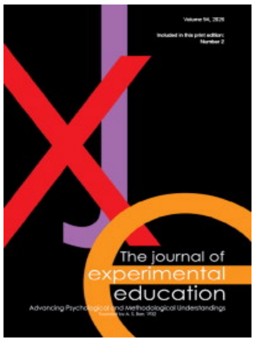


The Journal of Experimental Education

ISSN: 0022-0973 (Print) 1940-0683 (Online) Journal homepage: www.tandfonline.com/journals/vjxe20

Evidence for Both Task and Information Conflicts in the Color-Digit Stroop Task

Ronen Hershman, Lisa Beckmann, Eldad Keha, Avishai Henik & Ayelet Sapir

THE JOURNAL OF EXPERIMENTAL EDUCATION
https://doi.org/10.1080/00220973.2025.2565172

Routledge
Taylor & Francis Group

LEARNING, INSTRUCTION, AND COGNITION

OPEN ACCESS Check for updates

Evidence for Both Task and Information Conflicts in the Color-Digit Stroop Task

Ronen Hershman[a], Lisa Beckmann[b], Eldad Keha[c,d], Avishai Henik[e], and Ayelet Sapir[f]

[a]University of Innsbruck, Innsbruck, Austria; [b]Ludwig-Maximilians-Universität München, Munich, Germany; [c]The Hebrew University, Jerusalem, Israel; [d]Achva Academic College, Beer-Tuvia, Israel; [e]Ben-Gurion University of the Negev, Beer-Sheva, Israel; [f]University of Greenwich, London, UK

ABSTRACT

The color-digit Stroop task was used to examine information and task conflicts through response times (RTs) and pupil dilation. RTs showed typical interference and facilitation (slower incongruent, faster congruent), suggesting information conflict. Pupil data showed greater dilation for incongruent versus congruent trials later in the trial, indicating information conflict. Both congruent and incongruent trials also showed greater dilation than neutral trials, reflecting task conflict not visible in RTs. Task conflict appeared early, suggesting automatic processing of numerical meaning, whereas information conflict emerged later. This dissociation demonstrates that physiological measures capture hidden cognitive challenges. The color-digit Stroop task, being simple and language- (and culture-) independent, may serve as a promising tool for studying mathematical learning difficulties and guiding interventions.

KEYWORDS

Cognitive control; information conflict; numerical cognition; pupillometry; task conflict

1. Introduction

Since its inception in the 1930s, the color-word Stroop task (Stroop, 1935) has captured the attention of researchers and clinicians alike as they seek to examine the interactions between attention, inhibition, and automaticity (Henik et al., 2018). In addition to its contribution to theoretical knowledge, this cognitive paradigm has been used in a variety of contexts, from understanding cognitive development in healthy people to identifying and tracking cognitive deficits in clinical populations (MacLeod, 1991). In the frequently used color-word Stroop task (Henik et al., 2018), participants are presented with colored stimuli and are asked to name the color of the ink. The stimuli could be color words with a congruent color (i.e., the word BLUE written in blue ink), color words with an incongruent color (i.e., RED written in blue), or neutral non-word stimuli (i.e., XXXX written in blue). Presenting color words in incongruent colors leads to longer response times (RTs) when compared with the neutral condition. This RT difference has been termed "the interference effect" to reflect the cost imposed by reading the incongruent word. The congruent condition leads to faster RTs compared to the neutral conditions. This RT difference has been termed "the facilitation effect". The interference effect is often large and considered robust, while the facilitation effect is small and relatively fragile (Hershman & Henik, 2019; Kalanthroff & Henik, 2013; MacLeod, 1991). Since its early start, a variety of Stroop-like

CONTACT Ronen Hershman ronen.hershman@uibk.ac.at Department of Psychology, University of Innsbruck. Innrain 52f, 6020, Innsbruck, Austria.

paradigms have been introduced, such as the spatial Stroop task (Lu & Proctor, 1995), the auditory Stroop task (Jerger et al., 1988), the emotional Stroop task (Ben-Haim et al., 2016), and the numerical Stroop task (Besner & Coltheart, 1979; Henik & Tzelgov, 1982).

In the numerical Stroop task (Besner & Coltheart, 1979; Henik & Tzelgov, 1982), participants are presented with two digits that differ in physical size and/or numerical value and are asked to compare either the physical or the numerical dimension. The physical task resembles the color-word Stroop task, where participants are required to focus on the physical attribute (such as the font color or the physical size) while inhibiting a strong automatic response, such as reading the word or evaluating the magnitude of the digit. Similar to other Stroop-like tasks, the numerical Stroop task shows the typical finding of interference and facilitation effects, with the RTs to congruent conditions being faster than neutral conditions, which in turn are faster than incongruent conditions.

Another type of numerical Stroop-like task is the counting Stroop task is the counting Stroop task (Bush et al., 1998). During the counting Stroop, participants are presented with several numbers of words and asked to report the number of words on the screen (varied from one to four), regardless of the meaning of the words. In the neutral condition, non-number words are presented (e.g., the word "DOG" written one to four times), while in incongruent trials, number words are inconsistent with the correct response (e.g., the word "ONE" written four times). Healthy participants typically show a reliable interference effect. Recently, we introduced a new numerical Stroop-like task, the color-digit Stroop task (Hershman et al., 2025). In this task, participants are presented with digits made out of colored stripes and are asked to report the number of different colors while suppressing the irrelevant dimension of the numerical value of the digit. Similar to other Stroop-like tasks, this experiment revealed both facilitation and interference effects, supporting the automaticity of symbolic number processing.

