Acta Psychologica 226 (2022) 103583

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Conflict monitoring or multi-tasking? Tracking within-task performance in single-item and multi-item Stroop tasks

Laoura Ziaka[a,b,*], Athanassios Protopapas[a]

[a] Department of Special Needs Education, University of Oslo, P.O. box 1140 Blindern, 0318 Oslo, Norway
[b] Department of History and Philosophy of Science, University of Athens, Ano Ilissia University Campus, GR-157 71 Zografos, Greece

ARTICLE INFO

Keywords:
Cognitive control
Conflict monitoring
Multi-tasking
Stroop task

ABSTRACT

Cognitive control is applied in situations that require overriding a habitual and automatic response. The conflict monitoring hypothesis and the Expected Value of Control (EVC) theory as its extension posit a control system responsible for detecting conflicting occasions and adapting to them dynamically within a task. Here we evaluate this prediction in two versions of one of the most popular tasks in cognitive control, namely the Stroop task. We hypothesized that nearby-items interference combines with task interference in the multi-item version effectively turning it into a multi-task that may challenge cognitive control. Adopting an alternative methodology tracking within-task performance, we compared the classical multi-item version of the Stroop task and its single-item counterpart in adults and children. The results revealed a within-task performance decline only in the multi-item version of the task, in both incongruent and neutral conditions, modulated by the presumed maturity of the control system. These findings suggest capacity constraints in control implementation and allocation under conditions requiring parallel execution of multiple cognitive tasks. Task complexity and demands seem to modulate effects on performance. We discuss implications for cognitive control as well as substantial concerns regarding the calculation and use of indices of interference based on the commonly used multi-item version of the Stroop task.

1. Introduction

Control is defined as the ability to identify challenging and conflicting tasks in order to adapt to them and execute them successfully. Refraining from an impulsive, automatic, or default behavior is an act of control (Botvinick et al., 2001; Botvinick et al., 2004; Muraven & Baumeister, 2000). A critical question is how the cognitive system determines how much control is needed to accomplish a task.

To address this issue, Botvinick et al. (2001) proposed the conflict monitoring hypothesis. They assumed that control is required in conflicting situations (e.g., the incongruent condition of the Stroop task) and posited a conflict monitoring system that is responsible for detecting occasions of conflict and effecting on-line adjustments, leading to performance improvements. This was successfully simulated in a connectionist computational model, which, in accordance with the theory, produced greater interference in the initial trials of the task, compared to subsequent trials (Henik et al., 1997, as cited in Botvinick et al., 2001), and reduced interference if incongruent trials are frequent rather than rare (Carter et al., 2000; Lindsay & Jacoby, 1994), because in both cases control is adjusted and highly active. Botvinick et al. located the neuropsychological basis of the conflict monitoring function at the dorsal anterior cingulate cortex (dACC), adjacent to the corpus callosum on the medial surface of the frontal lobe. More recently, Shenhav et al. (2013) proposed the Expected Value of Control (EVC) theory, arguing that during conflicting tasks a cost-benefit analysis optimizes control allocation by increasing control while diminishing the costs of its implementation. This account is an extension of the conflict monitoring hypothesis aiming to approach control evaluation and allocation in greater detail.

Applied in the Stroop task, the EVC theory posits that three core processes are engaged, namely specification, monitoring, and regulation. Specification refers to the decision about the task goal (identity; here, color naming) and how intensively this goal must be pursued (intensity). Monitoring identifies the current state in terms of response conflict—an indicator of control adaptation need—for the system to adjust the dimensions of identity and intensity. Following monitoring and specification of the appropriate control signal, regulation adapts control and influences lower-level processing. For EVC theory, cost and

---

* Corresponding author at: Department of Special Needs Education, University of Oslo, 1140, Blindern, 0318 Oslo, Norway.
E-mail address: laoura.ziaka@isp.uio.no (L. Ziaka).

