No pupillometric evidence for effortful proactive control in the proportion-congruent Stroop paradigm

Journal:	*Quarterly Journal of Experimental Psychology*
Manuscript ID	QJE-SIP-23-149.R2
Manuscript Type:	Special Issue Paper
Date Submitted by the Author:	06-Dec-2023
Complete List of Authors:	Hasshim, Nabil; University of Salford Carruthers, Molly; De Montfort University Leicester FERRAND, Ludovic; CNRS, LAPSCO Augustinova , Maria ; Université de Rouen UFR Lettres et Sciences Humaines, Parris, Benjamin; Bournemouth University, Psychology
Keywords:	Cognitive control, Proportion congruency, Stroop task

DOI: 10.1177/17470218241235671

No pupillometric evidence for effortful proactive control in the proportion-congruent Stroop paradigm

Nabil Hasshim[1,2], Molly Carruthers[2], Ludovic Ferrand[3], Maria Augustinova[4] & Benjamin A. Parris[5]

[1] University of Salford, UK

[2] De Montfort University, UK

[3] Université Clermont Auvergne, CNRS, LAPSCO, Clermont-Ferrand, France

[4] Normandie Université, UNIROUEN, CRFDP, Rouen, France

[5] Bournemouth University, UK

Corresponding Author:

Nabil Hasshim

School of Health and Society, University of Salford

Email: nabil.hasshim@gmail.com

Authors' note

This research was supported by the Experimental Psychology Society via the EPS Small Grants Scheme. The writing of the report; and the process related to submission of this article for publication was additionally supported by Agence nationale de la recherche (ANR Grant ANR-19-CE28-0013) et Réseau d'Intérêt Normandie (RIN Tremplin Grant 19E00851).

The preregistered design and analysis plans are available at the following OSF registrations page: https://osf.io/syzhx/?view_only=21613b8e058941758e43681337e7aa8c. The raw data (.edf and .csv formats), analysis files (JASP file format) are available at the

DOI: 10.1177/17470218241235671

following OSF project page:

https://osf.io/bweg8/?view_only=36ab2ae24ad949d7ab09b26992fd9c77.

DOI: 10.1177/17470218241235671

Abstract

Cognitive control is the ability to allocate attention away from stimuli that are irrelevant to achieving a goal, towards stimuli that are. When conflict is anticipated, attention is biased in a global, top-down manner called proactive control and this effortful type of cognitive control is engaged before stimulus onset. The list-wise congruency proportion (LWPC) effect, where the Stroop congruency effect is reduced when there are more incongruent than congruent trials compared to vice versa, has been viewed as one of the prime signatures of this type of cognitive control. However, there has been recent debate about the extent to which this effect should be attributed to proactive control instead of alternative explanations such as simpler associative learning or reactive control. Thus, by using pupillometry (i.e., an indicator of cognitive effort), the present study investigated the extent to which LWPC effects result from effortful proactive control. Experiment 1 employed a classic proportion congruency manipulation while Experiment 2 replaced congruent trials with neutral trials to control for potential effects of associative learning. While in line with past findings, proportion congruency effects were obtained in response times of both experiments and pupillometry showed both proportion congruency and Stroop effects after stimulus onset, no differences in pupil sizes were found during the preparatory phase. Therefore, these results do not support the idea that the observed LWPC effects are due to participants engaging in effortful proactive control.

Keywords*: Cognitive control, Proportion congruency, Stroop task*

DOI: 10.1177/17470218241235671

No pupillometric evidence for effortful proactive control in the proportion-congruent paradigm

Cognitive control is the ability to orient our thoughts and actions towards an internal goal (Miller & Cohen, 2001). One of the most popular laboratory task that experimental psychologists use to study this ability is the Stroop task (Stroop, 1935; see MacLeod, 1991; Parris et al., 2021 for reviews) in which the goal is to identify the colour in which different colour-words are printed. Because for *incongruent* stimuli (e.g., the word 'blue' printed in green), attention needs to be shifted away from reading the word and applied to identifying the colour, poorer performance (e.g., slower responses, and more errors) is typically observed for these items as compared to control stimuli (e.g., *congruent* stimuli such as 'blue' printed in blue).

The efficiency of cognitive control is inferred from differences in magnitudes of this latter difference – referred to as the Stroop congruency effect such that smaller magnitudes are thought to reflect more efficient cognitive control (Braem et al., 2019). These differences can be observed both between participants (i.e., as a function of age for instance, see e.g., Bugg et al., 2007; Burca et al., 2022) and within participants. Indeed, magnitudes of the Stroop congruency effect are known to be substantially smaller when participants go through a block of trials which are mostly incongruent (e.g., 20% congruent and 80% incongruent) than when they go through a mostly congruent block (e.g., 80% congruent and 20% incongruent; see e.g., Lindsay & Jacoby, 1994; Logan & Zbrodoff, 1979).

This so-called list-wise congruency proportion effect (LWPC) along with other proportion congruency (PC) effects (see e.g., Bugg et al., 2008; Crump et al., 2006) suggests that cognitive control can be triggered in a top-down manner. Therefore, studies of these effects were crucial for subsequent theoretical development of cognitive control (e.g., Botvinick et al., 2001; Braver, et al., 2007; Norman & Shallice,

DOI: 10.1177/17470218241235671

1986; Posner, 2012). Recently however, there is a debate about the extent to which these effects can actually be accounted for in a simpler, perhaps a more parsimonious way (see below). Therefore, by shedding some additional light on list-wise congruency proportion effect (LWPC), the present study was aimed to contribute to this ongoing debate.

Alternative accounts to proactive control

Prominent theories of cognitive control often view LWPC (and other PC) effects as prime signatures of *proactive control* (e.g., Botvinick et al., 2001; Braver, et al., 2007; Norman & Shallice, 1986; Posner, 2012). This type of attentional control – triggered by participants' expectancies and motivations in response to different task environments – biases attention system in a global, top-down manner. For example, in an environment where conflicting information (e.g., an incongruent Stroop trial) is frequently encountered (in mostly incongruent or MI block), the conflict monitoring system (Botvinick et al., 2001) will signal for a shift towards greater proactive control (and away from reactive control) where more attentional resources to be utilised so as to aid task performance. The cognitive system will then react by engaging greater global cognitive control to bias attention away from the word, and towards the colour instead (Cheesman & Merikle, 1986; Lindsay & Jacoby, 1994; Lowe & Mitterer, 1982; West & Baylis, 1998). Since the influence of the word is consequently lessened, smaller magnitudes of the Stroop congruency effect are observed than in situations where conflict is not frequently encountered (i.e., in a mostly congruent or MC block) and the conflict monitoring system is not triggered.

However, there are also control accounts that do not assume a proactive component but posits the application of control can be stimulus driven instead. Much of the work comes from the demonstration of context-specific (e.g., fonts or

DOI: 10.1177/17470218241235671

locations, Bugg et al., 2008) and item-specific proportion congruency effects (see Bugg & Crump, 2012 for a review of PC effects). Although cognitive control is still assumed, these accounts state that it is applied at the trial level, and not sustained at the global level.

Since the frequencies of colour-word pairings in typical PC manipulations are not equal (e.g., in MC blocks, the colour-word appears in the corresponding colour most often), proponents of alternatives to cognitive control argue that this may explain LWPC and other PC effects instead. For example, Schmidt and Besner (2008) proposed a contingency learning explanation, driven by the fact that the frequencies of the colour-word pairs making up the Stroop stimuli are confounded in several PC paradigms (see also Schmidt, 2013, 2019). Because each colour-word is more frequently paired with its corresponding colour (e.g., ‘blue’ in blue) than the other colours (e.g., ‘blue’ in red, yellow, etc.), MC blocks lead to stronger (explicitly or implicitly) learned stimulus-response contingencies and aiding participants’ ability to predict responses to more frequent word/colour pairs.

