# QJEP

# The contribution of meaning to the detection of task conflict

**Ronen Hershman[1,2]** , **Yulia Levin[3], Joseph Tzelgov[1,2,4]** **and Avishai Henik[2,3]**

## Abstract

The colour-word Stroop task produces both information conflict (detection of the ink colour vs word meaning) and task conflict (respond to the ink colour vs read the word). In this study, we measured both reaction time and pupil dilation, and the neutral stimuli in our study were non-readable letter strings as well as meaningless non-readable stimuli (i.e., coloured patches and abstract character strings). Our results showed slowest responses in the incongruent trials and fastest responses in the congruent trials. However, no differences were found between the investigated neutrals. In contrast, pupil dilation was largest in the incongruent trials and smallest in the neutral trials. Moreover, the more the neutral stimuli were meaningless, the less the pupil dilation that was observed. Our results suggest that non-word meaningless stimuli reduced task conflict (compared with all the investigated conditions). Neutral equivalence should be taken into consideration in Stroop and Stroop-like tasks.

## Introduction

The Stroop task (Henik et al., 2018; Parris et al., 2019) is one of the most frequently used tasks to examine cognitive control. It examines the ability to focus on relevant information and ignore irrelevant information. Specifically, in the commonly used colour-word Stroop task (Henik et al., 2018), participants are presented with coloured stimuli and are required to name the ink colour. The ink colour could be congruent with the meaning of the word (e.g., the word RED written in red ink) or incongruent with the meaning of the word (e.g., the word BLUE written in red ink). In addition, there are cases when the stimulus is not a colour word (e.g., the word TABLE)—this is a neutral condition. The neutral condition does not have to be a readable word. It can also be a non-word stimulus. Non-word stimuli can be pseudo-words (e.g., TAKLE) (Kinoshita et al., 2017; Monsell et al., 2001), letter strings (e.g., XXXX) (Brown, 2011; Goldfarb & Henik, 2007; Hershman & Henik, 2019; Kalanthroff et al., 2013), or symbol strings (e.g., @@@@) (Augustinova & Ferrand, 2012; Kinoshita et al., 2018; Monsell et al., 2001). In the original Stroop experiment (Stroop, 1935), Stroop used coloured shapes as the control condition. In general, slower responses are observed in incongruent trials compared with neutral trials. This

difference in reaction time (RT) is called interference and it is large and reliable. In addition, RTs to the neutral trials are slower compared with congruent trials. This difference is called facilitation and it is usually small and fragile (Hershman & Henik, 2019; Kalanthroff & Henik, 2013; MacLeod, 1991).

### Stroop conflicts

The Stroop task presents a mismatch between word meaning and ink colour in incongruent stimuli. This mismatch is often called the information conflict and only arises when the stimulus is a colour-word or colour-related incongruent

[1]Department of Cognitive and Brain Sciences, Ben-Gurion University of the Negev, Be'er Sheva, Israel
[2]Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Be'er Sheva, Israel
[3]Department of Psychology, Ben-Gurion University of the Negev, Be'er Sheva, Israel
[4]Achva Academic College, Arugot, Israel

**Corresponding author:**
Ronen Hershman, Department of Cognitive and Brain Sciences, Ben-Gurion University of the Negev, P.O.B. 653, Be'er Sheva 84105, Israel.
Email: ronenhe@post.bgu.ac.il

stimulus (e.g., the word banana appearing in blue ink colour). However, it has been suggested that the Stroop task also gives rise to task conflict. In general, stimuli evoke tasks that are strongly associated with them (Rogers & Monsell, 1995; Waszak et al., 2003) and words tend to evoke reading (Monsell et al., 2001; Stroop, 1935). As a result, when one has to name the colour of the ink of a word stimulus, word reading competes with colour naming and creates task conflict. Interestingly, the task conflict appears regardless of ink-word congruency (i.e., in both congruent and incongruent trials) (Goldfarb & Henik, 2007; Hershman & Henik, 2019; Kalanthroff et al., 2018). When a stimulus word is a non-colour word or is a letter string, the observed interference expresses the task conflict and has nothing to do with informational compatibility (Levin & Tzelgov, 2014). Thus, the task conflict should represent a major component of Stroop interference.