https://doi.org/10.1016/j.actpsy.2022.103583
Received 2 May 2021; Received in revised form 23 March 2022; Accepted 29 March 2022

rewards are crucial because control signals must be optimal. That is, they must be intense enough, to maximize rewards, but not too intense, to keep cost down. If not, monitoring detects the undesired state, a new optimal signal is selected and specified and, ultimately, lower-level processes are regulated accordingly. Specifically, Musslick et al. (2015) stated that the proposed model, which adjusts control dynamically, can account for sequential adaptation findings and explain the Gratton effect (i.e., that Stroop interference is larger after a congruent trial than after an incongruent one), arguing that "after an incongruent trial the control system chooses to implement a higher control signal (in this case associated with increased drift rate toward the controlled response) leading to faster RTs and fewer errors" (p. 2). Therefore, a key prediction of this theory—and its predecessor, the conflict monitoring hypothesis—is that during a control-demanding task control adaptations occur as the identity and intensity of control are gradually adjusted.

1.1. *The issue of task*

Although real-life situations often require engagement in more than one task, research has primarily focused on single-item tasks for examination of control allocation and implementation (Schuch et al., 2019), leaving the question open as to whether the proposed underlying control mechanisms hold irrespective of the demands posed by task(s).

Specifically, popular tasks used to examine cognitive control, conflict adaptation, and ACC activation are the Stroop task, the Eriksen flanker task, and the Simon task (Nee et al., 2007). For all three tasks competing responses are presented and, consequently, control implementation is required. In particular, in the incongruent condition of the Stroop task participants have to name the color of the ink in which a word is printed, where the word denotes a different color (e.g., "red" in green ink). Performance in the incongruent condition is compared to a neutral condition, in which there is no competition for response selection (e.g., "XXX" in green). Similarly, in the Eriksen flanker task a central stimulus is presented and is flanked by other stimuli. A specific response, for example left keypress, is associated with the central stimulus. Flanking could be identical (e.g., S flanked by S), congruent (e.g., S flanked by H where H is associated with the same key response), or incongruent (e.g., S flanked by L where L is associated with the opposite key response). Finally, in the Simon task, the item could be a colored shape, which could appear in different locations (e.g., left or right) and for which the color dimension corresponds to a specific mapped response (e.g., green-left, red-right). The participants' task is to ignore the spatial location of the item, which could be conflicting (e.g., red shape appearing on the left), and focus only on its color for correct responding (Nee et al., 2007).

Due to the nature of the tasks, especially in the cases of the Eriksen flanker task and the Simon task, single-item presentation is required, that is, each stimulus appears individually on the screen, usually for a fixed period of time or until participants' response, with a temporal gap between trials, namely the interstimulus interval, the duration of which may depend on the experimenter's purposes. Although not obligatory, this is also typically the case for the Stroop task, at least in experimental settings (e.g., Carter et al., 1995; Egner & Hirsch, 2005a; Egner & Hirsch, 2005b; Hinault et al., 2019; Kerns et al., 2004; Pardo et al., 1990; Schulte et al., 2019; Teubner-Rhodes et al., 2019). In contrast, in reading (e.g., Di Filippo & Zoccolotti, 2011; Kapoula et al., 2010; Wang & Gathercole, 2015), self-control (e.g., Bray et al., 2008; DeWall et al., 2007; Neshat-Doost et al., 2008), neuropsychological and aging research (e.g., Arán Filippetti et al., 2021; Burger et al., 2020; Chen et al., 2019; Clemmensen et al., 2020; Ludwig et al., 2010), and, perhaps more importantly, in clinical settings (e.g., Bezdicek et al., 2015; Björngrim et al., 2019; Penner et al., 2012; Periáñez et al., 2021; Rabin et al., 2005; Salo et al., 2001; Scarpina & Tagini, 2017), the Stroop task is most often administered in a multi-item format, aiming to assess attentional and executive functions and deficits (Rabin et al., 2005; Strauss et al., 2006). In the original multi-item presentation used by Stroop (1935) all items appeared on a sheet of paper as an array (100 items arranged in 10 rows and 10 columns; MacLeod, 2005) and participants were asked to read or name the items sequentially as fast as possible. Nowadays, popular clinical versions are the Victoria (Strauss et al., 2006) and Golden version (Golden, 1978; Golden & Freshwater, 2002), with format and administration that are very similar to the original. What is of major importance here is that interference estimates are substantially different between the multi-item and the single-item Stroop tasks (MacLeod, 2005), with the single-item version failing to retain its initial "Stroop-like" (i.e., highly conflicting) nature (Penner et al., 2012).