The idea of how the imbalance of the word-colour pairings makes the word dimension informative is not a new one as it had previously been proposed by Dishon-Berkovits and Algom (2000). Unlike Schmidt and Besner’s account that works via pure associative learning, Dishon-Berkovits and Algom’s model works via adaptation. Since the colour-word correlation results in the word dimension of the Stroop stimuli being informative of which potential response is more likely, this encourages word reading, which is the process that we assume participants are trying to supress. In line with this reasoning, pairing frequencies of the stimuli has been shown to reduce (and in some cases eliminate) PC effects, and Stroop effects in general (see Algom & Chajut, 2019 and Schmidt, 2013; 2019 for reviews of the

DOI: 10.1177/17470218241235671

topic, Spinelli & Lupker, in press, and Hasshim & Parris, 2021) for a direct comparison of the two accounts).

In short, associative learning can explain PC effects such as LWPC without the need to evoke control accounts. Therefore, there has been growing acknowledgement that control processes are not always engaged or even necessary (see Algom & Chajut, 2019; Algom et al., 2022; Schmidt, 2013; 2019). In an attempt to reconcile these different explanations, Bugg (2014) tested the idea that top-down proactive control is a process that the cognitive system can engage, depending on the environmental context. This "associations as antagonists to top-down control" (also called "last-resort") account posits that proactive control does occur, but only in certain situations. Specifically, it was not observed when responses can be predicted via learned associations, but the cognitive system reverts to top-down control in situations when such reliable stimulus-response cues are unavailable.

Pertinent to the current research, Spinelli and Lupker (2021) modified the proportion congruency manipulation in LWPC by replacing congruent trials with neutral ones to isolate the effects of conflict within the task, as without congruent trials the word dimension never predicts the response. By demonstrating that in such an environment, global level control is still engaged, they provided further evidence that not only does proactive control exist, but it can be invoked even in case of unpredictable word-response relationships (but see Bugg, 2014 above). In their more recent work, Spinelli and Lupker (2022) provided additional evidence for the existence of cognitive control by demonstrating its engagement even in situations where contingency learning was encouraged.

To sum up, the aforementioned lines of work are asking the question of whether proactive cognitive control exists and if so, when exactly it is triggered and

DOI: 10.1177/17470218241235671

engaged. This question is still subject to much ongoing debate (see Abrahamse et al., 2016; Egner, 2008; and Henik, et al., 2018; for reviews on the general topic of cognitive control) with one of the reasons for this being that the arguments tend to be relatively circulatory (i.e., going back and forth from constant, conditional and no proactive control). One possible reason for this is the fact that the vast majority of studies mentioned simply infer cognitive control and its different features from the modulation of the Stroop congruency effects (although see e.g., Blais & Bunge, 2010 for the use of fMRI and West & Alain, 2000 for the use of EEG to address these issues). Since proactive control is thought to be a resource-demanding, engagement of effortful proactive control can be measured more directly and more sensitively via pupillometry. This was precisely the goal of the current study.

Current study

Pupillometry has previously been used as a measure of cognitive effort (Kahneman & Beattie, 1966; Laeng et al., 2012) in selective attention tasks such as Stroop paradigms (e.g., Hasshim & Parris, 2015; Hershman & Henik, 2019; Laeng et al., 2011) where more effort is typically exerted in trials that have more conflict (e.g., incongruent trials compared to congruent) leading to larger pupil sizes.

The current study is specifically looking into the effortful and proactive aspect of cognitive control, as described in the previous section, and past research have used pupillometry to achieve this goal. For example, Parris at al. (2021) showed that the attenuation of Stroop interference when participants were under post-hypnotic suggestion corresponded with larger pupil sizes, which was interpreted as the improvement in performance being due to participants engaging in more effortful control, a form of demand characteristic. While Parris et al. measured effort throughout the task, Chiew and Braver (2013) had a similar interest to the current

DOI: 10.1177/17470218241235671

study, in proactive control prior to target onset. They found pupil sizes within a pre-trial time window were larger on trials in which participants were expecting a reward for performing – demonstrating that pupillometry measures just before the onset of a target reflect top-down activation of effort.

Another example of the use of pupillometry in PC Stroop designs is Diede and Bugg (2017) who used pupillometry to specifically demonstrate context specific proportion congruency manipulation (CSPC), where congruent and incongruent trials each reliably appear in specific contexts. In this study using a Flanker task, targets appeared in locations where trials were MI or MC and indeed, larger pupil sizes were observed in the MI location[1] where more cognitive effort was expected. Somewhat similarly, the current study used pupillometry to measure effortful control in LWPC paradigms.

However, it is important to understand that the goal of the present study is not to simply extend results of Diede and Bugg (2017) to another PC paradigm (applied in the Stroop task). Indeed, CSPC manipulation used in their study provides evidence for control being exerted at the item level instead of (or in addition to) globally throughout the block as described in the proactive control accounts of LWPC effect. Since the application of proactive control prior to target onset cannot account for these latter effects, attentional processes are likely to be activated after the target is presented (i.e., after stimulus onset). The proactive account of LWPC effect on the other hand clearly anticipates effortful control to be engaged *before* the target is present (i.e., before stimulus onset, e.g., Botvinick et al., 2001; Braver et al., 2007; Kane & Engle, 2003), especially when conflict is anticipated (Braver et al., 2007;

[1] Trials in MI locations showed larger peak pupil sizes compared to MC. Average peak pupil sizes occurred ~800ms to ~1000ms after average response times.

DOI: 10.1177/17470218241235671

Gonthier et al., 2016; Spinelli & Lupker, 2021). Therefore, the current study investigated whether and the extent to which the expected smaller Stroop congruency effects MI blocks (as compared to MC blocks) were indeed preceded by effortful proactive control engaged during the preparatory phase (i.e., before target onset). To this end, in Experiment 1, participants' pupil sizes before target onset were compared in MI as compared to MC block of the classic LWPC paradigm using congruent and incongruent trials (e.g., Lindsay & Jacoby, 1994). In Experiment 2, neutral trials replaced congruent trials as per Spinelli and Lupker (2021), so the influence of stimulus-response contingencies and the informativeness of the word – operating in Experiment 1 – were both removed.

Experiment 1

Method

Ethical approval

Ethical approval for this study was granted by the Faculty Research Ethics Committee, Faculty of Health & Life Sciences, De Montfort University; Reference:3427.

Participants

44 individuals recruited from the university community participated in the experiment (data from 7 participants were excluded as they did not meet the pre-defined minimum accuracy of 90%) and received either course credit for their undergraduate course or £10 shopping vouchers. The intention was to recruit a minimum of 44 participants which was the suggested sample size calculated through the jpower module of jamovi software (The jamovi project, 2020). The parameters of a minimally

DOI: 10.1177/17470218241235671

interesting effect size (δ) of 0.5, power of 0.9, and type 1 error rate of 0.05 were based on the sizes of the effects of interest in previous laboratory Stroop research that utilised behavioural responses. This target sample size is larger than the sample sizes used in Blais and Bunge (2010) with 16 participants, and West and Alain (2000) with 18 participants, studies of similar designs that informed the current research. As mentioned above, both studies investigated similar research questions to the current experiments, but with the use of EEG and fMRI methodology.

While the sample size was estimated from behavioural measures of the PC effect, it is also larger than those reported in the earlier mentioned research that measured cognitive effort using pupillometry. Parris et al. (2021) reported a difference in pupil sizes averaged throughout trials with 16 participants, while Chiew and Braver (2013) showed pre-trial pupillary differences with 33 participants. If we were to assume that effortful proactive control was to be present in the current design and would have a similarly sized effect as these two studies, a Bayesian power analysis (estimating a minimum sample size required, if we expect to observe an effect of the same size as in the literature) suggests that the sample size required to show a BF>3 to be 6 participants and for a BF>6, to be 15 participants.