The color-digit task imposes relatively low cognitive demands, as it neither requires reading skills as in the widely used color-word Stroop task nor involves a comparison process, as in the numerical Stroop task (e.g., Besner & Coltheart, 1979; Henik & Tzelgov, 1982). This allows for the examination of interference processing without the confounding effects of attention or reading fluency. As a language-independent measure, the task is particularly well-suited for use with pre-literate children and individuals with limited or impaired reading abilities, such as certain patient populations. Notably, it was found that the same pattern of results emerged across samples from different countries and even different continents (Hershman et al., 2025). These findings highlight the task's cultural robustness and suggest that the color-digit Stroop task provides consistent and generalizable measures of interference control across diverse populations.

Interestingly, facilitation and interference are not the only effects found in Stroop-like tasks. It has been found that under certain conditions, two separate conflicts arise: information conflict and task conflict (Goldfarb & Henik, 2007; Littman et al., 2019; Monsell et al., 2001). The information conflict is the result of conflicting information on the relevant and irrelevant dimension (e.g., the color of the word and the meaning of the word in the color-word Stroop task), and it is typically assessed by the difference in RT between incongruent (consists of conflicting information) and congruent (consists of matching information) trials. The task conflict results from two competing tasks (e.g., naming the color of the word and reading the word in the color-word Stroop task) and can be revealed by the difference between conditions that consist of two tasks (both congruent and incongruent trials) and conditions that only consist of one task (non-word neutral trials). Typically, task conflict arises in situations in which one task is explicit (e.g., naming the color of the stimulus in the color-word Stroop task), and the second task is more implicit and automatic, and it is a result of automatic activation of driven behaviors which are strongly associated with specific stimuli (Kiesel et al., 2007; Littman et al., 2024; Rogers & Monsell, 1995; Waszak et al., 2003). For example, words tend to automatically evoke reading (Monsell et al., 2001). Task conflict is usually observed in the form of a reverse facilitation effect, namely, congruent trials are slower than neutral trials. However, it mainly occurs under specific restrictions. For example, Goldfarb and Henik (2007) demonstrated that task conflict is seen when control is dramatically reduced by increasing the proportion of non-word neutral trials. Likewise, increasing the difficulty level of processing the irrelevant dimension of the stimulus can also reveal task conflict (Hershman et al., 2025; Hershman et al., 2025) Alternatively, it's been shown that tracking pupil size can be used to expose task conflict without these restrictions (Hershman et al., 2024;; Hershman et al., 2021, 2021, 2022; Hershman & Henik, 2019, 2020).

While the pupil size is changed as a physiological response to the brightness of the stimuli (Ellis, 1981), it is also well known to reflect mental effort in cognitive control tasks, with pronounced dilation when more effort is required (van der Wel & van Steenbergen, 2018). This is true for the Stroop color-word task (Hershman & Henik, 2019; Laeng et al., 2011; Siegle et al., 2004) and for the numerical Stroop task (Hershman et al., 2022) as well. While the RTs in a Stroop (and Stroop-like) task often show facilitation and interference, the pattern of the pupil response is different: pupillometry results reveal an interference effect, characterized by more pronounced pupil dilation in the incongruent condition compared to the congruent condition (Hershman et al., 2024;; Hershman et al., 2021, 2021; Hershman & Henik, 2019), presumably indicating the additional effort required in these trials; but also reverse facilitation. Specifically, a larger pupil dilation is found in the congruent compared to the neutral condition when the neutral stimulus is meaningless (e.g., colored patches). Analyzing the pupil size across the duration of a trial indicates a task conflict relatively early on, at about 500 ms post-stimulus onset (Hershman and Henik 2019). Meanwhile, evidence for information conflict starts later at about 1,000 ms after the stimulus onset.

When pupillometry was used during a numerical Stroop task, the pupil dilation echoed the findings from the color-word Stroop task (Hershman et al., 2022). When participants were asked to report the physical dimension, evidence for task conflict could be seen earlier than the evidence for information conflict. This, along with the previous results from Hershman and Henik (2019), suggests that task conflict appears prior to information conflict and, therefore, further suggests that the numerical processing of symbolic numbers is an automatic process.

1.1. The present study

The purpose of the present study was to investigate the presence of both task and information conflicts in the color-digit Stroop. Specifically, we examined whether the color-digit Stroop task would show the same dissociation of RT and pupil dilation as was found previously in both the color-word Stroop task (Hershman et al., 2021; Hershman & Henik, 2019) as well as the numerical Stroop task (Hershman et al., 2022) that is, evidence for interference and facilitation with RTs and evidence for both interference and reverse facilitation (evidence for task conflict) with pupil size. Furthermore, we examined the time course of the effects, asking whether task conflict, emphasizing the automatic processing of the irrelevant dimension (i.e., symbolic number processing), would appear before information conflict.