Specifically, Ludwig et al. (2010) compared the single-item and multi-item version of the Stroop task to examine age-related differences in inhibition and suggested that nearby items in the multi-item version could act as distractors and increase the difficulty of the task, which is evident in the magnitude of interference, especially if distractors are not suppressed. This point is further strengthen by Salo et al. (2001), who proposed that one of the basic differences between the single-item and multi-item version is that items in the single-item version are presented at central fixation and no generation of eye-movements is needed. This can be taken to imply that in the single-item version the attentional and perceptual field is restricted to the currently presented item, whereas for the multi-item version it is expanded in order to process the upcoming target items, consistent with eye-movement studies of parafoveal processing in other multi-item naming tasks (Henry et al., 2018; Kuperman et al., 2016; Pan et al., 2013). This expansion opens up the possibility of interference among stimuli. Indeed, there is evidence to suggest that nearby stimuli interfere with the current response not only in the incongruent but also in the neutral condition. Specifically, in a practice study, MacLeod (1998) found that in the incongruent condition the integration problem (i.e., color and word are integrated) can be resolved rather quickly but the distraction problem (i.e., when color and word are spatially separated) persists. This means that one of the main problems in the multi-item version of the Stroop task is that two competing responses are simultaneously present, causing distraction, in accordance with the earlier claim of MacLeod (1991) that competing stimuli in spatial proximity cause interference. This applies to both the neutral and the incongruent condition in the multi-item version of the task, where all stimuli are simultaneously visible.

In support of this idea, Glaser and Glaser (1982, 1989) argued that interference occurs even when modally pure stimuli (e.g., color-color) are used. Their findings supported the presence of interference in the neutral condition of the task as well. In a similar vein, observations from the Eriksen flanker task indicate that spatially adjacent distractors belonging to the target set cause interference (Eriksen & Eriksen, 1974). Considering the confluence of these circumstances in the neutral condition of the multi-item version of the Stroop task, in which modally pure stimuli that belong to the target set are simultaneously present, it seems reasonable to hypothesize that interference occurs and, consequently, control is required. In addition, this implies that in the incongruent condition the complexity is not limited to within-stimulus interference between color and word responses but is further increased by interference from nearby stimuli.

Arguably, successful performance in the multi-item version of the Stroop task requires "simultaneous analysis of foveal and parafoveal information, programming eye movements, and synchronizing speech output to word decoding" (Zoccolotti et al., 2013, p. 641). In particular, in the incongruent condition, participants are parafoveally exposed to the upcoming item and need to identify its automatic dimension (i.e., word) as irrelevant, while simultaneously processing the two dimensions of the current item, namely, the—integrated, controlled, and slower—color dimension in parallel with the—task-irrelevant but automatic—word dimension, which must be inhibited. This ultimately leads to at least three simultaneous active responses, two to be identified and filtered out, and one to be articulated. (For simplification, preview of items further down is not taken into account, although it is plausible; Rayner, 1998; Rayner et al., 2005). And all this while uttering the response to the preceding item (for eye-voice span in other multi-item naming tasks see Gordon & Hoedemaker, 2016; Huang, 2018; Pan et al., 2013). Taken together, it seems that reading and naming, identification of the upcoming responses, and response selection, planning, and articulation run in parallel in the multi-item Stroop task. The same should hold for the neutral condition with the exception of the (automatic) word dimension. Therefore, when compared to the single-item version, the multi-item version appears far more complicated.

