Pupillometric contributions to deciphering Stroop conflicts

Ronen Hershman^{1,2} • Avishai Henik^{2,3}

Published online: 8 August 2019
© The Psychonomic Society, Inc. 2019

Abstract

The Stroop task gives rise to two major conflicts: the task conflict (respond to the color vs. read the word) and the information conflict that can result from the stimulus–response compatibility (SRC; difference between two responses) or from the stimulus–stimulus compatibility (SSC; difference between the two contradictory pieces of information). We conducted a two-to-one Stroop experiment (i.e., two colors are mapped to one response key) and measured reaction time and pupil dilation. The results showed clear evidence for informational conflict composed of both the SRC and SSC. In addition, pupil indications for task conflict appeared earlier than indications for both the SSC and the SRC, in line with the theory regarding task conflict.

Keywords Task conflict · Informational conflict · Stimulus–stimulus compatibility · Stimulus–response compatibility · Stroop effect · Pupillometry · Cognitive control

Introduction

Stroop conflicts

The Stroop task (Stroop, 1935) is one of the most investigated tasks to measure cognitive control. In the commonly used Stroop task (Henik, Bugg, & Goldfarb, 2018; MacLeod, 1991), participants are presented with a color word in color and asked to respond to the ink color and ignore the meaning of the word. The ink color and meaning of the word can be either congruent (e.g., RED written in red), incongruent (e.g., BLUE written in red), or neutral (e.g., XXXX written in red). The difference in reaction time (RT) between incongruent and neutral stimuli—the interference effect—is large and reliable, whereas the difference between neutral and congruent stimuli—the facilitation effect—is small and fragile (Hershman & Henik, 2019; Kalanthroff & Henik, 2013; MacLeod, 1991).

The Stroop task presents two major conflicts. On the one hand, stimuli evoke tasks that are strongly associated with them (Rogers & Monsell, 1995; Waszak, Hommel, & Allport, 2003). Words tend to evoke reading (Monsell, Taylor, & Murphy, 2001), and, as a result, when one has to name the color of the ink, word reading competes with color naming and creates task conflict. Interestingly, the task conflict appears regardless of ink–word congruency (Goldfarb & Henik, 2007; Hershman & Henik, 2019; Kalanthroff, Davelaar, Henik, Goldfarb, & Usher, 2018). On the other hand, the contradiction between word meaning and ink color in incongruent stimuli creates the information conflict. This information conflict can result from the stimulus–response compatibility (SRC) due to the difference between two responses triggered by the color and the word dimensions, or from the stimulus–stimulus compatibility (SSC) due to the difference between the two contradictory pieces of information of the color and the meaning of the stimulus (Chen, Bailey, Tiernan, & West, 2011; De Houwer, 2003; Hasshim & Parris, 2015; Kornblum, Hasbroucq, & Osman, 1990; Shichel & Tzelgov, 2018; van Veen & Carter, 2005).

✉ Ronen Hershman
ronenhe@post.bgu.ac.il

¹ Department of Cognitive and Brain Sciences, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva, Israel

² Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel

³ Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel

De Houwer's two-to-one paradigm

De Houwer's (2003) two-to-one task is one possible way to disentangle informational conflict into its components. In this task, multiple ink colors are mapped to one response key. For example, four colors—red, blue, green, and yellow—are

presented to participants, but there are only two keys for response: the colors red and blue are assigned to the “m” key while the colors yellow and green are assigned to the “b” key. A two-to-one task produces four possible conditions: neutral (represented by noncolor words), congruent (represented by color words whose meaning and ink color are the same) and two incongruent conditions. The first incongruent condition refers to the case where word meaning and ink color are different, but lead to the same response (e.g., “BLUE” in red in the given example). We call this condition incongruent with same response (IC–SR). The second incongruent condition refers to the case where word meaning and the ink color are different and lead to different responses (e.g., “BLUE” in green in the given example). We call this condition incongruent with different response (IC–DR). According to the terminology suggested above, there are two possible sources for conflict, the SRC indicated by the IC–DR (BLUE in green) and the SSC indicated by the IC–SR (BLUE in red).