### Stimuli meaning and task conflict

Klein (1964) and Fox et al. (1971) in their replication used various non-colour stimuli and reported that the interference effect depended on the readability of the neutral stimuli; common words interfered more than rare words, rare words interfered more than consonant strings, and consonant strings interfered more than strings of asterisks. Recent studies showed that during a vocal colour-word Stroop task, the readability of neutral stimuli influenced the interference effect in that readable (i.e., containing lexical material) stimuli elicited much slower RTs than completely unreadable geometric shapes. However, the difference between letter strings and real words was less pronounced and somewhat unstable across experiments (Levin & Tzelgov, 2016). In a recent work by Kinoshita et al. (2017), the researchers investigated the effect of response modality on interference produced by a large variety of neutral stimuli. Specifically, they showed gradually increasing interference as a function of the readability of the presented stimuli, relative to a series of Xs. This interference that appears due to reading and represents task conflict (Levin & Tzelgov, 2016) was observed only when vocal responses were required but not when manual (i.e., key pressing) responses were required. Recent studies have confirmed these results and showed that task conflict was observed only when vocal responses were required but not when manual responses were required (Augustinova et al., 2018; Parris et al., 2019).

### Pupil dilation

Pupil dilation is an unobtrusive high temporal resolution measure that is thought to change as a result of task difficulty (Kahneman & Beatty, 1966). Interestingly, pupil dilation seems to be a sensitive measurement for task conflict (Hershman et al., 2020; Hershman & Henik, 2019, 2020).

Recently, Hershman and Henik (2019, 2020) presented a new approach to detect temporal changes in pupil size by using Bayesian analysis. Examination of the temporal differences between the conditions revealed different effects (i.e., both task and information conflicts) in different time windows. Specifically, it was found that during the manual Stroop task, there are both task and information conflicts, and moreover, the task conflict appears before the information conflict is initiated. Hershman et al. (2020) used this temporal approach to look into task-related sub-components of the Stroop interference effect. They measured changes in pupil size using trials with various kinds of neutrals. Specifically, they found that comparisons of different neutral types (words, pseudo-words, letter strings, and symbols) did not show any meaningful difference in RT or pupil dilation. However, the indications for task conflict using the various types of neutrals were different. Namely, the less readable the neutral stimulus was, the earlier reverse facilitation appeared (i.e., less dilation for neutral compared with congruent trials) and the longer the duration (i.e., wider time window) of task conflict.

Words have both lexical and semantic meanings that may be processed, pseudo-words (as well as consonant strings) have quasi phonology that may be processed, and also symbols have a semantic meaning that may be processed (e.g., $ is associated with money, * and # are often associated with passwords, and random presentation of symbols like "#$%@" is strongly associated with swear words). In this study, we used the same paradigm as in Hershman et al.'s (2020) study to investigate the temporal changes of task conflict for different stimuli sets. Specifically, we used meaningless neutral stimuli sets that should not cause task conflict (or at least, should cause less task conflict compared to the typically used neutral stimuli). In particular, in addition to the standard letter strings (e.g., XXXX), we used coloured patches and abstract draws as neutrals. Colour patches do not have any semantical/lexical/phonological meaning that may trigger processing except for colour naming—the required task. Abstract draw stimuli also have no semantic/lexical/phonological meaning. Moreover, these stimuli were never presented to the participants before and therefore they cannot be associated with something else. While the abstract draws look like characters and strings of them may somehow be associated with words (at least in terms of their visual properties), coloured patches are completely unreadable—they do not possess any semantic/lexical/phonological/orthographical feature that can be read—in contrast to abstract draws as well as letter strings (e.g., XXXX has letters that can be identified, which involves a reading sub-process such as orthographic encoding).

## Pilot experiment

In a pilot experiment (see supplementary material), we conducted a colour-word Stroop task (MacLeod, 1991)

with four colour responses and measured both RT and pupil dilation. In our experiment, two types of neutrals were used: letter strings (e.g., XXXX) and coloured patches (filled horizontal rectangles). Our results showed that RT was faster in congruent trials compared with both letter strings and coloured patches. In the same vein, these two types of neutrals provided faster responses compared with incongruent trials. Similar to RT, pupil dilation was larger in the incongruent trials (indicating more effort) compared with the congruent trials. However, in contrast to the RT, pupil dilation was larger in the congruent trials compared with both letter strings and coloured patches. Moreover, pupil dilation was considerably smaller for coloured patches compared with letter strings.

The results suggested that the Stroop task includes two conflicts, namely, both information and task conflicts. The difference in pupil size response to incongruent compared with congruent trials (between about 1,250 ms to about 1,550 ms after the stimulus onset) is indicative of information conflict. In addition, in line with previous pupillometry results (Hershman et al., 2020; Hershman & Henik, 2019, 2020), before these changes (i.e., early on, about 500 ms post stimulus onset), our results showed larger pupil dilation in the congruent trials compared with the neutral trials (both for letter strings and coloured patches). These reverse facilitations (that stayed for about 400 ms) are typically used as a marker of task conflict (i.e., respond to the colour vs read the word).