1.2. Default behaviors in the single-item and multi-item Stroop task

In the absence of specific instructions, word reading as an automatic and habitual response is considered to be the "default" behavior in both Stroop task versions, that is, single-item and multi-item. The introduction of color naming as goal via task instructions requires the application of "default override", that is, a situation where task demands require suppression of a default behavior (here, word reading) in order to allow a more controlled process to guide performance (i.e., color naming). It is because of default override that control implementation and allocation is required in the context of the Stroop task (Shenhav et al., 2013).

However, for the multi-item Stroop task word reading is not the only default behavior that emerges. As discussed earlier, because of simultaneous presentation of multiple items, parallel processing by virtue of parafoveal preview is an additional emergent default behavior. This is the case in both conditions of the multi-item Stroop task, that is, incongruent and neutral. Due to our reading history, in particular our well-practiced skill in reading sentences (i.e., word sequences), parallel processing of multiple items in the absence of explicit instructions should manifest naturally (i.e., automatically). Practice has an effect on parallel processing making it the default strategy (Allport et al., 1972; Fischer & Plessow, 2015). Indeed, it seems that in the absence of specific instructions parallel processing is the preferred strategy, despite its performance costs and the increase of between-tasks interference (i.e., crosstalk; Lehle & Hübner, 2009).

These observations further highlight the difference in complexity between single-item and multi-item Stroop tasks and suggest that control processes may manifest themselves differently depending on task version.

1.3. Implications of multi-tasking

Recapping, it appears that fast, successful color naming in the multi-item Stroop task depends on the execution of multiple concurrent processes (i.e., reading and naming, identification of the upcoming responses, and response selection, planning) involving current and nearby items due to parallel processing as the default behavior. These are by no means unitary or simple processes. For example, it is well established that reading and color naming require multiple steps (i.e., subtasks; for a definition of "task" see Koch et al., 2018, and Monsell, 1996), that is, perception, conceptual identification (for color naming), lemma retrieval, and word-form encoding (Roelofs, 2003). Thus, because of the presence of adjacent items, the multi-item Stroop task involves different, overlapping task sets (i.e., preparation and organization of cognitive processes for goal-oriented behavior to emerge; Kiesel et al., 2010; Monsell, 1996). Therefore, carrying out the multi-item Stroop task requires *time sharing* among concurrent tasks and subtasks, ultimately resulting in additional *task interference* (Wickens, 2002) and crosstalk (Fischer & Plessow, 2015). Time sharing is intra-modal: identification of responses is visual for both word and color. The same holds for responding (i.e., usually vocal in the multi-item task), which involves both reading, as an automatic but task-irrelevant response (evident in errors), and color naming, as a controlled but task-relevant response (i.e., task conflict; Hershman & Henik, 2019).

In addition, semantic relatedness between items in spatial proximity to the target (i.e., nearby items) may act to increase demands and could be seen as an additional source of information conflict (Hershman et al., 2020; Hershman & Henik, 2019; Kalanthroff et al., 2013; Levin & Tzelgov, 2014), that is, conflict arising from contradictory meaning-related information between the current item and nearby items. The semantic dimension further elevates between-task interference because response identification for the concurrent tasks overlap (Fischer & Plessow, 2015). For example, Hirst and Kalmar (1987) found that in a dichotic listening task semantically similar competing messages hindered target detection, especially if presentation was simultaneous. This competition between target and simultaneous semantically related response should be even stronger in the multi-item Stroop task. This is because the non-target response belongs to the task set and consequently should not be filtered out but, rather, remain active as the upcoming response. This is true for both conditions, namely, incongruent and neutral.

If this conceptualization is on the right track, then we can refer to the multi-item Stroop task as a "multi-task", meaning that—although it is presented as a single task—its successful execution entails simultaneous and parallel activation of more than one task sets, thereby satisfying the conditions for multi-tasking (i.e., execution of concurrent tasks demanding the simultaneous maintenance of two or more task sets; Koch et al., 2018; Meyer & Kieras, 1997; Monsell, 1996). Indeed, time sharing (or else temporal overlap between task sets), as in the case of the multi-item Stroop task, is a defining feature of a multi-task context (Fischer & Plessow, 2015; Koch et al., 2018; Wickens, 2002). This approach is further supported by Samuels and Flor (1997), who discussed how and why a superficially single, simple task such as text reading requires multi-task operations.