By conducting this kind of experiment, we aimed to address a few important aspects regarding both cognitive control and numerical processing. First, we aimed to replicate our finding with the color-digit Stroop task. That is, we aimed to revalidate our behavioral findings as observed in the original task (Hershman et al., 2025). In the present study, we conducted the experiment in the lab (in contrast to the original online measurement). In addition, while the interference and facilitation effects are well replicated, there is a lack of literature regarding the observations of task conflict without designated manipulations. The presentation of the two conflicts (i.e., both information and task conflicts) in a different Stroop-like task might help generalize the concept of task conflict to additional domains. The combination of these two aspects (i.e., a replication of the findings in the color-digit Stroop task and the generalization of findings of task conflict as observed with pupillometry) might open the door to further studies that will examine cognitive control processes in terms of numerical cognition and specifically, it could examine task and information conflicts in the acquisition of mathematical skills, dyscalculia, and acalculia. The color-digit Stroop task is language-independent, and it does not require advanced reading abilities. As such, it is optimal for the examination of cognitive control conflicts with participants with reading disabilities.

Accordingly, we conducted the color-digit Stroop experiment and measured changes in pupil size. Participants were asked to report the number of colored stripes that made a digit stimulus. This could be congruent (e.g., digit 2 consists of two colored stripes), incongruent (e.g., digit 4 consists of two colored stripes), or neutral (e.g., a rectangle consists of two colored stripes). Both RT and pupil dilation served as dependent measures. The present experiment was conducted in the lab, allowing us to replicate findings from the original online version (Hershman et al., 2025) under controlled conditions and to further examine the presence and time course of both task and information conflicts using pupillometry. We aimed to contribute to the literature by replicating established interference/facilitation effects and by demonstrating task conflict without the need for specific manipulations, thus helping generalize the concept of task conflict across domains.

1.2. Hypotheses

1.2.1. Hypothesis 1

When looking at RTs, we hypothesized that both interference (slower responses to incongruent trials compared to neutral trials) and facilitation (faster responses to congruent trials compared to neutral trials) effects would be observed in RTs, replicating the behavioral patterns found in the original color-digit Stroop study (Hershman et al., 2025). These effects are consistent with findings from the color-word Stroop and other Stroop-like paradigms. Consequently, we also expected to observe a congruency effect, characterized by slower responses to incongruent trials relative to congruent ones. Replicating these effects under laboratory conditions would provide further evidence for the robustness and reliability of the color-digit Stroop task across different testing environments.

1.2.2. Hypothesis 2

When measuring pupil size, we hypothesized that both A. Information conflict (indexed by greater pupil dilation to incongruent trials compared to congruent trials, i.e., congruency effect) and B. Task conflict (indexed by greater pupil dilation to congruent trials compared to neutral trials, i.e., reverse facilitation) would be observed. These pupillary patterns replicate findings from the color-word Stroop task (Hershman & Henik, 2019) and other Stroop-like paradigms (Hershman et al., 2022).

1.2.3. Hypothesis 3

We hypothesized that the pupil size evidence for task conflict (i.e., reverse facilitation) would emerge earlier in the time course than evidence for information conflict (i.e., the congruency effect). This prediction is based on prior findings showing that task conflict reflects early automatic processing of the irrelevant dimension (symbolic number processing) and typically precedes information conflict in time (Hershman et al., 2022; Hershman & Henik, 2019).

2. Method

2.1. Participants

We requested a required sample size based on an expected effect size of $\eta_p^2 = .25$ (corresponding to a Cohen's $f = 0.577$), for 3 types of congruency. With α set at .05 and desired power at .99, the analysis indicated that a total of 19 participants would be required to detect a main effect. Twenty native Hebrew speakers (15 females, mean age 23.95 years, $SD = 1.1$) from Ben-Gurion University of the Negev participated in the experiment in return for 25 shekels (approximately $7) or course credit. The study was approved by the ethics committee of the Psychology Department. All participants signed an informed consent form prior to participation in the experiment. Each participant had normal vision (without glasses or contact lenses), as well as normal color vision, and no reported history of attention deficit disorder or any learning disabilities.

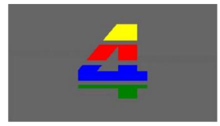
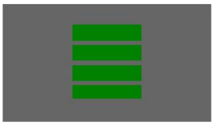
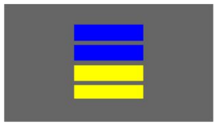
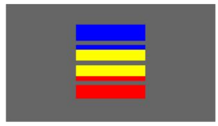
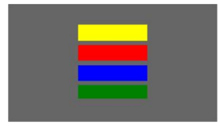
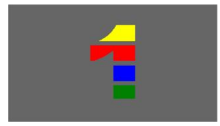
2.2. Stimuli

In line with Hershman et al. (2025), participants were presented with colored single-digit numbers in Arial font. The digits were 1, 2, 3, and 4 in a size of 640×740 pixels. The single digits could consist of a different number of colors between 1 and 4 (red, blue, green, yellow). Specifically, in each of the four colored digits, the stimuli were divided horizontally into four equal-sized parts. The number of colors could have been congruent with the numerical value of the digit (e.g., the digit 3 that is painted with three colors; mean RGB = 101.98) or incongruent (e.g., the digit 3 that is painted with two colors; mean RGB = 101.98). In addition to the colored single-digit numbers, participants were also presented with colored rectangles of the same size and colors (mean RGB = 101.96), which were used as neutral stimuli in the current experiment. Using rectangles as neutral stimuli was suggested to require less processing and, as such, is expected to cause less task conflict (Hershman et al., 2021, 2022). The conditions and the stimuli for each participant were selected randomly from a pool that included all the possible combinations. In line with Exp.2 in Hershman et al.'s study (2024), all the stimuli were also cut into four pieces by removing five pieces from the stimuli. Specifically, the upper and lower 20 pixels were removed from the stimuli. In addition, the pixels between 166 and 205, 351 and 390, and 536 and 575 were also removed. The stimuli appeared against a gray (RGB: 102, 102, 102) background (see Table 1 for examples of the stimuli). The color of the background was defined to reduce the luminance differences between the conditions.