Interestingly, the divergence at about 500 ms post stimulus onset appeared also between the two types of neutrals. Specifically, from about 500 ms post stimulus onset, there are meaningful differences between letter strings and coloured patches. These differences stayed stable until the end of the trial. The lack of meaningful information of the stimuli (e.g., coloured patches) did not trigger reading, leading to no or much smaller task conflict (Augustinova & Ferrand, 2014; Kinoshita et al., 2017; Levin & Tzelgov, 2016; Monsell et al., 2001).

This dissociation between RT and pupil dilation provides clear evidence for the influence of the meaning of the stimulus. While RT suggests that all neutrals have the same contribution to the Stroop conflict, pupil dilation (as an indicator of mental effort) suggests that the meaning of stimuli has a major influence on the Stroop conflict.

What would happen if a neutral stimulus without meaning would be a string of meaningless objects? Most of the presented stimuli in the pilot experiment were four-letter strings, which formed a colour word or a meaningless letter string. In contrast, the coloured patches were a single "object" (i.e., one coloured patch). It could be argued that the differences between coloured patches and the other stimuli were due to the number of presented objects (i.e., smaller dilation for a single object compared with a string of four objects), or due to the fact that coloured shapes have no orthographic structure compared with the other

stimuli. To examine this possible artefact, we carried out another experiment with one more type of neutral strings of meaningless abstract characters that had the same orthographic structure.

## This study

In this study, participants responded manually in a colour-word Stroop task (MacLeod, 1991) with four colour responses. The set of trials included (in addition to the usual congruent and incongruent trials) letters, coloured patches, and abstract character strings as neutrals. This study had two main aims: first, we expected to replicate the results of the pilot experiment (see supplementary material); namely, larger dilation in the incongruent trials compared with the congruent trials, larger dilation in the congruent trials compared with the letters, and considerably larger dilation in the letter strings compared with coloured patches. In addition, we wanted to find out if abstract character strings (that do not include meaningful information) would lead to smaller dilation than for the letter strings.

### Methods

*Participants.* Twenty-two undergraduate students (15 females, mean age = 24.18 years, $SD$ = 1.87) from Ben-Gurion University of the Negev participated in the experiment in return for 35 shekels (approximately $10) or partial fulfilment of course requirements or credit. All participants signed an informed consent form prior to their participation in the experiment. All participants had normal vision (without glasses or contact lenses) as well as normal colour vision and no reported history of attention-deficit disorder or any learning disabilities.

*Stimuli.* The stimuli in the experiment subtended a visual angle of 4.92° to 7.28° for height and 13.16° to 22.59° for width, from a viewing distance of about 50 cm. The ink colour was red (RGB: 255, 0, 0), blue (RGB: 0, 0, 255), green (RGB: 0, 130, 0), or yellow (RGB: 255, 255, 0).

Each stimulus consisted of one of 16 possible stimuli (see Table 1 for the list of stimuli). The stimuli were divided into five possible groups: congruent colour words, incongruent colour words, single character four-letter strings in Hebrew, strings of abstract meaningless draws, and coloured rectangles (patches). The draws were based on La Heij and colleagues' (2010) stimuli. There were 64 possible combinations of strings and ink colours: 4 congruent (mean luminance = 190.34), 12 incongruent (mean luminance = 190.23), 16 letters (mean luminance = 189.97), 16 abstract character strings (mean luminance = 189.86), and 16 patches (mean luminance = 187.8). The experimental part of the experiment included 864 trials. One-third of the presented trials were incongruent,

**Table 1.** Stimuli used in our experiment.

| Congruent | Incongruent | Letters (LET) | Patches (PTC) | Abstract (ABST) |
|---|---|---|---|---|
| Colour word | | | | |
| כחול (blue) | | לללל | ▬ | 🦋🦋🦋 |
| ירוק (green) | | עעעע | ▬ | 🐛🐛🐛🐛 |
| צהוב (yellow) | | רררר | ▬ | ☘☘☘☘ |
| אדום (red) | | ששש | ▬ | 🦂🦂🦂🦂 |

The stimuli were divided into five possible groups. Colour words were used for both incongruent (e.g., the word אדום [red] written in blue ink) and congruent (e.g., the word כחול [blue] written in blue ink) trials.

