP study of the temporal course of Stroop color-word interference effect. Neuropsychologia 38:701–711

Loewenfeld I (1993) The pupil: anatomy, physiology, and clinical applications. Wayne State University Press, Detroit

Loftus GR, Masson MEJ (1994) Using confidence intervals in within-subject design. Psychon Bull Rev 1:476–490

MacDonald AW III, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulated cortex in cognitive control. Science 288:1835–1838

MacLeod CM (1991) Half a century of research on the Stroop effect: an integrative review. Psychol Bull 109:163–203

Matthews SC, Paulus MP, Simmons AN, Nelesen RA, Dimsdale JE (2004) Functional subdivision within anterior cingulate cortex and their relationship to autonomic nervous system function. Neuroimage 22:1151–1156

Munsinger H, Banks MS (1974) Pupillometry as a measure of visual sensitivity among infants, young children, and adults. Dev Psychol 10:677–682

Nagai Y, Critchley HD, Featherstone E, Trimble MR, Dolan RJ (2004) Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level SCL, a physiological account of a "default mode" of brain function. NeuroImage 22:243–251

Nuthmann A, van der Meer E (2005) Time's arrow and the pupillary response. Psychophysiology 42:306–317

Pardo JV, Pardo PJ, Janer KW, Raichle ME (1990) The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. Proc Natl Acad Sci USA 87:256–259

Paulsen HG, Laeng B (2006) Pupillometry of grapheme-color synaesthesia. Cortex 42:290–294

Peavler WS (1974) Pupil size, information overload, and performance differences. Psychophysiology 11:559–565

Piquado T, Isaacowitz D, Wingfield A (2010) Pupillometry as a measure of cognitive effort in younger and older adults. Psychophysiology 47:1–10

Pratt RW (1970) Cognitive processing of uncertainty: its effect on pupillary dilation and preference ratings. Percept Psychophys 8:193–198

Rebai M, Bernard C, Lannou J (1997) The Stroop's test evokes a negative brain potential, the N400. Int J Neurosci 91:85–94

Richardson DC, Dale R, Spivey MJ (2007) Eye movements in language and cognition: a brief introduction. In: Gonzalez-Marquez M, Mittelberg I, Coulson S, Spivey MJ (eds) Methods in cognitive linguistics. John Benjamins, Amsterdam, pp 323–344

Richer F, Beatty J (1985) Pupillary dilations in movement preparation and execution. Psychophysiology 22:204–207

Sara S (2009) The locus coeruleus and noradrenergic modulation of cognition. Nat Rev Neurosci 10:211–223

Siegle CJ, Steinhauer SR, Thase ME (2004) Pupillary assessment and computational modelling of the Stroop task in depression. Int J Psychophysiol 52:63–76

Siegle CJ, Ichikawa N, Steinhauer S (2008) Blink before and after you think: blinks occur prior to and following cognitive load indexed by pupillary responses. Psychophysiology 45:679–687

Simpson HM (1969) Effects of a task-relevant response on pupil size. Psychophysiology 6:115–121

Steinhauer SR, Condray R, Kasparek A (2000) Cognitive modulation of midbrain function: task-induced reduction of the pupillary light reflex. Int J Psychophysiol 39:21–30

Stelmack RM, Siddle DAT (1982) Pupillary dilation as an index of the orienting reflex. Psychophysiology 19:706–708

Stroop JR (1935) Studies of interference in serial verbal reactions. J Exp Psychol 28:643–661

Van der Meer E, Beyer R, Horn J, Foth M, Bornemann B, Ries J, Kramer J, Warmuth E, Heekeren HR, Wartenburger I (2010) Resource allocation and fluid intelligence: insights from pupillometry. Psychophysiology 47:158–169

van Veen V, Carter CS (2002) The time of action-monitoring processes in the anterior cingulate cortex. J Cogn Neurosci 14:593–602

Verney SP, Granholm E, Marshall SP (2004) Pupillary responses on the visual backward masking task reflect general cognitive ability. Int J Psychophysiol 52:23–36

Võ MLH, Jacobs AM, Kuchinke L, Hofmann M, Conrad M, Schacht A et al (2007) The coupling of emotion and cognition in the eye: introducing the pupil old/new effect. Psychophysiology 44:130–140

West R (2003) Neural correlates of cognitive control and conflict detection in the Stroop and digit-location tasks. Neuropsychologia 41:1122–1135

West R, Alain C (1999) Event-related neural activity associated with the Stroop task. Cogn Brain Res 8:157–164

Winn B, Whitaker D, Elliot DB (1993) Factors affecting light-adapted pupil size in normal human subjects. Invest Ophthalmol Vis Sci 95:1132–1136