What is of importance here is that, if the multi-item Stroop task is indeed a multi-task in this sense, then performance costs should be expected, as reported in the multi-tasking literature, where it is well known that combining cognitive tasks for parallel execution results in severe performance decrements (Fischer & Plessow, 2015), indicating capacity constraints. This is in accordance with EVC theory, where "in principle, it is possible to specify more than one identity-intensity pairing, and thereby more than one task. However, in practice there are strict capacity constraints on control" (Shenhav et al., 2013, p. 220).[1] Thus, an open question remains: In the presence of different kinds of control demands (Schuch et al., 2019), can the control system find a balance between them for control implementation and allocation?

1.4. Present study

In our study we were interested in how conflict monitoring and EVC theory can be applied to conditions that require parallel execution of multiple cognitive tasks. To investigate this, we capitalize on the contrast between the two different versions of the Stroop task.

Hypothesis 1. Single-item and multi-item Stroop task require the same amount of control.

If the same amount of control is required for the single-item and multi-item Stroop tasks, then conflict monitoring and EVC theory predict control adaptations during the incongruent condition of the Stroop task, because control is adjusted by reference to the identity and the intensity of the control signal, taking payoffs and cost of control into account. No performance changes should be observed in the course of the neutral condition, in which control demands are minimal or even absent, depending on stimulus selection and/or readability of the material (e.g., colored letter strings such as 'XXX' or semantically unrelated colored words such as 'CAT'; Augustinova et al., 2018; Kalanthroff et al., 2013; Levin & Tzelgov, 2014; see also Botvinick et al., 2001, Fig. 1).

---

[1] Note that Shenhav et al. (2013) do not elaborate more on this point, arguing that simple control-demanding tasks are the most common circumstance (for a different view see Schuch et al., 2019), therefore no clear-cut hypotheses regarding the nature of the proposed constraints can be made.

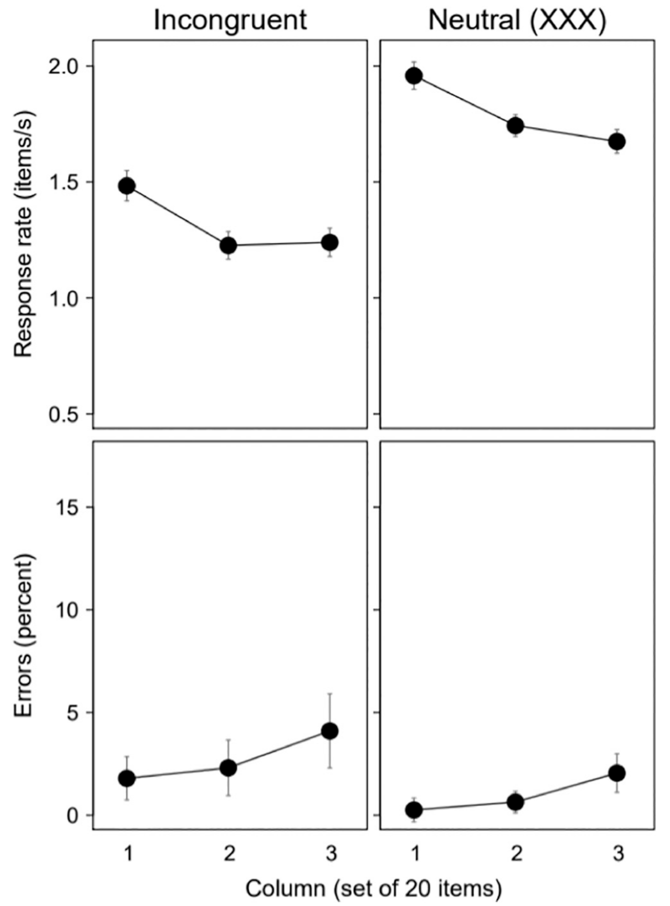

Fig. 1. Results of Experiment 1. Response rate (items per second) and accuracy (percent errors) in each column and condition. Error bars show within-participant 95% confidence intervals.