Many researchers have suggested that the Stroop effect is an example of a response selection effect (Flowers, Warner, & Polansky, 1979; McClain, 1983; Simon & Sudalaimuthu, 1979; Zakay & Glicksohn, 1985). A slowdown of responding in the IC–DR condition compared with the IC–SR condition would support the suggestion that the Stroop effect is triggered by SRC. However, a slowdown in responding in the IC–SR condition compared with the congruent condition would suggest that SSC could contribute to the Stroop effect as well. Moreover, it would suggest that the commonly measured Stroop interference (i.e., the difference between the commonly used incongruent and neutral conditions) is partly contributed to by conflicts triggered by conflicting stimulus features (i.e., SSC) rather than response features.

Previous studies (Chen et al., 2011; De Houwer, 2003; Shichel & Tzelgov, 2018; van Veen & Carter, 2005) that used De Houwer’s (2003) design found both behavioral and neuronal evidence for differences in both SRC and SSC. However, the questions about the temporal occurrences of these differences are still open—namely, when does each difference occur, and for how much time is the difference maintained. To this end, we used pupil dilation.

Stroop and pupil dilation

Pupil dilation is considered to be a good measure of mental effort or task difficulty (Kahneman & Beatty, 1966). Specifically, it is considered to be an index of effort in cognitive control tasks in general (van der Wel & van Steenbergen, 2018) and in the Stroop task in particular (Brown et al., 1999; Laeng, Ørbo, Holmlund, & Miozzo, 2011; Siegle, Ichikawa, & Steinhauer, 2008; Siegle,

Steinhauer, & Thase, 2004). In addition, in contrast to the standard RT measurement, pupil dilation has a temporal appearance that keeps the differences among conditions for long intervals. Hence, if there is a difference between conditions and if the difference is caused by a fast cognitive process, the pupils will react, and the differences will stay for a long duration. In addition, temporal analysis should be useful to avoid a Type II error—namely, it will avoid missing an effect when it actually exists. For example, assume that we measure pupil dilation for 2,000 ms poststimulus and a significant difference between two conditions appears for 400 ms only, at the end of this time interval. If we average pupil dilation across the whole 2,000 ms, we might not find a significant difference between the means of those two conditions. Hence, the effect that appeared for a short duration might be missed. In addition, if we look at a narrow time window in the middle of the whole time window, we also might not find a significant difference. As a result, again, the effect might be missed.

Several studies examined changes in pupil dilation during Stroop tasks (Brown et al., 1999; Laeng et al., 2011; Siegle et al., 2008; Siegle et al., 2004). These studies showed that pupil diameter increased during incongruent trials compared with both neutral and congruent trials. Another pupillometry study was provided by Hasshim and Parris (2015), who used De Houwer’s (2003) two-to-one design and measured pupil dilation using means of the maximum, average, and minimum pupil diameter during the trials. Surprisingly, their results did not show evidence for SSC and SRC. Importantly, by using the typical measures (i.e., average and maximum of pupil diameter during the trials), the researchers did not find any pupillometric evidence for a difference between neutral and incongruent trials, nor between congruent and incongruent trials with and without the same response, as already found in previous studies (Brown et al., 1999; Laeng et al., 2011; Siegle et al., 2008; Siegle et al., 2004).