Apparatus

Stimuli were presented to the participants using a standard PC (screen dimensions: 530mm x 300mm) running Experiment Builder software (SR Research Ltd.), which was displayed on a colour monitor displaying at 1920px x 1080px with a refresh rate of 60Hz. Pupillary data consisting of pupil and corneal reflection, as well as sampling at 1000 Hz was recorded using the EyeLink 1000 Plus eye-tracker (SR Research Ltd.) running in monocular mode. Data was processed offline using Data Viewer (SR Research Ltd.) Blinks were automatically identified by the EyeLink Online Parser

DOI: 10.1177/17470218241235671

used by the EyeLink 1000 eye-tracker (SR Research). Data during blinks, along with those withing the 10 frames (10ms) of the onset and offset of a blink event were ignored as recommended by Mathôt and Vilotijević (2022).

Design

The experiment was a 2 (block type: mostly congruent vs. mostly incongruent) x 2 (trial condition: congruent vs. incongruent) fully within-participant design. In the MC block, 80% of the trials were congruent trials, while the remaining 20% were incongruent trials, with these proportions reversed in the MI block. The stimuli consisted of two sets of colours (see Material section below) and participants encountered a different set in each block. The presentation order of the MC and MI blocks, and the colour set used in each block was randomly determined using a four-sided die. The presentation order of the trials within each block was randomised by the computer.

Material

Two sets of colours and their corresponding words were used as targets for the experiment: yellow, green, and red in one set, and blue, pink, and white in the other. The visual angle of each of the words on the screen were as follows: yellow (2.08° x 0.78°), green (1.89° x 0.65°), red (1.04° x 0.59°); and blue (1.17° x 0.52°), pink (1.11° x 0.65°), white (1.95° x 0.65°). The RGB values of each colour were: yellow – (208, 255, 93), green (0, 255,0), red (255, 0, 0), blue (0, 0, 255), pink (204, 0, 204), and white (255, 255, 255)[2]. The background of the screen was black (0, 0, 0) throughout the experiment.

[2] Luminosity of each stimulus was measured by converting each stimulus (word/letter-string and the background of the screen) into greyscale and aggregating the pixels as a proportion between 0 (completely black) and 1 (completely white). The mean luminosity of the congruent, incongruent, and neutral (used in experiment 2) stimuli were 0.000825, 0.000708, and 0.000677 respectively.

DOI: 10.1177/17470218241235671

Participants encountering colours from different sets in the two blocks. Congruent trials were made up of words spelling out the colour it appears in, while incongruent trials were made up of a word appearing in one colour but spelling out a different colour. Each word appears in its congruent colour on congruent trials and can appear in either of the two other colours in the set. As mentioned above, Experiment 1 was meant to follow the classic PC paradigm that does not control for stimulus contingencies. Indeed, each word stimulus appear more (less) often in its corresponding colour during the MC (MI) blocks. All visual stimuli were presented in the centre of the screen.

Procedure

Participants performed a manual version of the Stroop task, where they were asked to respond to the target by pressing keys on a full-sized QWERTY keyboard corresponding to the colour. The keys used were the1, 2, and 3, keys of the number pad of a standard keyboard. During eye-tracking participants positioned their heads on a headrest placed at an eye to screen distance of 880mm.

For the experiment, participants underwent 330 trials. In the first half of the experiment, participants underwent a block consisting of 15 practice trials (either MC or MI) followed by the corresponding experimental block of 150 trials, with the stimuli made up of one of the two aforementioned sets. The second half of the experiment repeats this process with the second set of stimuli presented. A break was administered in between the four blocks, and each block was preceded by a 5-point calibration and validation.

Each trial began with a 500ms blank screen, followed by a grey fixation cross presented for 800ms, before being replaced by the target. The target stayed in view for 500ms followed by a blank screen until the participant responded up till a

DOI: 10.1177/17470218241235671

maximum of 1000ms, when the trial would then be classified as a non-response and recycled as an upcoming trial. Another blank screen followed which was displayed for 1000ms. The time-windows for sampling pupil size are defined as follows: baseline pupil size was the average pupil size within 500ms before the onset of the fixation, and proactive control was the average pupil size within 500ms before target onset (see Figure 1). Since proactive control is a tonic response that is assumed to be activated generally during the block, pupillometry readings were measured as the average pupil size during the proactive time window.

To reduce noise in the pupillary measure, recommendations from Mathôt et al. (2018) that could be implemented, were reflected in the procedure. For example, all stimuli were presented in the centre of the screen to minimise artefacts from pupil foreshortening and to ensure that participants were looking at the centre of the screen during the critical time windows. On trials where the total fixation time on the fixation cross (as determined by an invisible circular interest area subtending 2.1°) was <300ms, a drift correction was conducted and the trial restarted as this indicated that either the participant was not focusing on the centre of the screen or accurate calibration was not maintained. Recalibration and validation were conducted in more severe cases. Furthermore, to reduce pupillary effects due to differences in exposure time to the visual stimuli, the duration for which each stimulus was presented was kept consistent for all trials.

The details of the procedure were carefully considered to accommodate the use of pupillometry in the PC Stroop paradigm. While a longer inter-trial-interval (ITI) is recommended to allow for pupil dilation to go back to a baseline level after each trial, long ITIs (e.g., > 2000ms) have been shown to be detrimental to performance in the Stroop task by reducing participants' ability to focus on the goal/task (De Jong et

DOI: 10.1177/17470218241235671

al., 1999; see also e.g., Augustinova et al., 2018; Parris, 2014). This would arguably affect maintaining proactive control since it would need to be effortfully sustained over a longer period in between trials. Thus, an interval of 1500ms was chosen as a compromise between these two concerns (also note that the pupillometry study of Chiew and Braver, 2013 reported only minimal behavioural differences between ITIs of 250 vs 4000ms). A baseline correction was also done to reduce carryover effects from the previous trial, where pupillometry was taken as the percentage change compared to the baseline period (Mathôt et al., 2018). To further elicit potential proactive control, the fixation cross acts as an implicit cue to anticipate the impending trial. The potential effects of these design choices are discussed with the findings.[3]

--- Insert Figure 1 about here ---

The preregistered design and analysis plans are available at the following OSF registrations page: https://osf.io/syzhx/?view_only=21613b8e058941758e43681337e7aa8c. The raw data (.edf and .csv formats), analysis files (JASP file format) are available at the following OSF project page: https://osf.io/bweg8/?view_only=36ab2ae24ad949d7ab09b26992fd9c77.

Results

Trials consisting of incorrect responses, or correct responses where responses were not within 200-1500 ms of target onset, or with more than 30% missing pupil data

[3] Although the 300ms period between the Baseline and Proactive periods was in place to minimise the effects of pupillary light reflex to the fixation cross, it is still a possibility. See the online Supplementary Material which shows the pattern for the pupillometry measures throughout the experiment that suggests that it is not due to pupillary light reflex.

DOI: 10.1177/17470218241235671

were excluded from the analysis (only gaze data within the invisible central interest area defined above were considered 'valid'). Also, data from 7 participants who did not attain at least 90% accuracy was excluded from the analysis. Since the main analyses of interest are the comparison of pupil sizes between the MI and MC blocks with two directly competing predictions (i.e., proactive control accounts predict larger pupil sizes in MI, while other accounts do not), a Bayesian t-test was performed to complement the frequentist analysis. Bayes factors were calculated in JASP (JASP Team, 2022), using a half-Cauchy prior distribution (i.e., pupil size in MI>MC) scaled to 0.707 and robustness regions were reported to determine whether any interpretation from the Bayes factors were sensitive to the choice of prior distribution. The default prior specifications in JASP were also used for the Bayesian ANOVA analyses (i.e., r-scale fixed, random, and covariates values of 0.5, 1, and 0.354 respectively). For replicability of the analyses reported in this manuscript, the random seeds were specified to be '999'. Table 1 summarises the RT and error data for both experiments.