Hypothesis 2. The multi-item Stroop task is a multi-task.

A conceptualization of the multi-item Stroop task as a multi-task implies that multiple control demanding tasks *run in parallel*. In this case, two alternative predictions can be made for the incongruent condition.

First, the pattern of results should be the same as in any single-item task. This will be expected if the identity and intensity of the control signal is gradually and adaptively adjusted, leading to stable performance. Successful adjustment can be expected for switching both between control demanding tasks (i.e., identity) and between tasks of different difficulty (intensity; Shenhav et al., 2013).

If, however, capacity constraints exist, as proposed by Shenhav et al. (2013) and multi-tasking theorists (e.g., Fischer & Plessow, 2015; Kahneman, 1973; Neigel et al., 2019; Wickens, 2002, 2008), they could lead to performance decrements (i.e., breakdowns; Wickens, 2002). This is primarily relevant for the more complex incongruent condition, but is not limited to that. Indeed, if the so-called "neutral" condition also requires exertion of intensive control, as we propose, then a pattern similar to the incongruent condition should be evident, namely, decrement as the task progresses. However, this performance decrement should be modest and more gradual, due to the absence of the irrelevant word response which raises the level of task complexity in the incongruent condition.

Notably, these two contrasting predictions concern only the multi-item version of the task. That is, we predict substantial qualitative differences in within-task performance patterns and control manifestations between the single-item and multi-item version of the Stroop task. In the following three experiments, we directly tested these predictions by adopting a methodology of within-task performance evaluation.

2. Experiment 1

2.1. Method

2.1.1. Participants

The sample consisted of 41 adults (32 women) 21–38 years old ($M = 29.0$, $SD = 4.2$), mainly undergraduate and graduate students. All were native speakers of Greek.[2]

2.1.2. Material and apparatus

The Greek words for red (κόκκινο/kocino/), green (πράσινο/prasino/), and yellow (κίτρινο/citrino/) were used, because they have the same number of letters and syllables and comparable written frequency (33, 34, and 9 per million, respectively, from the IPLR; Protopapas et al., 2012). The corresponding colors are familiar and easily distinguishable. Stimuli for the neutral color condition were made up of 7 repetitions of the letter X (no spaces) in red, green, and yellow color (RGB #FF0000, #00FF00, and #FFFF00, respectively). For the incongruent condition the Greek words for red, green, and yellow appeared in a non-matching color.

Each condition was presented on a 15.5″ laptop screen in a single-screen array of three columns of 20 stimuli, for a total of 60 stimuli per condition, displayed in 20-pt Arial font on a full-screen Miscrosoft Powerpoint with black background. Each item extended approximately 5 × 20 mm on the screen. The vertical distance between successive items (nearest edges) was 6 mm and the horizontal distance between columns was 40 mm. The distance between the last item of a column and the first item of the subsequent column was 135 mm. For the incongruent condition there were 20 repetitions of each word and 20 repetitions of each color with colors being counterbalanced over columns. For the neutral condition 20 repetitions of each color were used. Colors and color words were randomly ordered with the constraint that adjacent items were not the same, similarly to the frequently used Golden (1978; Golden & Freshwater, 2002), Comalli et al. (1962), and Comalli-Kaplan (Strauss et al., 2006) versions of the Stroop task using three colors in cards of 100 items.