Stroop models and pupil dilation

Several researchers suggested computational models for the Stroop effect. Several models (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Braver, 2012; Cohen, Dunbar, & McClelland, 1990) dealt with one type of conflict, and a more recent model (Kalanthoff et al., 2018) also tackled task conflict. Kalanthoff et al.’ (2018) departure point was that the indications of task conflict appear when proactive control is low. Low cognitive control might appear as a result of an increase in cognitive load (Kalanthoff, Avnit, Henik, Davelaar, & Usher, 2015;

Kalanthroff & Henik, 2013) or manipulating expectations regarding the need for control (Goldfarb & Henik, 2007; Tzelgov, Henik, & Berger, 1992). Accordingly, in Kalanthroff et al.' (2018) model, “when proactive control is high, there is enough top-down bias to the color-naming task demand unit to prevent task conflict and allow for a net Stroop facilitation effect” (p. 65, Fig. 5). However, our previous results (Hershman & Henik, 2019) suggested that pupil indications for task conflict appear even when control is as high as in any common Stroop experiment (i.e., when the neutral proportion was the same as the other Stroop conditions). We wanted to replicate this finding. Interestingly, Goldfarb and Henik (2007) hypothesized that task conflict might appear earlier than the information conflict, but they did not have any direct evidence for this suggestion. According to our previous study (Hershman & Henik, 2019), we expect task conflict to appear earlier than the information conflict. Kalanthroff et al.' (2018) model dealt with all conflicts as if they appeared at the same time, although they did suggest that task conflict might be handled relatively early when proactive control is high. Hence, changes in the temporal appearance of the various conflicts are important and would help specify new requirements regarding the theory and modeling of the Stroop task. To summarize, we wanted to confirm that task conflict appears even when proactive control is relatively high, and that it appears before the information conflict appears. Moreover, we wanted to find out if the information conflict is unitary or if it is actually composed of an SRC and SSC. And last but not least, assuming the existence of SRC and SSC, do these two conflicts appear together, or do they differ in the time of appearance?

The current study

In the current study, we used Hershman and Henik's (2019) approach to detect temporal changes in pupil size in the three conflicts discussed above—namely, task conflict, SSC, and SRC. In our experiment, participants carried out a two-to-one color response mapping Stroop task (De Houwer, 2003) with four colors and two possible responses. שששש (the repeated Hebrew letter “shin” that replaced the commonly used XXXX) was used as a neutral condition. We used RT and changes in pupil dilation as dependent measures.

Materials and methods

Participants Twenty-six undergraduate students (19 females, mean age = 22.73 years, $SD = 1.04$) from Ben-Gurion

University of the Negev participated in the experiment in return for partial fulfillment of course requirements or credit. The sample size was based on the sample size from Hershman and Henik (2019), who used 19 participants. Taking into consideration dropout rates, we increased our sample size to 26 participants. The study was approved by the university's behavioral ethics committee. All participants signed an informed consent form prior to their participation in the experiment. All participants had normal or corrected-to-normal vision (without glasses or contact lenses) and no reported history of attention deficit disorder or any learning disabilities.

Stimuli Each stimulus consisted of one of four Hebrew color words—כחול (blue), אדום (red), ירוק (green), צהוב (yellow)—or a single four-letter string in Hebrew—שששש (meaningless repetition of a Hebrew letter, parallel to XXXX in the English version of the Stroop task). The combined four letters of each stimulus subtended a visual angle of 2.28° to 3.16° for height and 10.44° to 10.8° for width, from a viewing distance of about 63.5 cm. The ink color was either red (RGB: 255, 0, 0), blue (RGB: 0, 0, 255), green (RGB: 0, 130, 0), or yellow (RGB: 255, 255, 0). There were 20 combinations of words/letter strings and ink colors: four congruent (mean luminance = 191.5), four neutral (mean luminance = 191.15), four incongruent with the same response (mean luminance = 191.49) and eight incongruent with different responses (mean luminance = 191.49). One-third of the trials were neutral, another third were congruent, and the remaining third were divided equally into incongruent trials with same response and incongruent trials with different response. The words were presented at the center of a screen on a silver background (RGB: 192, 192, 192; mean luminance = 192). The conditions and the stimuli within the conditions were selected randomly. The stimuli were printed in bold-faced, 150-point, Ariel font.