2.3. Procedure

The experiment was conducted in a dimly illuminated room, and the participants were tested individually. A keyboard for measuring responses was placed in between the participant and a monitor on a desk. Participants positioned their heads on a chinrest support to prevent head movement. The experiment included ten practice trials that were excluded from the analysis. After each practice trial, participants received feedback on their accuracy. Participants were required to reach an accuracy rate of at least 80% to proceed to the experimental part; otherwise, they had to repeat the practice block. This criterion was intended to ensure that participants adequately understood the task before continuing. In the experimental part, participants carried out two blocks of 216 experimental trials each. In total, 144 trials were presented per congruency condition. At the beginning of each trial (see *Figure 1* for a visual example), there was a 1,000 ms fixation in the form of a black cross at the center of the screen. The experimental stimulus then appeared for 400 ms and was followed by a blank screen for a maximum of 1,100 ms or until a keypress. Each trial ended with a 1,500 ms inter-trial interval (ITI) of a blank (gray) screen.

Table 1. Examples for the presented stimuli in Exp. 1 (the entire set of stimuli can be downloaded from: https://osf.io/evgm4/?view_only=3f0112fef54f4e669c1a6a7fe99a9d72).

Condition	1 color	2 colors	3 colors	4 colors
Congruent				
Neutral				
Incongruent				

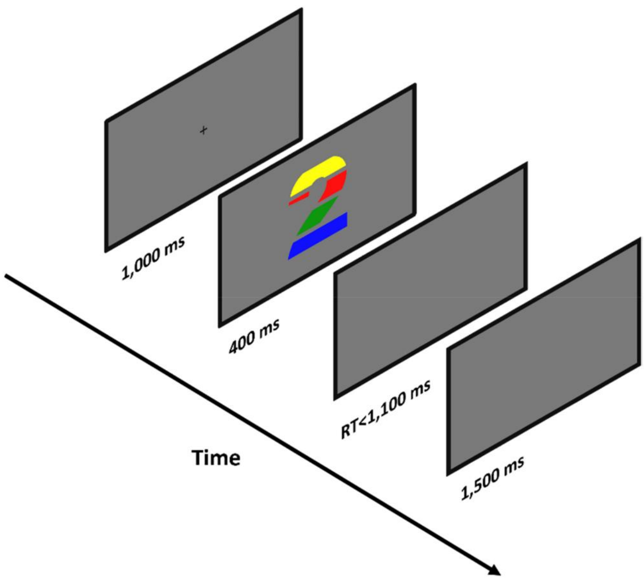


Figure 1. Example of an incongruent trial.
Note. Participants had to report the number of colors in the presented digit (the correct response is 4 in this example).

Participants were asked to hit the “Z” key if the stimulus consisted of 1 color, the “X” key if the stimulus consisted of 2 colors, the “N” key if the stimulus consisted of 3 colors, and the “M” key if the stimulus had 4 colors. Participants responded with both hands, in a horizontal arrangement of responses that is regularly used in Stroop and Stroop-like tasks (Hershman et al., 2022; Hershman & Henik, 2019). RT was calculated from the appearance of the visual stimulus to the reaction in the form of a key press. Due to the length of the experiment, the participants had one break in which they were told that they could close their eyes but keep their head static, resting on the chinrest.

2.4. 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 chinrest, and the distance from the eyes to the monitor was set at about 50 cm. Participants were required to keep their eyes fixated on the center of the screen and to avoid eye movements for the entire task. The pupil area was determined using the Eye Tribe algorithm.

2.5. Pre-processing of the pupillometry data

Two participants were excluded from the analysis because they did not have a minimum of 80 valid trials (correct response trials with no more than 30% of missing pupillometric samples) in each condition. For the 18 remaining participants (13 females and five males, mean age = 24 years old, SD = 1.14), 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, outlier samples with Z-scores larger than 2.5 (based on the mean and standard deviation calculated for each trial) were removed. For each participant, we excluded from the analysis the trials with more than 30% missing values, as well as trials with no response or with incorrect responses. This pre-processing eliminated, on average, 7.26% of trials. Next, we detected eye blinks by using Hershman et al. (2018) algorithm and filled the missing values by using linear interpolation (Hershman & Henik, 2019). Finally, time courses were aligned with the onset of the stimulus and divided by the baseline, which was defined as the average pupil size 500 ms before the stimulus onset.