--- Insert Table 1 about here ---

Response times

Response times were analysed to determine whether the pattern of results was in line with the expectations from research that use similar techniques, namely the classic LWPC effect.

The 2 (block: MC vs MI) × 2 (trial condition: congruent vs incongruent) analysis of variance (ANOVA) of RT showed the block × trial condition interaction, $F(1, 36) = 34.94$, $p < .001$, $\eta_p^2 = 0.49$, and main effect of trial condition, $F(1, 36) = 58.38$, $p < .001$, $\eta_p^2 = 0.62$, to be statistically significant. This meant that while the

DOI: 10.1177/17470218241235671

Stroop congruency effect was statistically significant in both MI, $t(36) = 3.85$, $p < .001$, $d = 0.63$, and MC, $t(36) = 7.97$, $p < .001$, $d = 1.31$, blocks, it was significantly reduced in the former; a result that is consistent with the classic LWPC effect (see figure 2 for a visualisation of the RTs).The main effect of block was not statistically significant, $F(1, 36) = 0.551$, $p = .463$, $\eta_p^2 = 0.015$.

The corresponding Bayesian ANOVA analysis showed that the model which included block, trial condition, and the interaction between the two, to be the best model with $BF_{10} = 8.0e+9$ when compared to the null model. The analysis of effects comparing this model to one without the interaction showed decisive evidence in favour of including it with a $BF_{inclusion}$ of 407.08.

--- Insert Figure 2 about here ---

Error rates

The 2 × 2 ANOVA for error rates revealed a statistically significant block × trial condition interaction $F(1, 36) = 8.74$, $p = .005$, $\eta_p^2 = 0.20$, and main effect of trial condition, $F(1, 36) = 22.85$, $p < .001$, $\eta_p^2 = 0.39$, while the main effect of block was non-significant, $F(1, 36) = 1.31$, $p = .260$, $\eta_p^2 = 0.035$. This meant that although the Stroop congruency effect was significant in the error rates for both MC, $t(36) = 4.36$, $p < .001$, $d = 0.717$, and MI, $t(36) = 2.41$, $p = .021$, $d = 0.397$, blocks, the effect in the former was larger. The results from the error rates mirrored that of the RT analysis.

Pre-target Pupillometry

The comparison of preparatory pupil diameter between the two blocks is the primary analysis of interest of this research. The results showed that the difference in dilation between MI (0.44%) and MC (0.66%) blocks were statistically non-significant, $t(36) =$

DOI: 10.1177/17470218241235671

-0.690, p = .495, d = -0.113, BF_{01} = 8.91. The Bayes factor showed evidence supporting the null hypothesis and robustness check showed that for the *BF* to fall below 3, an interpretation of only anecdotal support for the null, the prior distribution would have to be scaled to be narrower than 0.193[4].

Post-response Pupillometry (exploratory analysis)

Although congruency effects in pupillometry are well established, predictions for reactive control accounts on LWPC effects have been less researched. Thus, post-response pupil sizes[5] were examined as exploratory analyses. The average pupil size during the 1000ms blank screen at the end of the trial (see “Post-response” in Figure 1) were sampled for each of the four conditions and compared, with the same baseline as the one used in the pre-response measures. This post-response time-window should be sensitive to congruency effects, following the time-series findings of Hershman and Henik (2019) who indicated pupil effects begin to emerge around 500ms (between neutral and incongruent trials) and 1000ms (between congruent and incongruent trials) post stimulus onset.

[4] It is possible that proactive control could be affecting pupil sizes during the baseline period as well if it is engaged throughout a block and thus occluding any effect. Although correction is recommended to reduce noise and improve statistical power, this is a valid concern. To check this, we compared the raw pupil sizes (pixels recorded by the eye-tracker) during the baseline period which showed no statistical difference between the blocks – Experiment 1: $t(36) = 0.814$, p = .421 d = 0.134 BF_{01} = 4.16; Experiment 2: $t(35) = 0.342$, p = .734 d = 0.057, BF_{01} = 5.29. This suggests that the pupil sizes during baseline were not different across blocks.

[5] The original pre-registration of this analysis indicated that the time-window for the exploratory post-target analysis starts immediately at target onset (i.e., while the target is still visible). However, while conducting the study, we concluded that this may not be a valid comparison since the pupillometry readings could be affected by the different luminosity of the stimuli on the screen. The time-window of data analysed and reported in the main manuscript excludes the period where any stimuli was visible. For transparency, analyses of pupil size in the time window between target onset and response showed the main effects and interactions to be non-significant for both experiments (all *p*s > .271). Bayesian t-tests comparing each of the different conditions showed support for the null on all comparisons (BF_{01}s > 3) apart from the difference between the neutral and incongruent trials in the MN block of Experiment 2, where BF_{01} = 2.97. This suggests that the different stimuli (colour and words) used in the different conditions did not differently affect pupil sizes. Observing pupillary Stroop effects only after response is also consistent with findings in the literature (e.g., Laeng et al., 2012; Hershman et al., 2023) showing the effects to only emerge after response.

DOI: 10.1177/17470218241235671

As this was not the main analysis of interest and the design of the experiment was geared towards the pre-target analysis, it should be noted that some conditions (e.g., congruent trials in a Mostly Incongruent block) have a maximum of 30 trials presented to participants which meant that some conditions had very few valid observations. Thus, in addition to the trial and participant exclusion criteria that were previously specified, participants who did not have at least 10 valid trials for each condition (a further 7 participants) were also excluded from the analyses[6].

The 2 (block: MC vs MI) × 2 (trial condition: congruent vs incongruent) ANOVAs of the mean pupil size showed the main effect of block to be statistically significant, $F(1, 29) = 4.66$, $p = .039$, $\eta_p^2 = 0.14$, as was the main effect of trial condition, $F(1, 29) = 9.61$, $p = .004$, $\eta_p^2 = 0.25$. This suggests larger post-response pupil dilation occurred during the MC block, and on incongruent trials. The block × trial condition interaction was non-significant, $F(1, 29) = 0.834$, $p = .369$, $\eta_p^2 = 0.028$. The Bayesian analysis indicated the model containing the two factors being the best model, with a BF_{10} of 12.05 when compared to the null model. Compared to the null model, models including only the block; only trial condition; and both factors and their interaction returned BFs_{10} of 4.69, 2.13, and 4.12 respectively.

--- Insert Figure 3 about here ---

Time-course Pupillary Changes

More recent research using pupillometry has moved to favour temporal analysis of pupil size changes (e.g., Hershman & Henik, 2019; Hershman et al., 2023) as it has several advantages, such as revealing when effects differ in their onset instead of

[6] Analyses that included these participants showed similar patterns, although only the congruency effect in Experiment 2 was statistically significant.

DOI: 10.1177/17470218241235671

magnitude, or those that only appear in a very small time-period compared to the overall time window that is analysed. However, these are not the characteristics of the proactive control effect that the current research endeavours to detect, but instead the effortful proactive control is hypothesised to be sustained throughout the preparatory period. Thus taking the average pupil sizes across the specified time-windows were preferred instead.

Nonetheless, it is possible that the temporal dynamics might reveal additional insights beyond what was hypothesised. Thus, Figure 4 shows visualisations of pupil size during the critical pre-target period, from the two experiments. The initial pupil size is larger than baseline, which is likely attributed to a phasic response to the fixation cross. The overall pattern does not suggest any hidden small effects within the time-period. Although there is a visual difference between the two lines, note that the analyses showed no significant differences between the blocks, and that the pattern on neither graph is consistent with the hypothesis of generally larger pupil sizes in the MI condition.