Procedure The experiment was conducted in a dimly illuminated room. A keyboard was placed on a table between the participant and the monitor. Participants were tested individually. The experiment included 10 practice trials (which ended when the participants had a success rate of more than 90%. These trials were not analyzed) and four blocks of 144 experimental trials each. During practice, participants received feedback on accuracy. Each trial (see Fig. 1 for a visual example) started with a 1,000-ms fixation (a black “+” sign in the center of the screen), followed by a Stroop stimulus (i.e., a color word/string printed in color). The participants were instructed to press the “b” key on the keyboard if the ink color was blue or green, and to press the “m” key if the ink color was red or yellow (pairing of response keys to pairs of colors was

Fig. 1 An example for a typical trial. Participants had to respond to the ink color of the stimulus. (Color figure online)

counterbalanced across participants). The participants were asked to ignore the meaning of the word and to press the correct key as fast as possible, without making mistakes. The visual stimulus stayed in view for 400 ms and was followed by a blank screen for a maximum of 1,100 ms or until a key press. RT was calculated from the appearance of the visual stimulus to the onset of a response. Each trial ended with a 1,500-ms intertrial interval.

Apparatus Pupil size was measured using a video-based desktop-mounted eye tracker (The Eye Tribe) with a sampling rate of 60 Hz (16.66-ms intersampling time). Stimulus presentation and data acquisition were controlled by Psychtoolbox software (Version 3.0.14) on MATLAB (The MathWorks, Version 9.4.0.813654 [R2018a]). Stimuli were displayed on a 23-inch LED monitor (Dell E2314Hf) at a resolution of $1,920 \times 1,080$ pixels, with a refresh rate of 60 Hz. The participant's head was positioned on a chin rest, and the distance from the eyes to the monitor was set at about 63.5 cm. To maintain an accurate measurement of pupil size during the task, participants were required to keep their eyes fixated on the center of the screen and to avoid eye movements for the entire task. Pupil area was determined using the Eye Tribe algorithm.

Pupil data analysis Two participants were excluded from the analysis because they did not have at least 60 valid trials (correct responses with no more than 30% of missing values) in each condition. Pupil data was processed using CHAP software (Hershman, Henik, & Cohen, 2019). First, pupil data was extracted from the Eye Tribe (pupil size in arbitrary units). Then, we removed outlier samples with Z scores larger than 2.5 (by using Z scores based on the mean and standard deviation calculated for each trial). Next, for each participant, we excluded from analysis the trials with more than 30% of missing values. We also excluded trials with

no response or with incorrect responses. This preprocessing eliminated 12.78% of trials on the average (for the 24 participants included in the final sample; 18 females, mean age = 22.66 years, $SD = 1.05$). Next, we detected eye blinks by using Hershman, Henik, and Cohen's (2018) algorithm and filled missing values by using a linear interpolation (Hershman & Henik, 2019). Next, time courses were aligned with the onset of the Stroop stimulus and divided by the baseline (baseline was defined as the average pupil size 200 ms before the stimulus onset).

Results

Reaction time In order to verify that the Stroop task worked as expected, mean RTs of correct (pupil valid) trials (i.e., 87.22%; $C = 89.21\%$, $N = 87\%$, $IC-SR = 88.98\%$, and $IC-DR = 83.68\%$) for each participant in each condition were subjected to a one-way repeated-measures analysis of variance (ANOVA) with congruity (C , N , $IC-SR$, & $IC-DR$) as an independent factor. As we expected, an omnibus analysis produced a significant effect, $F(3, 69) = 24.34$, $p < .001$, $\eta_p^2 = .51$ (mean RTs in the various conditions are presented in Fig. 2).

Mean RT was slower in $IC-DR$ trials compared with $IC-SR$ trials, $F(1, 23) = 16.52$, $p < .001$, $BF_{10} = 67.07$; mean RT was slower in $IC-SR$ trials compared with congruent trials, $F(1, 23) = 30.69$, $p < .001$, $BF_{10} = 1,749.39$; but there was no significant difference in RT between neutral to congruent trials, $F(1, 23) < 1$, $BF_{01} = 4.28$.