3. Results

3.1. Reaction time

Mean RTs of correct trials with valid pupil data for each participant in each condition were subjected to a one-way repeated measures analysis of variance (ANOVA) with congruency (congruent, incongruent, and neutral) as an independent factor (mean RTs in the various conditions are presented in Figure 2). As expected, an omnibus analysis produced a significant effect for congruency, $F\ (2,34) = 74.24, p < .001, \eta^2_p = .81, BF_{10} > 10^9$. Specifically, mean RT in incongruent trials was slower than in neutral trials, $t(17) = 8.78, p < .001, BF_{10} > 1,000, Cohens'd = 0.64$; mean RT in neutral trials was slower than in congruent trials, $t(17) = 2.94, p = .006, BF_{10} > 10^4, Cohens'd = 0.21$; and mean RT in incongruent trials was slower than in congruent trials, $t(17) = 11.71, p < .001, BF_{10} > 10^6, Cohens'd = 0.86$. It means that RT analysis revealed both interference and facilitation effects. In other words, the results indicated the existence of an information conflict but not a task conflict. Assumption testing revealed that certain assumptions were violated, and therefore, both Greenhouse-Geisser and Huynh-Feldt corrections were applied to adjust for these violations. Notably, both corrections yielded the same results, indicating the robustness of the findings.

3.2. Pupil size

Mean relative changes in pupil size in each condition are presented in Figure 3. In order to examine the temporal differences among the three conditions, we follow Hershman et al.'s (Hershman et al., 2023, 2024) approach, where we compared each pair of the three conditions over the entire time

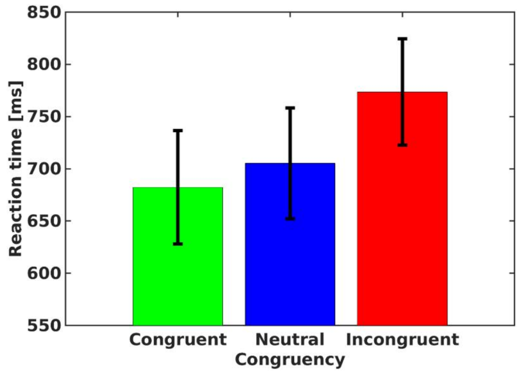


Figure 2. *Mean reaction time for each congruency condition.*
Note. Error bars represent a 95% confidence interval from the mean.

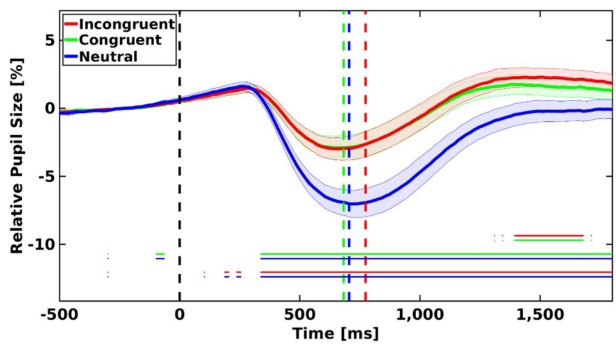


Figure 3. Mean relative pupil size for each congruency condition.
Note. Mean relative pupil size (compared with pupil size at the stimulus onset) for the three congruency conditions (participants had to indicate how many colors were presented). 0 represents stimulus onset, and the colored vertical lines represent mean response times (around 700 ms post-stimulus onset) for each condition. The three 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 meaningful differences (i.e., $BF_{10} \geq 3$) for each contrast (e.g., the red and green horizontal lines indicate meaningful differences in pupil response between the incongruent and the congruent conditions).

course of pupil measurement. Meaningful differences ($BF_{10} \geq 3$) are illustrated in Figure 3 by the horizontal lines at the bottom of the graph. The analysis (Figure 3) indicates that pupil size was larger in both incongruent and congruent trials compared with neutral trials. The meaningful difference is indicated by the double horizontal lines showing that the neutral condition (blue) was different than the congruent (green) and incongruent (red) conditions. These differences were observed at about 340 ms after the stimulus onset and maintained until the end of the trials. In addition, pupil size was larger in incongruent trials compared with congruent trials, as indicated by the red (incongruent) and green (congruent) double horizontal lines. This difference appeared a lot later, at about 1,400 ms after the stimulus onset, and lasted for about 280 ms.

4. Discussion

In the present study, we conducted the color-digit task (Hershman et al., 2025), and we measured both RTs and changes in pupil size. In line with previous behavioral results observed using this task (Hershman et al., 2025; Hershman et al., 2024; Hershman et al., 2025) and in line with most of the Stroop and Stroop-like studies (Henik et al., 2018; Stroop, 1935), both interference (i.e., slower responses for incongruent trials than for neutral trials) and facilitation (i.e., faster responses for congruent trials than for neutral trials) effects were found when we measured RTs

(Hypothesis 1). In addition, and in line with previous Stroop and pupillometry studies (Hershman et al., 2024; Hershman et al., 2022; Hershman & Henik, 2019), a clear dissociation between RTs and pupil dilation was found. First, pupil dilation was greater for incongruent trials compared to congruent ones, particularly during the later stages of the trial. Second, and importantly, an early and very large difference in pupil size was found between neutral trials and both congruent and incongruent trials, so that the neutral trials elicited less pupil dilation. More pupil dilation for incongruent trials than for congruent trials suggests that the two contradicting pieces of information (one from the task-relevant dimension—the actual number of colors, and another one from the meaning of the presented stimulus) cause information conflict (Hypothesis 2 A). In addition, more pupil dilation for congruent trials than neutral trials provides us with evidence for task conflict (i.e., count the number of colors or process the meaning of the task-irrelevant digit) (Hypothesis 2B). The difference in effects onsets is in line with previous Stroop and pupillometry studies (Hershman et al., 2021, 2021, Hershman et al., 2022; Hershman & Henik, 2019, 2020), and it is in line with Goldfarb & Henik (Goldfarb & Henik, 2007) hypothesis regarding the temporal priority of task conflict to that of the information conflict (Hypothesis 3).