--- Insert Figure 4 about here ---

Discussion

The RT data from the experiment replicated the classic PC effect, where Stroop interference was smaller in the MI block compared to the MC block. The post-response pupillometry analysis showed a congruency effect of larger pupil dilation after accurately responding to incongruent trials compared to congruent ones, replicating typical pupillary Stroop effects. The main effect of block was also significant with generally larger post-response pupil dilation during the MC block which is the opposite of what would be expected if more effort was exerted in the MI

DOI: 10.1177/17470218241235671

blocks. It is possible that post-response pupil sizes reflect the amount of conflict encountered in the trial. That is, pupil Stroop effects are commonly observed ~500ms-600ms after participants make a response (see e.g., Hershman & Henik, 2019; 2020; Laeng et al., 2011), indicating that pupils reflect a delayed reaction to the conflict encountered when attempting to respond (Simpson, 1969), and thus pupil sizes would be overall larger in the MC block in which the RT congruency effect was larger. However, if pupil sizes were a delayed reflection of pre-response conflict encountered, we might expect to see a main effect of block in the RT data, but this was not observed (see also the lack of main effects in the pupil data in Experiment 2). Thus, unfortunately, the results from this exploratory analysis were not consistent with either account.

For the main analysis of interest, the pre-target pupillometry data showed evidence for no difference in pupil dilation during the preparatory phase between MI and MC blocks. This means that while the LWPC effect was observed in the behavioural data, it was not accompanied by a pupillary marker of effortful attention control in anticipation of conflict.

While a pupillary effect would have suggested that proactive control is still engaged in the classic LWPC paradigm despite the associative learning confounds highlighted earlier, the results from Experiment 1 do not necessarily provide evidence against the proactive control account. As proponents of proactive control accounts have argued, it might be a process that only emerges when other more effective strategies are unavailable (e.g., Bugg, 2014; Spinelli & Lupker, 2021).

DOI: 10.1177/17470218241235671

Experiment 2

Experiment 2 aimed at tackling one of the major associative learning confounds found in the classic LWPC paradigm, the effect of word-response contingency (Schmidt, 2013, 2019). This was done by modifying the design of the experiment in two main ways. Firstly, the congruent trials were used in Experiment 1 were replaced by neutral trials (e.g., string of 'ssss' in red). As suggested by Spinelli and Lupker (2021), the advantage of utilising neutral trials over congruent ones in this context is that the former reduces the informativeness of the word/letter-string and negates any advantage of gleaning information from the text. Additionally, the effect of word-response contingency was further controlled for by mapping each (non)word equally often to two of the response options (and never to the third). This meant that for each letter-string and word making up the neutral and incongruent trials, there was an equal probability of it appearing in one of two colours (and never the third).

Method

Participants

41 individuals recruited from the university community participated in the experiment (data from 5 participants were excluded as they did not meet the pre-defined minimum accuracy of 90%) and received either course credit for their undergraduate course or £10 shopping vouchers.

Design, Procedure, & Material

The design and procedure were identical to that of Experiment 1. The material was similar except that that congruent trials were replaced by repeated-letter neutral trials. Thus, instead of an MC block, participants went through a mostly neutral (MN) block. The repeated-letter sequences were zzzzzz (2.28° x 0.46° of visual angle),

DOI: 10.1177/17470218241235671

xxx (1.17° x 0.39°), jjjjj (0.91° x 0.78°) for one half of the trials, and the other half involved the targets ssss (1.30° x 0.39°), qqqq (1.43° x 0.52°), and hhhhh (1.69° x 0.52°).

To control for contingency effects, each repeated-letter string only appeared in two of the three colours, and equally often in each (e.g., ‘xxx’ appeared in green and yellow equally often, but never in red; ‘jjjjj’ appeared equally often in yellow and red, but never green; etc.). This mirrors the properties of incongruent trials.

Results

Data were processed in the same way as in Experiment 1, with data from 5 participants excluded for not meeting the 90% minimum accuracy threshold.

Response times

The 2 (block: MN vs MI) × 2 (trial condition: neutral vs incongruent) ANOVA of RT showed the block × trial condition interaction, $F(1, 35) = 25.09$, $p < .001$, $\eta_p^2 = 0.42$, and main effect of trial condition, $F(1, 35) = 44.15$, $p < .001$, $\eta_p^2 = 0.56$, to be statistically significant. The main effect of block was not statistically significant, $F(1, 35) = 0.314$, $p = .579$, $\eta_p^2 = 0.009$. This meant that while the Stroop interference effect was statistically significant in both MI, $t(35) = 2.49$, $p = .018$, $d = 0.42$, and MN, $t(35) = 6.76$, $p < .001$, $d = 1.13$, blocks, it was significantly reduced in the former. Similar to Experiment 1 this is consistent with the classic LWPC effect (see Figure 5 for descriptive statistics).

The Bayesian ANOVA analysis revealed the model which included block, trial condition, and the interaction between the two, to be the best model with BF_{10} = 99658.60 when compared to the null model. The analysis of effects comparing this model to one without the interaction showed strong evidence in favour of including it with a $BF_{inclusion}$ of 44.03.

DOI: 10.1177/17470218241235671

--- Insert Figure 5 about here ---

Error rates

Unlike the RT data, none of the effects in the error rate analysis were statistically significant, interaction: $F(1, 35) = 3.33$, $p = .077$, $\eta_p^2 = 0.087$, main effect of block: $F(1, 35) = 0.070$, $p = .794$, $\eta_p^2 = 0.002$, and trial condition: $F(1, 35) = 0.732$, $p = .401$, $\eta_p^2 = 0.020$. This could be due to lower error rates in the incongruent trials, compared the incongruent trials in Experient 1.

Pre-target Pupillometry

The comparison of preparatory pupil sizes showed that difference in dilation between MI (-0.01%) and MN (0.15%) were statistically non-significant, $t(35) = -0.468$, $p = .639$, $d = -0.078$, $BF_{01} = 7.72$. Similar to the results from Experiment 1, the Bayes factor showed evidence supporting the null hypothesis with a robustness check showing that for the *BF* to fall below 3, the prior distribution would have to be scaled to be narrower than 0.228.

Post-response Pupillometry (exploratory analysis)

Following the same exclusion criteria as in Experiment 1, a further four participants were excluded from this analysis. The 2 (block: MC/MN vs MI) × 2 (trial condition: congruent vs incongruent) ANOVAs of the mean pupil size showed only the main effect of trial-condition was statistically significant, $F(1, 31) = 8.12$, $p = .008$, $\eta_p^2 = 0.21$, with incongruent trials displaying larger pupil dilations compared to neutral trials. Neither the main effect of block, $F(1, 31) = 1.44$, $p = .240$, $\eta_p^2 = 0.044$, nor the block × trial-condition interaction, $F(1, 31) = 1.24$, $p = .274$, $\eta_p^2 = 0.038$, were statistically significant. Although the pattern of results is similar to that of Experiment

DOI: 10.1177/17470218241235671

1 (see Figure 6), the main effect of block was non-significant (unlike in Experiment 1) while the Bayesian analysis indicated the null model to be the best model with a model that included only trial condition having a BF_{10} = 0.935. This suggests that this exploratory analysis was underpowered and conclusions should not be drawn from this specific analysis.

--- Insert Figure 6 about here ---

Discussion

The primary results of interest from Experiment 2 closely followed those of Experiment 1, showing that although the LWPC effect was reflected in participants' RT, where the Stroop effect was larger in the MI compared to the MN block (replicating the results of Spinelli & Lupker, 2021), this was not accompanied by larger pupil dilation in the former. The analysis of the pupillometry data showed evidence for the null hypothesis. However, it should be noted that in the exploratory post-response pupillometry analysis, although trial congruency effects were shown in the frequentist analysis, the data was not sensitive enough to show similar evidence in the Bayesian analysis.