Error rate Error rates for pupil valid trials for each participant in each condition are relative to the number of trials with responses. $IC-SR$ trials had an error rate of 4.22%,

Fig. 2 Mean reaction time for each congruency condition of Stroop trials in the experiment. Error bars represent one confidence interval from the mean. C = congruent; N = neutral; $IC-SR$ = incongruent–same response; $IC-DR$ = incongruent–different response. (Color figure online)

congruent trials had an error rate of 4.82%, neutral trials had an error rate of 6.6%, and IC–DR trials had an error rate of 8.53%. Error rates were subjected to a one-way analysis of variance (ANOVA) with congruity (C, N, IC–SR, and IC–DR) as an independent factor. The ANOVA produced a significant effect, $F(3, 69) = 9.205, p < .001, \eta_p^2 = .286$. No difference was found between congruent and IC–SR trials, $F(1, 23) < 1, BF_{01} = 3.08$. Mean error rate was larger for neutral trials compared with congruent trials, $F(1, 23) = 10.31, p = .003, BF_{10} = 10.82$. No difference was found between neutral and IC–DR trials, $F(1, 23) = 3.34, p = .08, BF_{01} = 1.11$.

Pupillometry Mean relative changes of the pupil size in each condition are presented in Fig. 3. In order to examine the temporal differences among the four Stroop conditions, we used Hershman and Henik’s (2019) approach. Specifically, we ran a series of Bayesian paired-samples t tests between each of the two conditions over the whole time course of pupil measurement. Figure 3 presents these comparisons also. This analysis indicates that there are significant differences between all the investigated conditions.

Specifically, our analysis (see Fig. 3) indicates that significant differences between the neutral (blue) and nonneutral (congruent: green, IC–SR: magenta, and IC–DR: red) conditions appeared early at about

Fig. 3 Mean relative pupil size (compared with size at the stimulus onset) for the four congruency conditions in the experiment. Participants had to respond to the ink color of the stimulus. The vertical line at zero represents stimulus onset, and the other vertical lines represent mean response times (around time 500 ms poststimulus onset) for each condition. The four line curves present changes in pupil dilation as a function of time. The shaded areas represent one standard error from the mean. The horizontal lines represent significant comparisons for each contrast (e.g., the red and green lines indicate significant differences in pupil response between the incongruent–different response and the congruent conditions). IC–DR = incongruent–different response; IC–SR = incongruent–same response; C = congruent; N = neutral. (Color figure online)

500 ms after the stimulus onset. The nonneutral conditions started to separate at 1,000 ms after stimulus onset. The difference between IC–DR and both IC–SR and congruent started at about 1,000 ms after the stimulus onset. In contrast, the difference between congruent and IC–SR appeared only at about 1,400 ms after the stimulus onset (see Fig. 4 in the Appendix for the detailed Bayes factor figures).

Discussion

We conducted De Houwer’s (2003) two-to-one Stroop task and measured both RT and pupil dilation. Our results showed that RT was slower in IC–DR compared with IC–SR trials, slower in the IC–SR trials compared with congruent trials, and there was no difference in congruent compared with nonword neutral trials. In contrast, pupil dilation produced large differences among all the investigated conditions (i.e., the pupil was larger in the IC–DR compared with IC–SR trials, larger in the IC–SR compared with congruent trials, and larger in the congruent compared with neutral trials).

Our results suggest that the Stroop task includes three conflicts—namely, task conflict, SSC, and SRC. The difference in pupil size response to congruent compared with neutral trials is indicative of the task conflict (i.e., respond to the color vs. read the word) and is similar to results found in anterior cingulate cortex (ACC) activations in neuroimaging experiments (Bench et al., 1993; Carter, Mintun, & Cohen, 1995) and in pupillometry studies (Hershman & Henik, 2019). Importantly, pupil indications for task conflict appeared at the beginning of the whole time period that was measured. Early on (about 500 ms after stimulus onset), congruent and both incongruent (i.e., IC–DR and IC–SR) conditions did not significantly differ in pupil response, but all three conditions resulted in a larger pupil size than for the neutral condition. In the subsequent time samples (starting at about 1,000 ms after stimulus onset), pupil size remained smallest for the neutral condition, and the congruent and incongruent (i.e., IC–DR and IC–SR) conditions produced different pupil sizes. Specifically, pupil size in response to incongruent trials was largest. The current results are line with Goldfarb and Henik’s (2007) suggestion that task conflict appears earlier than the information conflict. Moreover, they suggested that because it appears early, it can be dealt with earlier and thus contributes to the production of (small) facilitation rather than reverse facilitation (namely, slower responses in congruent trials compared with neutral trials).