one-third were congruent, and one-third were composed of three conditions of neutrals—one-ninth were letters, one-ninth were abstract character strings, and one-ninth were coloured patches. The stimuli were presented at the centre 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 150-point, Arial 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 started with a short pre-experimental part. In this part, the experimental procedure was the same as the main task but here the participants were presented with a coloured background instead of a Stroop stimulus. The colour of the background was blue, green, yellow, or red and each background appeared eight times. In total, this part included 32 trials. Participants were asked to press the appropriate key for each presented colour (e.g., z for blue). This pre-experimental trial was repeated until the participants had a success rate of more than 80% (namely, a maximum of six mistakes). The experimental part included 10 practice trials (which were repeated and ended when the participants had a success rate of more than 80%; these trials were not analysed) and 864 experimental trials. After every 144 trials, the participants took a rest for few seconds. The different stimuli were presented in random order in both the practice and the experimental blocks. During practice, participants received feedback on accuracy. Each trial (see Figure 1 for a visual example) started with a 1,000 ms fixation (a black "+" sign in the centre of the screen), followed by a Stroop stimulus (i.e., a colour word/string/patch printed in colour). The participants were instructed to press the "z" key on the keyboard if the ink colour was blue, the "x" key if the ink colour was green, the "n" key if the ink colour was yellow, and the "m" key if the ink colour was red. They were asked to ignore the meaning of the stimulus and to press the correct key as fast as possible, without making
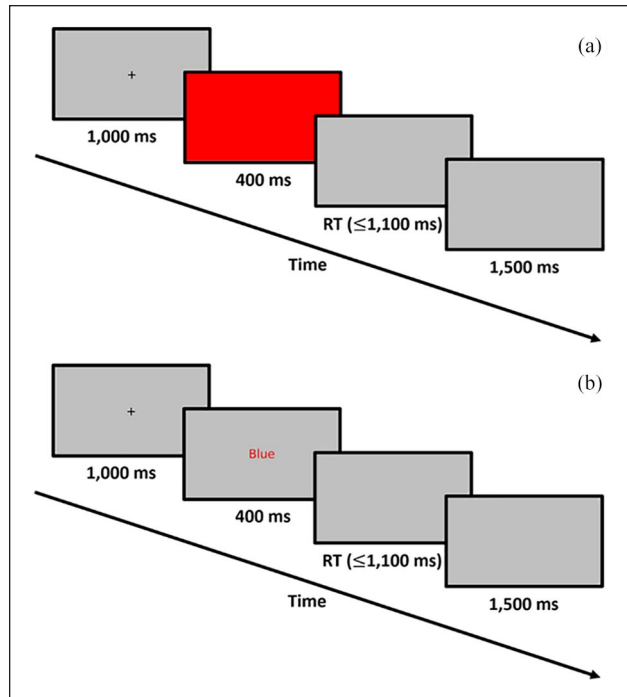


**Figure 1.** Examples of a typical trial. Participants had to respond to the ink colour of the stimulus/background. (a) An example of pre-experimental practice. (b) An example of an experimental trial.

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 inter-trial 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 inter-sampling time). Stimulus presentation and data acquisition were controlled by Psychtoolbox software (version 3.0.14) on MATLAB (MathWorks version 9.4.0.813654 [R2018a]). Stimuli were displayed on a 23-inch LED monitor (Dell E2314Hf) at a resolution of 1920 × 1080 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 50 cm. To maintain an accurate measurement of pupil size during the task, participants were required to keep their eyes fixated on the centre of the screen and to avoid eye movements for the entire task. Pupil area was determined using the Eye Tribe algorithm.

## Results

*Pre-processing of pupillometry data.* One participant was excluded from the analysis because she did not have at least 50 valid trials (correct responses with no more than
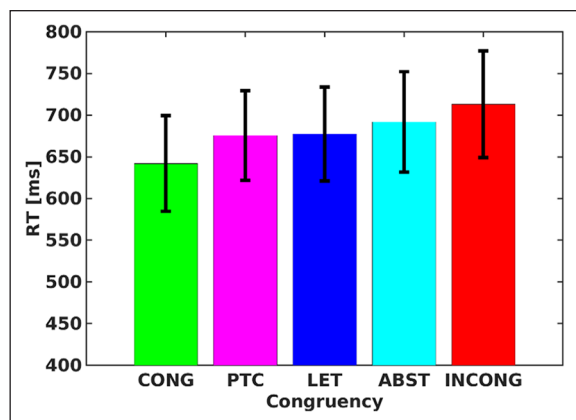
**Figure 2.** Mean reaction time for each congruency condition of Stroop trials. Error bars represent 1 confidence interval from the mean.
CONG: congruent; PTC: patches; LET: letters; ABST: abstract character strings; INCONG: incongruent.