General discussion

The current research investigated whether there is evidence of effortful control of attention being proactively applied when conflict is expected in a selective attention task. Following previous work utilising pupillometry to evidence proactive control (Chiew & Braver, 2013; Parris et al., 2021), the intention was to demonstrate a putative marker of proactive control in the time-frame just before the presentation of the target. If the attention control mechanism is engaged effortfully in anticipation of

DOI: 10.1177/17470218241235671

conflict, larger pupil sizes were expected before the Stroop stimulus was presented during the MI block, where most of the trials were conflicting – providing convincing evidence for proactive control. We did not observe this effect in our data.

In both experiments, while RTs displayed the classic LWPC effect of a smaller Stroop congruency effect in the MI block, no pupillary effects were observed prior to target onset. Experiment 1 used the typical LWPC paradigm of comparing an MI block to an MC block (Lindsay & Jacoby, 1994), while Experiment 2 aimed at addressing the potential confound of contingency effects by comparing an MI block to an MN block. The results in the RT data are consistent with Botvinick et al. (2001)'s model of proactive control since replacing congruent trials with neutral trials will not affect the PC effect because the conflict monitoring module works via sensitivity to the frequency of conflict trials (i.e., incongruent). The pattern of results is also consistent with Tzelgov et al. (1992) who found larger interference effects with increased proportion of neutral trials. However, the pupillometry results failed to detect an accompanying marker for increased effort when participants were anticipating target onset in the MI blocks. Indeed, descriptive statistics showed that the amount of pupil dilation in general to be very small and even in the opposite direction (albeit non-significant) to that which proactive control accounts would predict. Additionally, the Bayes factors for both experiments favoured evidence against the hypothesis of larger pupil sizes in the MI (as compared to MC/MN) block.

Interestingly, it is notable that the exploratory analysis showed that the pattern of results from the (post-response) pupil sizes data do not mirror that of participants RTs (non-significant block × trial condition interaction in pupil sizes) or conform to the expectation of more effort required in the MI block. This suggests that the post-response pupil size might be sensitive to block level effects in addition to trial-level

DOI: 10.1177/17470218241235671

ones, something that future research using an appropriately powered design should explore.

Limitations of the current research

In isolation, interpretation of the findings from the two experiments seems straightforward: while the LWPC effect was observed in the RTs, the absence of pupillary differences in the preparatory phase suggests that the effect was not due to more proactive control processes being engaged when conflict was more likely. However, although demonstrating an effect would have provided evidence for proactive control, the converse does not necessarily mean evidence against it as even though pupillary effects have been used in similar investigations (e.g., Chiew & Braver, 2013; Parris et al, 2021) they might not be sensitive enough to index effortful proactive control in the current study. Relatedly, the lack of effects could also be due to a lack of power for the analysis. However, the sample sizes were larger than studies that demonstrated pupillometry effects comparable to the ones of interest, and the BFs reported do not suggest an issue with sensitivity. Even if the results were to be taken at face-value i.e., evidence against greater effort, the only conclusion that can be made is that the mechanism (control or otherwise) behind the LWPC effect does not involve sustained effort.

Another potential limitation of the present study is the use of only manual (button-press) responses. Larger Stroop effects are typically found in vocal, compared to manual, versions of the task (e.g., Augustinova et al., 2019; White, 1969) and can differently affect processes within Stroop task performance (e.g., Sharma & McKenna, 1998). However, behavioural LWPC effects were observed in both experiments as were the typical pupillary Stroop effects, suggesting that the paradigm did manage to capture the effects of interest. Although there is no strong

DOI: 10.1177/17470218241235671

indication that different anticipatory control mechanisms should be involved when the subsequent response is vocal vs. manual, response modality should be something that future research should consider as well.

Also, and importantly, control accounts (such as Braver et al., 2007 and Crump et al., 2006) do highlight control that is applied reactively during the task. Somewhat in line with reactive control, the secondary, exploratory analyses of post-response pupil sizes hint at the influence of proportion congruency on the effort applied after the target is presented. However, the data from the current study, which was primarily aimed at detecting the larger block-level effects, were not sensitive enough to examine the interactive effects of trial condition and proportion congruency on post-response pupil dilation. Given this important limitation, further research is also required to tease apart these processes that can independently affect pupil sizes. In doing so, these studies also need to integrate pupillometry to other variations that target reactive control (e.g., context-specific PC, Crump et al., 2006).

Conclusions

The current research intended to examine often assumed, but largely untested idea that effortful, proactive control is engaged prior to the presentation of stimulus in the LWPC Stroop paradigm. This was done using pupillometry as a direct measure of cognitive effort during the preparatory phase of the task. It is important to acknowledge that the rich literature within the wider debate about the role of control in the PC paradigm has involved more elaborate designs, the current research is an initial exploration utilising pupillometry, and any conclusions are limited to the classic version of the task. Therefore, it is hoped that the results from the two experiments demonstrate a simple point: the nature of the processes that results in the LWPC

DOI: 10.1177/17470218241235671

effect, at least for those observed in the classic PC paradigm, was not shown to involve *effortful* control engaged *in anticipation* of conflict.

Supplementary Material

The Supplementary Material is available at: qjep.sagepub.com

DOI: 10.1177/17470218241235671

References

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142*(7), 693 – 728.

Algom, D., & Chajut, E. (2019). Reclaiming the Stroop effect back from control to input-driven attention and perception. *Frontiers in Psychology*, *10*, 1683.

Algom, D., Fitousi, D., & Chajut, E. (2022). Can the Stroop effect serve as the gold standard of conflict monitoring and control? A conceptual critique. *Memory & Cognition*, *50*(5), 883-897.

Augustinova, M., Parris, B. A., & Ferrand, L. (2019) The Loci of Stroop Interference and Facilitation. Effects with Manual and Vocal Responses. *Frontiers in Psychology*, *10*, 1786

Augustinova, M., Silvert, L., Spatola, N., & Ferrand, L. (2018). Further investigation of distinct components of Stroop interference and of their reduction by short response-stimulus intervals. *Acta Psychologica, 189*, 54-62.

Blais, C., & Bunge, S. (2010). Behavioral and neural evidence for item-specific performance monitoring. *Journal of Cognitive Neuroscience*, *22*(12), 2758-2767.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624 - 652.

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J., Weissman, D. H., Notebaert, W., & Egner, T. (2019). Measuring adaptive control in conflict tasks. *Trends in Cognitive Sciences*, *23*(9), 769-783.

DOI: 10.1177/17470218241235671

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. *Variation in Working Memory*, 76 - 106.

Bugg, J. M. (2014). Conflict-triggered top-down control: Default mode, last resort, or no such thing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 567 - 587.

Bugg, J. M., DeLosh, E. L., Davalos, D. B., & Davis, H. P. (2007). Age differences in Stroop interference: Contributions of general slowing and task-specific deficits. *Aging, Neuropsychology, and Cognition*, *14*(2), 155-167.

Bugg, J. M., & Hutchison, K. A. (2013). Converging evidence for control of color–word Stroop interference at the item level. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 433 - 449.

Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition*, *36*(8), 1484-1494.

Burca, M., Chausse, P., Ferrand, L., Parris, B. A., & Augustinova, M. (2022). Some further clarifications on age-related differences in the Stroop task: New evidence from the two-to-one Stroop paradigm. *Psychonomic Bulletin & Review*, *29*(2), 492-500.

Cheesman, J., Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, 40(4), 343–367.

Chiew, K. S., & Braver, T. S. (2013). Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry. *Frontiers in Psychology*, *4*, 15.

DOI: 10.1177/17470218241235671

Crump, M. J., Gong, Z., & Milliken, B. (2006). The context-specific proportion congruent Stroop effect: Location as a contextual cue. *Psychonomic Bulletin & Review*, *13*(2), 316-321.

De Jong, R., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: Dissociable causes of interference effects in conflict situations. *Acta Psychologica*, *101*(2-3), 379-394.

Diede, N. T., & Bugg, J. M. (2017). Cognitive effort is modulated outside of the explicit awareness of conflict frequency: Evidence from pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 824 - 835.