The difference in pupil size between the congruent and incongruent (specifically IC–DR) results starting at

about 1,000 ms indicates the existence of a general information conflict. In addition, the difference in pupil size between IC–DR and IC–SR results at the same time (i.e., starting at about 1,100 ms after the stimulus onset) indicates that the information conflict is composed of SSC and SRC (Chen et al., 2011; De Houwer, 2003; Hasshim & Parris, 2015; Kornblum et al., 1990; Shichel & Tzelgov, 2018; van Veen & Carter, 2005). Moreover, in contrast to the suggestions that the Stroop effect is an example for a response selection effect (Flowers et al., 1979; McClain, 1983; Simon & Sudalaimuthu, 1979; Zakay & Glicksohn, 1985), our results suggest that the Stroop effect is triggered partly by SSC. Pupil dilation in responding in the IC–SR compared with the congruent condition was enlarged starting at about 1,400 ms after the stimulus onset. This adds to the suggestion that SSC contributes to the Stroop effect. Specifically, our results suggest that both SSC and SRC appear at the same time (about 1,100 ms after the stimulus onset).

Pupil dilation is considered to be an index of effort in cognitive control tasks in general (van der Wel & van Steenbergen, 2018), and in the Stroop task in particular (Brown et al., 1999; Laeng et al., 2011; Siegle et al., 2008; Siegle et al., 2004). Hence, the temporal analysis of changes in pupil dilation allows us to examine the inception and decay of the cognitive conflicts mentioned above. Task conflict, indicated by the response difference between neutral and congruent conditions, appears as early as 500 ms after stimulus onset, whereas the information conflict, indicated by the divergence of the results in the incongruent and congruent conditions, starts later, at about 1,000 ms after stimulus onset. This pattern of results is in line with Hershman and Henik's (2019) and with Goldfarb and Henik's (2007) suggestion that task conflict appears before the information conflict. Interestingly, task conflict decays earlier; the difference between congruent and neutral conditions is not significant from about 1,300 ms post-stimulus onset. In contrast, the information conflict, indicated by the difference between the IC–DR (which is the commonly used incongruent condition) and congruent conditions, continues to be significant for the rest of the measurement (until 2,000 ms poststimulus onset). Importantly, the components of the information conflict—the SSC (indicated by the response difference between the IC–SR and the congruent conditions) and the SRC (indicated by the response difference between the IC–DR and the IC–SR conditions)—stop being significant at about 1,700 ms poststimulus onset. This suggests that the robust information conflict, which is usually discussed as being the major feature of the Stroop

effect (e.g., MacLeod, 1991), and seems to last for a long period of time, is not only dependent on the stimulus–response conflict but builds partly on the stimulus–stimulus conflict for its robustness and slow decay.

In the Introduction, we mentioned that the computational model proposed by Kalanthroff et al. (2018) suggested that task conflict appears when cognitive control is low. The current results show that pupil indications for task conflict appear even when control is as high as in any common Stroop experiment (i.e., the neutral proportion was the same as the other Stroop conditions; see also Hershman & Henik, 2019). Moreover, Kalanthroff et al.' (2018) model dealt with all conflicts as if they appeared at the same time. In contrast, our current results show that different conflicts onset at different times poststimulus. Our results suggest that such models need to make some changes in order to be able to predict actual results.