The behavioral findings of the present study are aligned with those of previous color-digit Stroop studies (Hershman et al., 2025; Hershman et al., 2024; Hershman et al., 2025). This revalidation of the task that was conducted in-lab (compared to the previous online measurements) suggests that the task is robust and valid. The color-digit Stroop task (Hershman et al., 2025) has several advantages compared to other Stroop and Stroop-like tasks. First, in contrast to the frequently used color-word Stroop task (Henik et al., 2018; Stroop, 1935), the color-digit Stroop task is language- (and culture-) independent. Therefore, it could be used with pre-literate children and individuals (patients) with low/deficient reading without confounding factors of reading fluency and/or attention. In addition, compared to other numerical Stroop tasks like the numerical Stroop task (Henik & Tzelgov, 1982) and the counting Stroop task (Bush et al., 1998), the color-digit Stroop task is visually less complex, and it contains only one object that should be processed. Moreover, compared to the numerical Stroop task (Henik & Tzelgov, 1982) no comparison is required. As a result, the color-digit Stroop task places less load on (visual-) attentional processes. This advantage makes the color-digit Stroop task well-suited for studying interference processing in individuals with working memory difficulties, such as those with neurodevelopmental, neuropsychiatric, or neurological populations.

In the color-digit task, as in the color-word Stroop task (Stroop, 1935) and in the numerical Stroop task (Henik & Tzelgov, 1982), both information and task conflicts were found. It has already been suggested that task conflict is the primary component of the Stroop effect because informational conflict can only arise after engaging in an irrelevant task (Levin & Tzelgov, 2014). Here, we present another (Stroop-like) task that presents this concept. Hence, the results of the present study suggest that task conflict that is raised when stimuli automatically activate associated task (Kiesel et al., 2007; Meiran & Kessler, 2008; Rogers & Monsell, 1995; Waszak et al., 2003; Wendt et al., 2013), is a general phenomenon that extends across a variety of cognitive control tasks.

In the color-digit Stroop task, the task-relevant task is counting. Hence, modulation of counting by the numerical values examines the automaticity of processing the counting or cardinality aspect of the digits. The presentation of task conflict in the present study suggests that the cardinality aspect of the digits is processed early on (also if it might interfere with the completion of the main of counting the number of presented colors). These findings pave the way for future research on cognitive control within the context of numerical cognition. Specifically, they may help explore how task and information conflicts play a role in the development of mathematical skills, as well as in conditions like dyscalculia and acalculia.

In addition to the frequent use of RTs, we also used pupillometry with the aim of examining both information and task conflicts. Pupil dilation is an unobtrusive high-temporal resolution measure that is thought to change as a result of task difficulty (Kahneman & Beatty, 1966). This measurement was used in Stroop studies (Brown et al., 1999; Laeng et al., 2011; Siegle et al., 2004, 2008) and in Stroop studies that examined information and task conflicts (Hershman & Henik, 2019, 2020). Interestingly, information and task conflicts were found using pupillometry, also when an actual motoric response was not required (Hershman et al., 2024). Therefore, pupillometry may be a suitable measurement tool for participants with motor disabilities, such as those with Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis (ALS), Huntington's disease, or individuals recovering from a stroke. The combination of both the language- (and culture-) independent task that requires less amount of mental and visual effort and the unobtrusive sensitivity to task conflict measurement makes the present design suitable for clinical populations that could be challenging to examine by other tasks and measurements.

In contrast to previous Stroop and pupillometry studies, in the present study, instead of an increment of the pupil as a result of the exposure to the stimulus, pupil decrement was observed. This pattern was observed because the presented stimuli were brighter than the background. Pupils respond to light intensity (Ellis, 1981), and therefore, when the stimuli were presented, the pupils responded to light intensity. However, for more demanding trials (in terms of both task and information conflict), a smaller decrease in pupil size was observed.

Although task conflict is assumed to occur whenever two tasks are activated at the same time (Kalanthroff et al., 2018), in most cases, the existence of two matched pieces of information (as in the congruent condition) will lead to a facilitation effect (Parris et al., 2023). That is, when two indications on the same answer appear, the response to the stimuli will be faster than in the case that only one indication appears (as in the neutral condition). However, more sensitive measures, such as changes in pupil size, may capture the cognitive effort required to complete an irrelevant task, even when behavioral performance appears unaffected (Hershman & Henik, 2019; van der Wel & van Steenbergen, 2018). Measurements like pupillometry could be particularly valuable in the study of learning and learning difficulties (Shechter & Share, 2021), as they can uncover underlying cognitive challenges that may not be evident through observable behavior alone.

In the present study, we found that although participants were instructed to count the number of presented colors, they also inadvertently processed the numerical values of the digits. This type of task conflict is independent of reading and reflects a fundamental level of interference (Levin & Tzelgov, 2016). Specifically, the challenge is of selectively attending to the relevant task while suppressing the irrelevant one. The ability to maintain focus on the relevant task is particularly important in situations where the irrelevant task may occasionally appear beneficial or tempting to engage with (Inzlicht et al., 2015; e.g., in the case of congruent trials). Future research should extend these findings by examining how task conflict manifests in individuals with learning difficulties, to better understand the cognitive barriers they may face and how to tailor interventions accordingly. It can help us understand how certain learning difficulties develop, and more importantly, point to the specific cognitive processes that need support or intervention.