Dishon-Berkovits, M., & Algom, D. (2000). The Stroop effect: It is not the robust phenomenon that you have thought it to be. *Memory & Cognition*, *28*(8), 1437-1449.

Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends in cognitive sciences*, *12*(10), 374-380.

Egner, T. (2014). Creatures of habit (and control): A multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology*, *5*, 1247.

Egner, T. (2017). Past, present, and future of the congruency sequence effect as an index of cognitive control. In *The Wiley handbook of cognitive control* (pp. 64-78). John Wiley & Sons.

Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. *Memory & Cognition*, *44*(5), 778-788.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480 - 506.

DOI: 10.1177/17470218241235671

Hasshim, N., & Parris, B. A. (2015). Assessing stimulus–stimulus (semantic) conflict in the Stroop task using saccadic two-to-one color response mapping and preresponse pupillary measures. *Attention, Perception, & Psychophysics*, *77*(8), 2601-2610.

Hasshim, N., & Parris, B. A. (2021). The role of contingency and correlation in the Stroop task. *Quarterly Journal of Experimental Psychology*, *74*(10), 1657-1668.

Henik, A., Bugg, J. M., & Goldfarb, L. (2018). Inspired by the past and looking to the future of the Stroop effect [Editorial]. *Acta Psychologica, 189,* 1–3.

Hershman, R., & Henik, A. (2019). Dissociation between reaction time and pupil dilation in the Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1899 - 1909.

Hershman, R., Milshtein, D., & Henik, A. (2023). The contribution of temporal analysis of pupillometry measurements to cognitive research. *Psychological Research*, *87*(1), 28-42.

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*(3), 638-644.

JASP Team (2022). JASP (Version 0.16.3)[Computer software]

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583-1585.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132(1),* 47–70.

DOI: 10.1177/17470218241235671

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary stroop effects. *Cognitive Processing*, *12*(1), 13-21.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious?. *Perspectives on Psychological Science*, *7*(1), 18-27.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: the relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 219 - 234.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & cognition*, *7*(3), 166-174.

Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided attention in a Stroop task. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 36*(4), 684–700.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin*, *109*(2), 163 - 203.

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94-106.

Mathôt, S., & Vilotijević, A. (2022). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*, 1-23.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*(1), 167-202.

Norman, D. A., & Shallice, T. (1986). Attention to action. In *Consciousness and self-regulation* (pp. 1-18). Springer, Boston, MA.

DOI: 10.1177/17470218241235671

Parris, B. A., Hasshim, N., & Dienes, Z. (2021). Look into my eyes: Pupillometry reveals that a post-hypnotic suggestion for word blindness reduces Stroop interference by marshalling greater effortful control. *European Journal of Neuroscience*, *53*(8), 2819-2834.

Parris, B. A., Hasshim, N., Wadsley, M., Augustinova, M., & Ferrand, L. (2021). The loci of Stroop effects: a critical review of methods and evidence for levels of processing contributing to color-word Stroop effects and the implications for the loci of attentional selection. *Psychological Research*, 1-25.

Posner, M. I. (2012). Attentional networks and consciousness. *Frontiers in Psychology*, *3*, 64.

Schmidt, J. R. (2013). The Parallel Episodic Processing (PEP) model: Dissociating contingency and conflict adaptation in the item-specific proportion congruent paradigm. *Acta Psychologica, 142*(1), 119–126.

Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: An updated review. *Psychonomic Bulletin & Review, 26*(3), 753–771.

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 514 - 523.

Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition*, *26*, 1033-1040.

Simpson, H. M. (1969). Effects of a task-relevant response on pupil size. *Psychophysiology, 6*, 115–121.

DOI: 10.1177/17470218241235671

Spinelli, G., & Lupker, S. J. (2021). Proactive control in the Stroop task: A conflict-frequency manipulation free of item-specific, contingency-learning, and color-word correlation confounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(10), 1550 1562.

Spinelli, G., & Lupker, S. J. (2022). Robust evidence for proactive conflict adaptation in the proportion-congruent paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(5), 675–700. https://doi.org/10.1037/xlm0001144.

Spinelli, G., & Lupker, S. J. (in press). Target-distractor correlation does not imply causation of the Stroop effect. *Quarterly Journal of Experimental Psychology*.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662.

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition*, *20*(6), 727-735.

Verguts, T., & Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive control. *Trends in Cognitive Sciences, 13*(6), 252-257.

West, R., & Baylis, G. C. (1998). Effects of increased response dominance and contextual disintegration on the Stroop interference effect in older adults. *Psychology and Aging*, *13*(2), 206.

West, R., & Alain, C. (2000). Effects of task context and fluctuations of attention on neural activity supporting performance of the Stroop task. *Brain research*, *873*(1), 102-111.

White, B. W. (1969). Interference in identifying attributes and attribute names. *Perception & Psychophysics*, *6*, 166-168.

DOI: 10.1177/17470218241235671

Figure Captions

Figure 1. Series of events during a single trial

Note: Text in bold indicate the sampling time-windows used in the pupillometry analysis. Participants responded by pressing the 1,2, or 3 keys on a standard QWERTY keyboard.

Figure 2. Mean response times for each condition in Experiment 1. Error bars represent 95% confidence intervals.

Figure 3. Mean increase in pupil size after response was made, for each condition in Experiment 1. Error bars represent 95% confidence intervals.

Figure 4. Pupil size (percentage change) relative to baseline Note: The data were downsampled to 100Hz as recommended by Mathôt (2022); and excludes samples not within 2.5 SD following Hershman & Henik (2019). The vertical axes represent the percentage change relative to baseline, while the horizontal axes show the time, in milliseconds, to target onset (0 representing onset).

Figure 5. Mean response times for each condition in experiment 2. Error bars represent 95% confidence intervals.

Figure 6. Mean increase in pupil size in the exploratory post-response analysis in Experiment 2. Error bars represent 95% confidence intervals.

DOI: 10.1177/17470218241235671

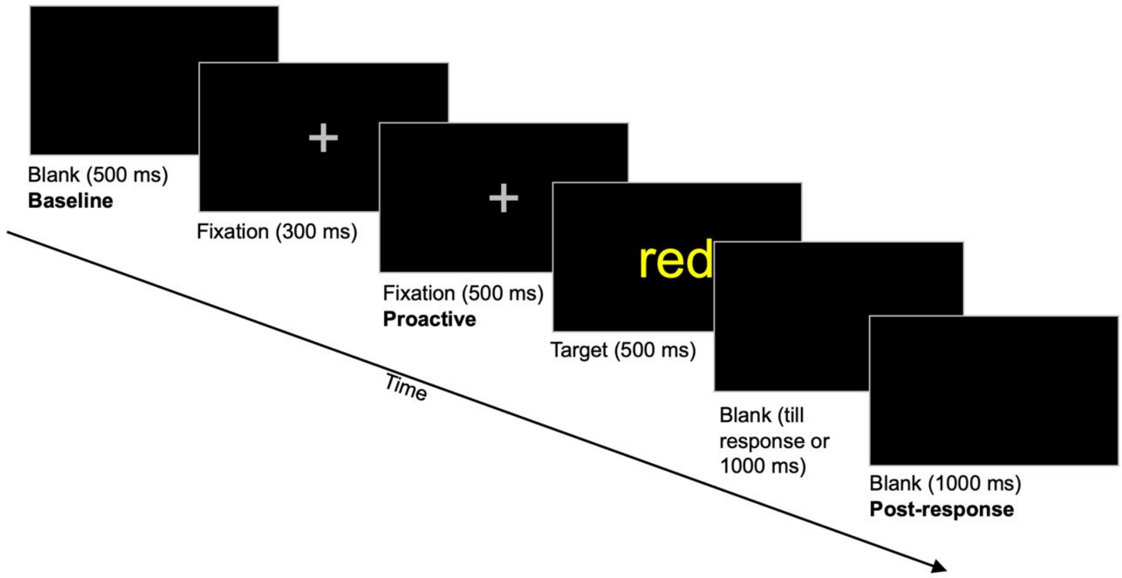

Figure 1. *Series of events during a single trial*

Note: Text in bold indicate the sampling time-windows used in the pupillometry analysis. Participants responded by pressing the 1,2, or 3 keys on a standard QWERTY keyboard.