Summary

The current study replicated the temporal development of both task and information conflicts that we found in our previous study (Hershman & Henik, 2019). Specifically, we showed that task conflict (represented by more dilation in congruent trials compared with neutral trials) occurred before the information conflict (represented by more dilation in incongruent trials compared with congruent trials). In addition, the current study suggested that the SSC contributes to the information conflict on top of the contribution of the SRC. Our temporal analysis suggested that both SSC and SRC were triggered at the same time. These findings provide evidence for the temporal development of the various conflicts that compose the Stroop effect and improve our understanding of the temporal development of these conflicts. This temporal appearance of the various conflicts is important and would help specify new requirements regarding the theory and modeling of the Stroop task.

Author note We wish to thank Ms. Desiree Meloul for helpful comments and useful input on this article. We also wish to thank Professor Joseph Tzelgov and Mr. Ido Shichel for helpful information and advice.

This work was supported by funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC Grant Agreement No. 295644.

Open practices statement None of the data or materials for the experiments reported here are available, and none of the experiments were preregistered.

Appendix

Detailed Bayes factor analysis

Fig. 4 Bayes factors (BFs) as function of time for each comparison in the experiment from 500 ms before the stimulus onset to 2,050 ms after the stimulus onset. Blue curves represent BF_{10} (namely, evidence for the alternative hypothesis that the two conditions are not the same), and red curves represent BF_{01} (namely, evidence for the null hypothesis that the two conditions are the same). The horizontal bold black line on the

bottom represents significant differences in the time window ($BF_{10} \geq 3$). The horizontal green line on the bottom indicates a significant result of no difference between the conditions in this time window ($BF_{01} \geq 3$). N = neutral; C = congruent; IC-SR = incongruent with same response; IC-DR = incongruent with different response. (Color figure online)