Beyond providing valuable insights into the mechanisms of cognitive control during learning more broadly, this study also offers direct implications for mathematical education specifically. By revealing that task conflict arises early and automatically when processing digits, the findings suggest students may engage with numerical information, even when it is irrelevant to the task. This may be relevant for understanding learning difficulties such as dyscalculia and could inform strategies for supporting learners who struggle with number processing (Ashkenazi et al., 2009; Rubinsten & Henik, 2005). Future research should extend these findings by examining how task conflict manifests in individuals with learning difficulties, to better understand the cognitive barriers they may face and how to tailor interventions accordingly.

In addition to the theoretical implications derived from our findings, the results offer clear evidence for automatic numerical processing. This evidence might be difficult to capture through behavioral measures alone. The current task, which taps into basic numerical processing, holds potential as a diagnostic tool for identifying mathematical learning difficulties, such as dyscalculia (Haberstroh & Schulte-Körne, 2019; Kaufmann & Von Aster, 2012) or acalculia (Willmes, 2008). The color-digit Stroop task (Hershman et al., 2025), independent from reading demands, combined with the relatively low cognitive effort required to perform the relevant task, makes it particularly suitable for young children at the early stages of mathematical learning, such as those in kindergarten or first grade. Future developmental research examining both informational and task conflicts, as those explored in the color-digit Stroop task, may further support the diagnostic utility of this paradigm.

5. Summary

In the present study, we used the color-digit Stroop task (Hershman et al., 2025) alongside RTs and changes in pupil size to investigate automatic numerical processing and cognitive control. Participants were asked to count the number of presented colors, but they also unintentionally processed the numerical values of the digits, revealing both task and information conflicts. While RTs showed classic interference and facilitation effects, pupil dilation provided deeper insights demonstrating early task conflict and later information conflict, even in the absence of noticeable behavioral differences. These results support the idea that automatic activation of task-irrelevant information can significantly influence cognitive performance. The findings have both theoretical and practical implications. The color-digit Stroop task is language- and culture-independent, visually simple, and cognitively accessible, making it well-suited for young children and individuals with learning or motor difficulties. Pupillometry, as a sensitive, noninvasive measure of cognitive effort, can detect hidden processing demands and may serve as a valuable tool in identifying learning difficulties such as dyscalculia. Together, the task and measurement approach offer a promising framework for research and potential early diagnosis in numerical cognition and education.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Open practices statement

The code that was used to analyze the pupillometry data is available at: http://in.bgu.ac.il/en/Labs/CNL/chap. All experimental data can be found on the OSF (Open Science Framework):https://osf.io/evgm4/?view_only=3f0112fef54f4e669c1a6a7fe99a9d72. This study was not preregistered.

Funding

This research was supported by the Israel Science Foundation grant No. 359/22 to Avishai Henik.

References

Ashkenazi, S., Rubinsten, O., & Henik, A. (2009). Attention, automaticity, and developmental dyscalculia. *Neuropsychology*, *23*(4), 535–540. https://doi.org/10.1037/a0015347

Ben-Haim, M. S., Williams, P., Howard, Z., Mama, Y., Eidels, A., & Algom, D. (2016). The emotional stroop task: Assessing cognitive performance under exposure to emotional content. *Journal of Visualized Experiments: JoVE*, *2016*(112), e53720. https://doi.org/10.3791/53720

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, *17*(5), 467–472. https://doi.org/10.1016/0028-3932(79)90053-8

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: JINS*, *5*(4), 308–319. https://doi.org/10.1017/s1355617799544020

Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., Mcinerney, S. C., & Rauch, S. L. (1998). The counting Stroop: An interference task specialized for functional neuroimaging-validation study with functional MRI. *Human Brain Mapping*, *6*(4), 270–282. https://doi.org/10.1002/(SICI)1097-0193(1998)6:4<270::AID-HBM6>3.0.CO;2-0

Ellis, C. J. (1981). The pupillary light reflex in normal subjects. *The British Journal of Ophthalmology*, *65*(11), 754–759. https://doi.org/10.1136/BJO.65.11.754

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

Haberstroh, S., & Schulte-Körne, G. (2019). The Diagnosis and Treatment of Dyscalculia. *Deutsches Arzteblatt International*, *116*(7), 107–114. https://doi.org/10.3238/ARZTEBL.2019.0107

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395. https://doi.org/10.3758/BF03202431

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., Beckmann, L., & Henik, A. (2022). Task and information conflicts in the numerical Stroop task. *Psychophysiology*, *59*(9), e14057. https://doi.org/10.1111/psyp.14057

Hershman, R., Beckmann, L., Keha, E., Wagner, M., Kaufmann, L., & Henik, A. (2025). A color-digit Stroop task shows numerical influence on numerosity processing. *Memory & Cognition*, *53*(4), 1055–1066. https://doi.org/10.3758/s13421-024-01631-7

Hershman, R., Dadon, G., Kiesel, A., & Henik, A. (2024). Resting Stroop task: Evidence of task conflict in trials with no required response. *Psychonomic Bulletin & Review*, *31*(1), 353–360. https://doi.org/10.3758/S13423-023-02354-7

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., Keha, E., Beckmann, L., Henik, A., & Sapir, A. (2025). Irrelevant task difficulty modulates the emergence of task conflict. [Manuscript Submitted for Publication]. Department of Psychology, University of Innsbruck.