DOI: 10.1177/17470218241235671

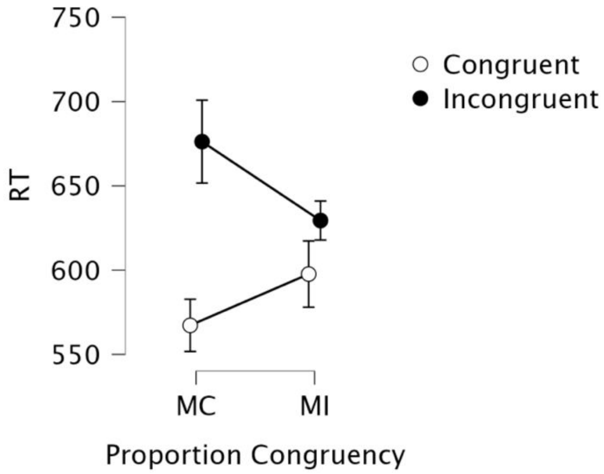

Figure 2. *Mean response times for each condition in Experiment 1. Error bars represent 95% confidence intervals.*

DOI: 10.1177/17470218241235671

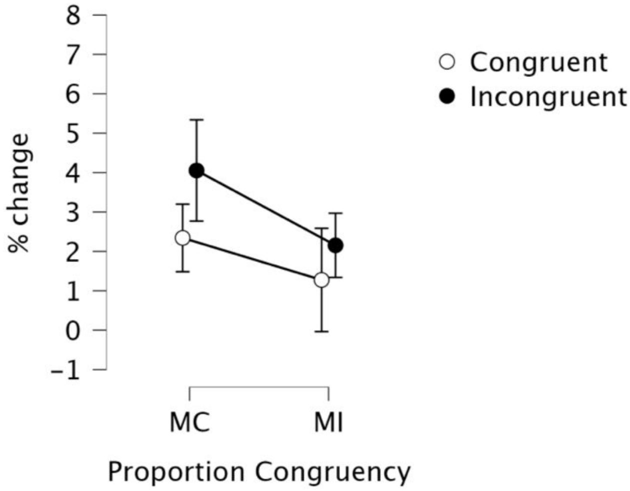

Figure 3. *Mean increase in pupil size after response was made, for each condition in Experiment 1. Error bars represent 95% confidence intervals.*

DOI: 10.1177/17470218241235671

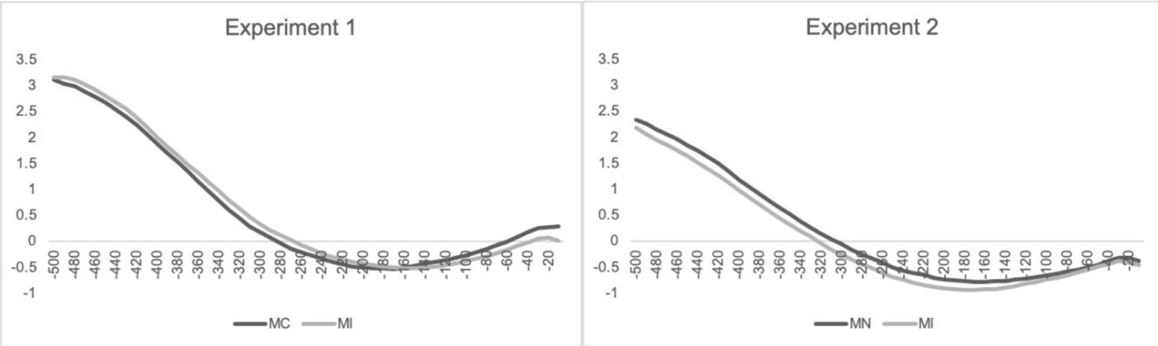

Figure 4. *Pupil size (percentage change) relative to baseline Note: The data were downsampled to 100Hz as recommended by Mathôt (2022); and excludes samples not within 2.5 SD following Hershman & Henik (2019). The vertical axes represent the percentage change relative to baseline, while the horizontal axes show the time, in milliseconds, to target onset (0 representing onset).*

DOI: 10.1177/17470218241235671

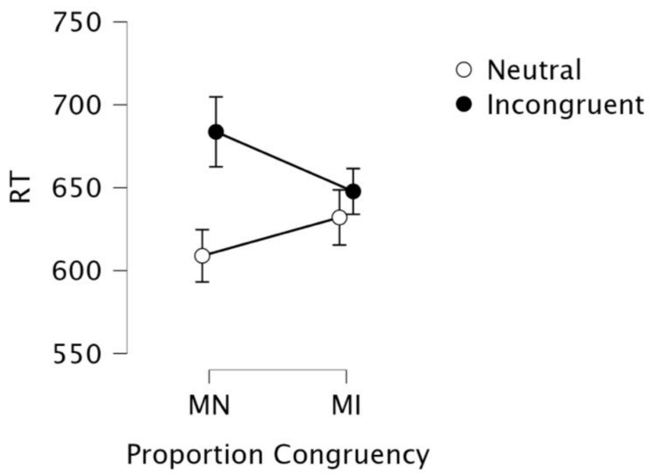

Figure 5. *Mean response times for each condition in experiment 2. Error bars represent 95% confidence intervals.*

DOI: 10.1177/17470218241235671

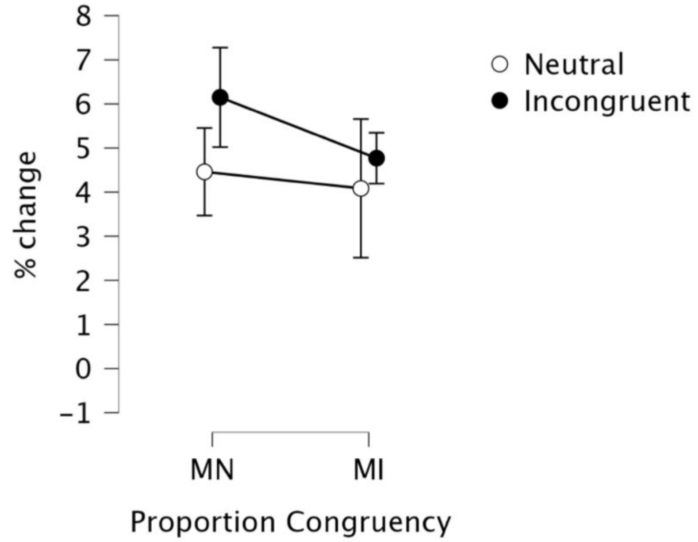

Figure 6. *Mean increase in pupil size in the exploratory post-response analysis in Experiment 2. Error bars represent 95% confidence intervals.*

DOI: 10.1177/17470218241235671

Table 1
Means (SD) of response times and error rates across both experiments

Experiment 1	MC		MI	
	Congruent	Incongruent	Congruent	Incongruent
RT (ms)	567.24 (94.40)	676.22 (146.52)	597.67 (129.01)	629.43 (122.05)
error rate (%)	2.70 (2.34)	7.78 (7.67)	3.78 (3.87)	5.18 (3.07)

Experiment 2	MN		MI	
	Neutral	Incongruent	Neutral	Incongruent
RT (ms)	608.91 (120.44)	683.66 (138.01)	632.05 (144.69)	647.75 (122.23)
error rate (%)	3.17 (2.19)	4.80 (5.72)	4.49 (5.56)	3.86 (3.40)

DOI: 10.1177/17470218241235671