40% of missing values) in each condition. Pupil data was processed using CHAP software (Hershman et al., 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 40% of missing values. We also excluded trials with no response or with incorrect responses. This resulted in exclusion of one participant. For the 21 remaining participants (14 females, mean age = 24.14 years, $SD$ = 1.9) included in the analysis, pre-processing of pupil data eliminated 7.1% of trials on average. Next, we detected eye blinks by using Hershman et al.'s (2018) algorithm and filled missing values by using a linear interpolation (Hershman & Henik, 2019). Next, the time courses were aligned with the onset of the Stroop stimulus and divided by the baseline (baseline was defined as the average pupil size 500 ms before the stimulus onset).

*RT.* Mean RTs of correct (pupil valid) trials for each participant in each condition were subjected to a one-way repeated-measures analysis of variance (ANOVA) with congruity (congruent, incongruent, neutral letter strings, abstract character strings, and coloured patches) as an independent factor. As expected, an omnibus analysis produced a significant effect, $F(4, 80) = 14.9, p < .001,$ $\eta_p^2 = .43, BF_{10} = 2.523 \times 10^6$ (mean RTs in the various conditions are presented in Figure 2).

Mean RT was faster in congruent trials compared with patches, $F(1, 20) = 12.1, p = .002, BF_{10} = 17.14$. RT to patches and letters was similar, $F(1, 20) < 1, BF_{01} = 4.34$. RT to letters tended to be similar to that of abstract character strings, $F(1, 20) = 2.34, BF_{01} = 1.61$, and RT to abstract

character strings tended to be faster than response in the incongruent trials, $F(1, 20) = 5.11, p = .035, BF_{10} = 1.81$. In addition, the standard interference effect was found (i.e., RTs to letter strings were faster than responses in incongruent trials), $F(1, 20) = 14.86, p < .001, BF_{10} = 36.59$.

*Error rate.* Incongruent trials had an error rate of 13.96%, congruent trials had an error rate of 10.35%, letter trials had an error rate of 14.38%, abstract character strings had an error rate of 13.84%, and patch trials had an error rate of 13.39%. Error rates were subjected to a one-way ANOVA with congruity as an independent factor. The analysis produced a significant effect, $F(4, 80) = 5.121, p = .001, \eta_p^2 = .168, BF_{10} = 30.219$. Mean error rate was smaller for congruent trials compared with other trials, $F(1, 20) = 22.82, p < .001, BF_{10} = 244.141$. No difference was found between letter, abstract character strings, patch and incongruent trials, $F(3, 60) < 1, BF_{01} = 10.96$.

*Pupillometry.* Mean relative changes of the pupil size in each condition are presented in Figure 3a (see Figure 3b for the detailed Bayes factor figures). Significant differences are presented in Figure 3a by the horizontal lines (e.g., the top two horizontal lines present significant differences between congruent and incongruent conditions).

Our analysis indicates significant differences between all the investigated conditions. Specifically, the differences between incongruent (red line) and congruent (green line) conditions appeared at about 1,260 ms after the stimulus onset. These differences stayed for about 240 ms (until about 1,500 ms after the stimulus onset). Moreover, the differences between the congruent (green line) trials and letters (blue line) appeared at about 620 ms after the stimulus onset (before the appearance of the differences between congruent and incongruent trials). These differences stayed for about 250 ms (until about 870 ms after the stimulus onset). Differences were also found between letters (blue line) and abstract character strings (light blue line) around 470 ms after the stimulus onset for about 760 ms. In addition, differences between abstract character strings (light blue line) and patches (purple line) appeared early on at about 400 ms after the stimulus onset and stayed until the end of the trial. Consistent with previous studies, differences were also found between incongruent (red line) and letter (blue line) conditions. These differences appeared early on at about 1,360 ms after the stimulus onset and stayed until about 1,560 ms post stimulus onset (for about 200 ms).