2.1.3. Procedure

The neutral condition was administered first, followed by the incongruent condition. Maintenance of a fixed order of conditions is compatible with most of the commonly used tests in which the non-conflicting conditions is administered first (e.g., Victoria version, Golden version; Golden, 1978; Golden & Freshwater, 2002; Strauss et al., 2006). We asked the participants to name the color as quickly as possible and to try to avoid errors, moving from top to bottom in columns. Prior to the tasks the production of the intended colors and understanding of the instructions were verified using sample cards. There were no practice trials. Naming responses were recorded via a headset using SIL Speech Analyzer 3.0.1 (SIL International; Speech Analyzer (Version 3.0.1) [Computer Software], n.d.).

2.1.4. Data preparation and dependent measures

The accuracy and total naming duration for each of the three columns (i.e., 3 blocks of 20 items) in each condition was measured on the waveform offline using Praat (Boersma & Weenink, 2012). Mispronunciations, substitutions, and self-corrections were considered errors. Our choice to process the multi-item version by column was

---

[2] As no previous research was available on which to base standard deviation and effect size estimates, no power analysis was conducted prior to the research. Sample size for the current and subsequent experiments was based on common practice in the study of Stroop tasks (e.g., Dulaney & Rogers, 1994; Ellis & Dulaney, 1991; Salo et al., 2001).

dictated by the physical grouping of items into three columns (see Fig. S.24 in the Supplemental Material), which implies a disruption of parallel processing when shifting gaze between columns.[3] A similar approach has been adopted in past studies of the Stroop task (Klein et al., 1997) and naming tasks (Amtmann et al., 2007). It would not be possible to carry out analyses based on individual items because in the multi-item version there is no item-specific onset of processing, as successive items are typically processed in parallel.

Naming times for each column were inverted and multiplied by the number of items to produce a measure of "items per second". Specifically, naming times for each of the three columns were divided by 20 (the number of items in each column). Hence, results reported below refer to *response rate* (i.e., number of items named per second). This transformation aimed to better approximate a normal distribution, compared to the frequently used mean response time), resulting in an interpretable ratio scale of measurement with meaningful parametric indices of central tendency and dispersion (i.e., mean and standard deviation, respectively). Graphical and statistical analyses of normality for both the original and transformed times are listed in the Supplemental Material (pp. 33–58). In addition, all analyses reported below have also been conducted using the raw (untransformed) response times and have produced the same pattern of pairwise differences; these are also listed in the Supplemental Material (pp. 4–25) to facilitate comparisons with the Stroop task literature.

2.1.5. *Statistical analysis*

Data were analyzed using two-way repeated-measures analysis of variance (ANOVA) with column (first, second, third) and condition (incongruent, neutral) as within-subjects factors and response rate/errors as the dependent measures in R 3.5.1 (R Core Team). Post-hoc pairwise comparisons with Bonferroni adjustment were performed using functions lme of the nlme package (Pinheiro et al., 2014) and glht of the multcomp package (Hothorn et al., 2008).

2.2. *Results*

Two participants did not complete the task and were excluded, leaving 39 for analysis. Table 1 presents the descriptive statistics for response rate and accuracy in each column and condition and Fig. 1 plots the means of both measurements for each condition as a function of column. Error bars show within-participant 95% confidence intervals adjusted using Morrey's (2008) correction, calculated using function summarySEwithin of Cookbook for R (www.cookbook-r.com).

2.2.1. *Response rate (items/s)*

2.2.1.1. Analysis of variance. In 3 (column) × 2 (condition) repeated-measures analysis of variance (ANOVA) the interaction was not significant, $F(2, 76) = 1.60$, $\eta^2 = 0.004$, $p = .207$. There was a main effect of column, $F(2, 76) = 74.50$, $\eta^2 = 0.187$, $p < .001$, and condition, $F(1, 38) = 182.21$, $\eta^2 = 0.482$, $p < .001$.

Table 1
Response rate and accuracy per column in each condition of Experiment 1.