Hershman, R., Keha, E., Sapir, A., Weiss, E. M., Henik, A., & Kaufmann, L. (2024). Evidence for two types of task conflict in a color-digit Stroop task. *Journal of Cognition*, *7*(1), 54. https://doi.org/10.5334/JOC.386

Hershman, R., Levin, Y., Tzelgov, J., & Henik, A. (2021). Neutral stimuli and pupillometric task conflict. *Psychological Research*, *85*(3), 1084–1092. https://doi.org/10.1007/s00426-020-01311-6

Hershman, R., Levin, Y., Tzelgov, J., & Henik, A. (2021). The contribution of meaning to the detection of task conflict. *Quarterly Journal of Experimental Psychology*, *74*(9), 1553–1561. https://doi.org/10.1177/17470218211001331

Hershman, R., Milshtein, D., & Henik, A. (2023). The contribution of temporal analysis of pupillometry measurements to cognitive research. *Psychological Research*, *87*(1), 28–42. https://doi.org/10.1007/S00426-022-01656-0

Hershman, R., Milshtein, D., & Henik, A. (2024). Processing and analyzing of pupillometry data. In: Papesh, M.H. & Goldinger, S.D. (eds). *Modern pupillometry*: cognition, neuroscience, and practical applications. *Springer, Cham*. 431–466. https://doi.org/10.1007/978-3-031-54896-3_15

Hershman, R., Sapir, A., Keha, E., Wagner, M., Weiss, E. M., & Henik, A. (2025). The contribution of difficulty of an irrelevant task to task conflict. *Quarterly Journal of Experimental Psychology (2006)*, *78*(5), 956–962. https://doi.org/10.1177/17470218241228709

Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive control. *Trends in Cognitive Sciences*, *19*(3), 126–132. https://doi.org/10.1016/J.TICS.2015.01.004

Jerger, S., Martin, R. C., & Pirozzolo, F. J. (1988). A developmental study of the auditory stroop effect. *Brain and Language*, *35*(1), 86–104. https://doi.org/10.1016/0093-934X(88)90102-2

Kahneman, D., & Beatty, J. (1966). Pupil size as related to interest value of visual stimuli. *Science*, *140*(3423), 349–350. https://doi.org/10.1126/science.132.3423.349

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

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

Kaufmann, L., & Von Aster, M. (2012). The diagnosis and management of dyscalculia. *Deutsches Arzteblatt International*, *109*(45), 767–777; quiz 778. https://doi.org/10.3238/ARZTEBL.2012.0767

Kiesel, A., Wendt, M., & Peters, A. (2007). Task switching: On the origin of response congruency effects. *Psychological Research*, *71*(2), 117–125. https://doi.org/10.1007/s00426-005-0004-8

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

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

Levin, Y., & Tzelgov, J. (2016). Contingency learning is not affected by conflict experience: Evidence from a task conflict-free, item-specific Stroop paradigm. *Acta Psychologica*, *164*, 39–45. https://doi.org/10.1016/j.actpsy.2015.12.009

Littman, R., Hochman, S., & Kalanthroff, E. (2024). Reliable affordances: A generative modeling approach for test-retest reliability of the affordances task. *Behavior Research Methods*, *56*(3), 1984–1993. https://doi.org/10.3758/S13428-023-02131-3

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

Lu, C. h., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, *1995 2:22*(2), 174–207. https://doi.org/10.3758/BF03210959

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

Meiran, N., & Kessler, Y. (2008). The task rule congruency effect in task switching reflects activated long-term memory. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(1), 137–157. https://doi.org/10.1037/0096-1523.34.1.137

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

Parris, B. A., Hasshim, N., Ferrand, L., & Augustinova, M. (2023). Do task sets compete in the stroop task and other selective attention paradigms? *Journal of Cognition*, *6*(1), 23. https://doi.org/10.5334/JOC.272

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Rubinsten, O., & Henik, A. (2005). Automatic activation of internal magnitudes: A study of developmental dyscalculia. *Neuropsychology*, *19*(5), 641–648. https://doi.org/10.1037/0894-4105.19.5.641

Shechter, A., & Share, D. L. (2021). Keeping an eye on effort: A pupillometric investigation of effort and effortlessness in visual word recognition. *Psychological Science*, *32*(1), 80–95. https://doi.org/10.1177/0956797620958638

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: Official Journal of the International Organization of Psychophysiology*, *52*(1), 63–76. https://doi.org/10.1016/J.IJPSYCHO.2003.12.010

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

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

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

Wendt, M., Kiesel, A., Mathew, H., Luna-Rodriguez, A., & Jacobsen, T. (2013). Irrelevant stimulus processing when switching between tasks. *Zeitschrift Für Psychologie*, *221*(1), 41–50. https://doi.org/10.1027/2151-2604/a000129

Willmes, K. (2008). Acalculia. *Handbook of clinical neurology*, 88, 339–358. https://doi.org/10.1016/S0072-9752(07)88017-1