## Discussion

We conducted a colour-word Stroop task (MacLeod, 1991) with four colour responses and measured both RT and pupil dilation. This experiment used the same procedure of
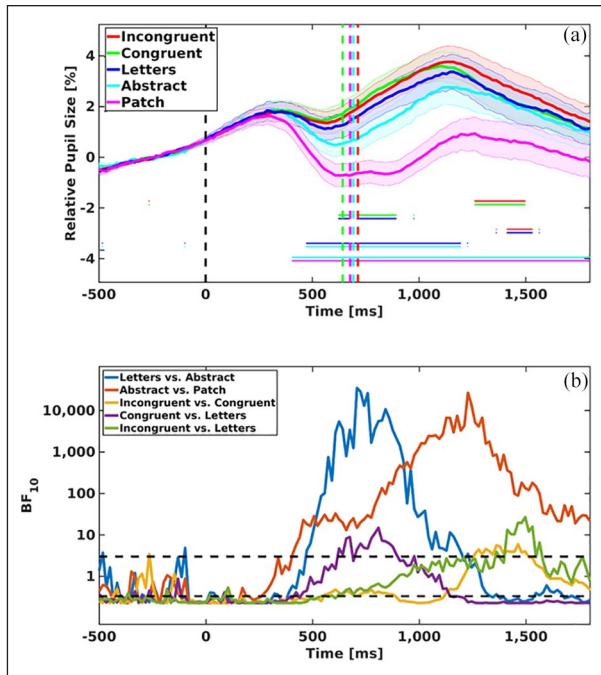
**Figure 3.** (a) Mean relative pupil size (compared to average pupil size 500 ms before the stimulus onset) for the five congruency conditions. Participants had to respond to the ink colour of the stimulus. Time 0 represents stimulus onset and the vertical lines represent mean manual RT (around time 650 ms post stimulus onset) for each condition. The five line curves present changes in pupil dilation as a function of time. The shaded areas represent 1 standard error from the mean. The horizontal lines represent meaningful comparisons ($BF_{10} \geqslant 3$) for each contrast (e.g., the top red and green lines indicate meaningful differences in pupil response between the incongruent and the congruent conditions). (b) BFs as function of time for each comparison, from 500 ms before the stimulus onset to 1,800 ms after the stimulus onset. Each curve represents $BF_{10}$ (namely, evidence for the alternative hypothesis that two conditions are not the same) for the specific comparison. The horizontal black lines on 3 and on 0.3 represent the thresholds for the decision making ($BF_{10}$ values above 3 provide evidence for the alternative hypothesis and $BF_{10}$ values below 1/3 provide evidence for the null hypothesis). Please note that the scale for the Y-axis is exponential.

Hershman et al.'s (2020) study that tested different types of neutrals. In contrast to the neutrals that were used in Hershman et al.'s study, here we used three types of neutrals: letter strings, abstract character strings, and coloured patches. Although the letter strings have phonological meaning, the abstract character strings and the coloured patches have no phonological, lexical, or semantic meaning. Moreover, the coloured patches have no structure that may be associated with readable words (i.e., coloured patches have no orthographical features). Our results showed that RT was faster in congruent trials compared with all the investigated neutral trials (namely, letter strings, coloured patches, and abstract character strings).

In addition, the neutral trials provided faster responses compared with incongruent trials. Analysis of the changes in pupil dilation provided a similar pattern to that of the pilot experiment. Specifically, pupil dilation was larger in the incongruent trials compared with the congruent trials, which was larger than for the letter strings. In line with the pilot experiment, pupil dilation was smaller for the coloured patches compared with all the other investigated conditions. Moreover, abstract character strings led to more dilation compared with coloured patches but smaller dilation compared with letter strings.

In line with previous Stroop and pupillometry studies (Hershman et al., 2020; Hershman & Henik, 2019, 2020), our results indicate the existence of both information and task conflicts. Specifically, the difference in pupil size response to incongruent compared with congruent trials is indicative of information conflict. In addition, our results show larger dilation in the congruent trials compared with the neutral trials (letter strings, abstract character strings, and coloured patches). These reverse facilitations are typically used as markers of task conflict (i.e., respond to the colour vs read the word). In addition to the evidence for the existence of these conflicts, our results provide evidence for the temporal priority of the task conflict compared with the information conflict. Although the evidence for the information conflict (i.e., the difference between congruent and incongruent conditions) appears at an advanced stage of the trial (about 1,200 ms post stimulus onset), the evidence for the task conflict (i.e., the difference between congruent and neutral conditions) appears earlier (about 500 ms post stimulus onset).

Interestingly, the time point of the divergence between the neutrals and the congruent trials was also the time point of the divergence between the neutrals themselves. Specifically, at this time point (about 500 ms post stimulus onset), pupil dilation showed differences between the different types of neutrals. The differences between letter strings and coloured patches were initiated at this time point. Moreover, pupil dilation showed differences between letter strings, coloured patches, and also abstract character strings. These results eliminate the artefact of the presentation type suggested by the pilot results (i.e., larger dilation due to the larger number of objects or due to significant different visual properties). In other words, the differences between coloured patches and any other meaningless stimulus cannot be explained only by the presentation of the stimuli. Specifically, when the presented stimuli were strings of an abstract character, smaller dilation was observed compared with strings of familiar letters. Hence, we believe that our results suggest that meaningful stimuli (and specifically, stimuli with orthographic or lexical meaning such as a series of Xs) produce task conflict. Moreover, our results suggest that not all stimuli produce the same amount of task conflict. Specifically, the more a stimulus is abstract and meaningless (i.e., coloured

patch)—and as a consequence requires less mental effort—the smaller the task conflict will be (Augustinova & Ferrand, 2014; Kinoshita et al., 2017; Levin & Tzelgov, 2016; Monsell et al., 2001).