References

- Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Paulesu, E., Frackowiak, R. S. J., & Dolan, R. J. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia*, 31(9), 907–922. [https://doi.org/10.1016/0028-3932\(93\)90147-R](https://doi.org/10.1016/0028-3932(93)90147-R)
- 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. <https://doi.org/10.1037/0033-295X.108.3.624>
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106–113. <https://doi.org/10.1016/j.tics.2011.12.010>
- Brown, G. G., Kindermann, S. S., Siegle, G. J., Granholm, E., Wong, E. C., & Buxton, R. B. (1999). Brain activation and pupil response during covert performance of the Stroop color word task. *Journal of the International Neuropsychological Society*, 5(4), 308–319. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10349294>
- Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An H215O PET study of Stroop task performance. *NeuroImage*, 2(4), 264–272. <https://doi.org/10.1006/NIMG.1995.1034>
- Chen, A., Bailey, K., Tiernan, B. N., & West, R. (2011). Neural correlates of stimulus and response interference in a 2–1 mapping Stroop task. *International Journal of Psychophysiology*, 80(2), 129–138. <https://doi.org/10.1016/J.IJPSYCHO.2011.02.012>
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, 97(3), 332–361. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2200075>
- De Houwer, J. (2003). On the role of stimulus–response and stimulus–stimulus compatibility in the Stroop effect. *Memory & Cognition*, 31(3), 353–359. <https://doi.org/10.3758/BF03194393>
- Flowers, J. H., Warner, J. L., & Polansky, M. L. (1979). Response and encoding factors in “ignoring” irrelevant information. *Memory & Cognition*, 7(2), 86–94. <https://doi.org/10.3758/BF03197589>
- Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology: Human Perception and Performance*, 33(5), 1170–1176. <https://doi.org/10.1037/0096-1523.33.5.1170>
- 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. <https://doi.org/10.3758/s13414-015-0971-9>
- Henik, A., Bugg, J. M., & Goldfarb, L. (2018). Inspired by the past and looking to the future of the Stroop effect. *Acta Psychologica*, 189, 1–3. <https://doi.org/10.1016/j.actpsy.2018.06.007>
- Hershman, R., & Henik, A. (2019). Dissociation between reaction time and pupil dilation in the Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000690>
- Hershman, R., Henik, A., & Cohen, N. (2018). A novel blink detection method based on pupillometry noise. *Behavior Research Methods*, 50(1), 107–114. <https://doi.org/10.3758/s13428-017-1008-1>
- Hershman, R., Henik, A., & Cohen, N. (2019). CHAP: Open-source software for processing and analyzing pupillometry data. *Behavior Research Methods*, 51(3), 1059–1074. <https://doi.org/10.3758/s13428-018-01190-1>
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science (New York, N.Y.)*, 154(3756), 1583–1585. <https://doi.org/10.1126/science.154.3756.1583>
- Kalanthroff, E., Avnit, A., Henik, A., Davelaar, E. J., & Usher, M. (2015). Stroop proactive control and task conflict are modulated by concurrent working memory load. *Psychonomic Bulletin & Review*, 22(3), 869–875. <https://doi.org/10.3758/s13423-014-0735-x>
- Kalanthroff, E., Davelaar, E. J., Henik, A., Goldfarb, L., & Usher, M. (2018). Task conflict and proactive control: A computational theory of the Stroop task. *Psychological Review*, 125(1), 59–82. <https://doi.org/10.1037/rev0000083>
- Kalanthroff, E., & Henik, A. (2013). Individual but not fragile: Individual differences in task control predict Stroop facilitation. *Consciousness and Cognition*, 22(2), 413–419. <https://doi.org/10.1016/J.CONCOG.2013.01.010>
- Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus–response compatibility—A model and taxonomy. *Psychological Review*, 97(2), 253–270. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2186425>
- Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing*, 12(1), 13–21. <https://doi.org/10.1007/s10339-010-0370-z>
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203. <https://doi.org/10.1037/0033-2909.109.2.163>
- McClain, L. (1983). Stimulus–response compatibility affects auditory Stroop interference. *Perception & Psychophysics*, 33(3), 266–270. Retrieved from <https://link.springer.com/content/pdf/10.3758/BF03202864.pdf>
- Monsell, S., Taylor, T. J., & Murphy, K. (2001). Naming the color of a word: Is it responses or task sets that compete? *Memory & Cognition*, 29(1), 137–151. <https://doi.org/10.3758/BF03195748>
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2), 207–231. <https://doi.org/10.1037/0096-3445.124.2.207>
- Shichel, I., & Tzelgov, J. (2018). Modulation of conflicts in the Stroop effect. *Acta Psychologica*, 189, 93–102. <https://doi.org/10.1016/J.ACTPSY.2017.10.007>
- Siegle, G. J., 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(5), 679–687. <https://doi.org/10.1111/j.1469-8986.2008.00681.x>
- Siegle, G. J., Steinhauer, S. R., & Thase, M. E. (2004). Pupillary assessment and computational modeling of the Stroop task in depression. *International Journal of Psychophysiology*, 52(1), 63–76. <https://doi.org/10.1016/J.IJPSYCHO.2003.12.010>
- Simon, J. R., & Sudalaimuthu, P. (1979). Effects of S–R mapping and response modality on performance in a Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, 5(1), 176–187. <https://doi.org/10.1037/0096-1523.5.1.176>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662. <https://doi.org/10.1037/h0054651>
- Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition*, 20(6), 727–735. <https://doi.org/10.3758/BF03202722>
- van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, 25(6), 2005–2015. <https://doi.org/10.3758/s13423-018-1432-y>
- van Veen, V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: A functional MRI study. *NeuroImage*, 27(3), 497–504. <https://doi.org/10.1016/J.NEUROIMAGE.2005.04.042>
- Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus–task bindings in task-shift

costs. *Cognitive Psychology*, 46(4), 361–413. [https://doi.org/10.1016/S0010-0285\(02\)00520-0](https://doi.org/10.1016/S0010-0285(02)00520-0)

Zakay, D., & Glicksohn, J. (1985). Stimulus congruity and S–R compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology/Revue Canadienne de*

Psychologie, 39(3), 414–423. <https://doi.org/10.1037/h0080069>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.