In this study, the RTs suggest that all the investigated neutrals had similar influence on the response. Moreover, in the pilot experiment, the analysis of the RTs indicated longer responses to coloured patches than to letter strings. In contrast, pupil dilation provided a clear-cut different pattern of dilation for each investigated neutral type. In fact, pupil dilation was considerably smaller for coloured patches compared with all the other investigated conditions and specifically, compared with the letter strings. The same pattern (i.e., smaller dilation compared with all the investigated conditions and specifically compared with the letter strings) occurred for the abstract character strings as well.

In line with our previous studies (Hershman & Henik, 2019, 2020), our results indicate an interference effect (i.e., pupil dilation in the incongruent trials was larger compared with the neutral trials). However, in contrast to the duration of the time window of the effect, here the effect was pretty short (about 200 ms compared with longer effects of more than 1,000 ms). In this study, we used three kinds of neutrals (instead of one in our previous studies). These results may suggest that the more the variability of the neutral stimuli, the more the cognitive control that will appear.

We would like to mention the relationship between RT and pupil changes. The reverse facilitation that serves as a marker of task conflict is usually obtained with measurement of RT when task conflict is high enough (for a detailed discussion, see Entel et al., 2015). In contrast, the current and previous Stroop and pupillometry studies (Hershman et al., 2020; Hershman & Henik, 2019, 2020) showed that there is no need to increase the likelihood of task conflict to observe the reverse facilitation when using measurement of pupil dilation. Moreover, in previous RT experiments (Hershman et al., 2020; Kinoshita et al., 2017), response to neutrals did not differ from one another when manual responding was used. In contrast, the current experiments showed significant differences among neutrals when pupil changes were measured. Hence, we can conclude that task conflict and additional differences between conditions that are concealed in RT measurements are easily detected with pupillometry.

Recently, task conflict has attracted attention (Littman et al., 2019) and researchers tend to suggest that task conflict is a major component of Stroop interference (Levin & Tzelgov, 2014). Interestingly, in a recent study by Ferrand et al. (2020), it was shown that for the first-grader participants, the magnitude of the Stroop interference was a result of only task conflict, whereas for third and fifth graders, more cognitive conflicts were present (i.e., task, semantic, and response conflicts).

Here, in addition to the evidence for the existence of the task conflict, we found that multiple levels of task conflict could be observed by using different levels of meaningfulness. In the last two decades, researchers have tried to produce task conflict by manipulation of the readability of the stimulus (Kinoshita et al., 2017; Monsell et al., 2001) or by manipulating the semantic meaning of the stimulus (Kinoshita et al., 2018; Levin & Tzelgov, 2016). Here, we suggest that task conflict can also be revealed (and maybe more clearly) by using meaningless stimuli-like series of letters (and not only with words or word-like stimuli). Specifically, by using the meaningless original Stroop (1935) neutrals (i.e., coloured patches), we significantly reduced the task conflict as reflected by smaller pupil dilation. Our findings suggest that coloured patches (meaningless stimulus) as well as strings of abstract characters trigger less task conflict compared with colour words, and moreover, compared with meaningless neutral strings. Hence, researchers who run Stroop tasks, and specifically, researchers who want to eliminate task conflict from the task, should consider using these meaningless stimuli. At the same time, researchers who are interested in eliminating task conflict in other Stroop-like tasks (as well as object-interference tasks, La Heij et al., 2010, and affordance tasks, Phillips & Ward, 2010) should consider adopting equivalent types of neutrals in their experiments.

## Author contributions

R.H. contributed to conceptualisation, data collection, analysis, software, validation, and writing of the manuscript. Y.L. contributed to conceptualisation, writing, and reviewing. J.T. contributed to statistical consultation and reviewing. A.H. contributed to supervision, consultation, reviewing, and editing.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## ORCID iDs

Ronen Hershman https://orcid.org/0000-0002-2601-4393
Avishai Henik https://orcid.org/0000-0002-5665-6004

## Supplementary material

The supplementary material is available at qjep.sagepub.com.

## References

Augustinova, M., & Ferrand, L. (2012). The influence of mere social presence on Stroop interference: New evidence from the semantically-based Stroop task. *Journal of Experimental Social Psychology*, *48*(5), 1213–1216. https://doi.org/10.1016/J.JESP.2012.04.014

Augustinova, M., & Ferrand, L. (2014). Automaticity of word reading: Evidence from the semantic Stroop paradigm. *Current Directions in Psychological Science*, *23*(5), 343–348. https://doi.org/10.1177/0963721414540169

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. https://doi.org/10.1016/j.actpsy.2017.03.009

Brown, T. L. (2011). The relationship between Stroop interference and facilitation effects: Statistical artifacts, baselines, and a reassessment. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 85–99. https://doi.org/10.1037/a0019252

Entel, O., Tzelgov, J., Bereby-Meyer, Y., & Shahar, N. (2015). Exploring relations between task conflict and informational conflict in the Stroop task. *Psychological Research*, *79*(6), 913–927. https://doi.org/10.1007/s00426-014-0630-0

Ferrand, L., Ducrot, S., Chausse, P., Maïonchi-Pino, N., O'Connor, R. J., Parris, B. A., . . .Augustinova, M. (2020). Stroop interference is a composite phenomenon: Evidence from distinct developmental trajectories of its components. *Developmental Science*, *23*(2), e12899. https://doi.org/10.1111/desc.12899

Fox, L. A., Shor, R. E., & Steinman, R. J. (1971). Semantic gradients and interference in naming color, spatial direction, and numerosity. *Journal of Experimental Psychology*, *91*(1), 59–65. https://doi.org/10.1037/h0031850

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

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*, *45*(10), 1899–1909. https://doi.org/10.1037/xlm0000690

Hershman, R., & Henik, A. (2020). Pupillometric contributions to deciphering Stroop conflicts. *Memory & Cognition*, *48*(2), 325–333. https://doi.org/10.3758/s13421-019-00971-z

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

Hershman, R., Levin, Y., Tzelgov, J., & Henik, A. (2020). Neutral stimuli and pupillometric task conflict. *Psychological Research*. Advance online publication. https://doi.org/10.1007/s00426-020-01311-6

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

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., Goldfarb, L., & Henik, A. (2013). Evidence for interaction between the stop signal and the Stroop task conflict. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 579–592. https://doi.org/10.1037/a0027429

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

Kinoshita, S., De Wit, B., & Norris, D. (2017). The magic of words reconsidered: Investigating the automaticity of reading color-neutral words in the Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(3), 369–384. https://doi.org/10.1037/xlm0000311

Kinoshita, S., Mills, L., & Norris, D. (2018). The semantic Stroop effect is controlled by endogenous attention. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(11), 1730–1742. https://doi.org/10.1037/xlm0000552

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *The American Journal of Psychology*, *77*(4), 576–588. https://doi.org/10.2307/1420768

La Heij, W., Boelens, H., & Kuipers, J.-R. (2010). Object interference in children's colour and position naming: Lexical interference or task-set competition? *Language and Cognitive Processes*, *25*(4), 568–588. https://doi.org/10.1080/01690960903381174

Levin, Y., & Tzelgov, J. (2014). Conflict components of the Stroop effect and their "control." *Frontiers in Psychology*, *5*, Article 463. https://doi.org/10.3389/fpsyg.2014.00463

Levin, Y., & Tzelgov, J. (2016). What Klein's "semantic gradient" does and does not really show: Decomposing Stroop interference into task and informational conflict components. *Frontiers in Psychology*, *7*, Article 249. https://doi.org/10.3389/fpsyg.2016.00249

Littman, R., Keha, E., & Kalanthroff, E. (2019). Task conflict and task control: A Mini-review. *Frontiers in Psychology*, *10*, Article 1598. https://doi.org/10.3389/FPSYG.2019.01598

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

Monsell, S., Taylor, T. J., & Murphy, K. (2001). Naming the color of a word: Is it responses or task sets that compete? *Memory and Cognition*, *29*(1), 137–151. https://doi.org/10.3758/BF03195748

Parris, B. A., Augustinova, M., & Ferrand, L. (2019). Editorial: The locus of the Stroop effect. *Frontiers in Psychology*, *10*, Article 2860. https://doi.org/10.3389/fpsyg.2019.02860

Phillips, J. C., & Ward, R. (2010). Visual cognition S-R correspondence effects of irrelevant visual affordance: Time course and specificity of response activation. *Visual Cognition*, *9*(4–5), 540–558. https://doi.org/10.1080/13506280143000575

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

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

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