	Response rate (items/s)				Proportion of errors			
	Incongruent		Neutral		Incongruent		Neutral	
Column	M	SD	M	SD	M	SD	M	SD
1	1.48	0.22	1.95	0.27	0.01	0.03	<0.01	0.01
2	1.22	0.20	1.74	0.27	0.02	0.04	<0.01	0.01
3	1.23	0.23	1.67	0.27	0.04	0.05	0.02	0.02

2.2.1.2. Post-hoc pairwise comparisons. Table 2 presents post-hoc pairwise comparisons with Bonferroni adjustment for response rate and errors. For response rate post-hoc analysis indicated that, in the incongruent condition, the first column differed from the second and third column, but there was no statistically significant difference between second and third column. For the neutral condition, the first column differed from the second and third column; the second and third column also differed significantly.

2.2.2. *Accuracy*

2.2.2.1. Analysis of variance. In the analysis of errors, the interaction of column and condition was not significant, $F(2, 76) = 0.10$, $\eta^2 = 0.001$, $p = .902$. There was a main effect of column, $F(2, 76) = 7.78$, $\eta^2 = 0.061$, $p < .001$, and condition, $F(1, 38) = 17.07$, $\eta^2 = 0.061$, $p < .001$.

2.2.2.2. Post-hoc pairwise comparisons. Post-hoc analysis, as presented in Table 2, revealed no significant difference between columns in the incongruent condition. In the neutral condition, the first and second columns did not differ but there was a significant increase between first and third column and the second and third.

2.3. *Discussion*

In the incongruent condition performance was not stable. In contrast, performance decrement was observed as the task progressed, in line with our conceptualization of the Stroop task as a multi-task, leading to performance costs. In addition, performance decrement was observed in the neutral condition as well, indicating demands that are intensive enough to require control implementation. The performance decrement observed in the neutral condition, interpreted as need for control, paralleled the performance decrement in the incongruent condition, resulting in a lack of interaction between conditions. Under the standard approach this lack of interaction might be taken as indicating lack of change in interference and therefore no effect on control. However, such an interpretation is no longer appropriate once the need for control in the neutral condition is appreciated, leading to control failures within the course of the neutral condition itself. This suggests that the neutral condition is not in fact free from interference, as typically assumed, and therefore cannot serve as a baseline measure relative to which interference can be measured by subtraction.

In the conflict monitoring hypothesis and EVS theory, interference is estimated as a difference between conditions and is used as an index of control, on the assumption that performance in the neutral condition remains stable and any fluctuations of performance can be attributed to the incongruent condition. However, Experiment 1 suggests that the so-

Table 2
Post-hoc pairwise comparisons with Bonferroni adjustment for response rate and accuracy in each condition of Experiment 1.

Column comparisons	Response rate (items/s)			Proportion of errors		
	β	z	*p*	*β*	z	*p*
Incongruent						
2–1	**−0.25**	**−6.82**	**<0.001**	<0.01	0.50	>0.999
3–1	**−0.24**	**−6.47**	**<0.001**	0.02	2.26	0.070
3–2	0.01	0.35	> 0.999	0.01	1.76	0.234
Neutral						
2–1	**−0.21**	**−7.83**	**<0.001**	<0.01	0.94	>0.999
3–1	**−0.28**	**−10.33**	**<0.001**	**0.01**	**4.39**	**<0.001**
3–2	**−0.06**	**−2.49**	**0.037**	**0.01**	**3.45**	**0.001**

Note. Statistically significant comparisons are marked in bold.

---

[3] Evidence from eye-movement studies, including return sweeps, indicates that if the target lies out of the attentional field it cannot be lexically preprocessed during the prior fixation (Slattery & Parker, 2019).

called neutral condition is not in fact a stable condition, immune to control processes, as typically assumed. This issue and its implications for interference estimation are taken up in more detail in the *General Discussion*.

Our findings are consistent with the idea that the multi-item Stroop task is a multi-task in which more than one tasks are active in parallel challenging the cognitive control system and indicating capacity constraints. If this interpretation is correct, then performance decrements should only be observed in the multi-item version of the Stroop task but not in a comparable single-item version with the same number and composition of items, due to its low complexity level because of single-item presentation and interstimulus intervals. This hypothesis was tested in the next experiment.

An additional objection could be raised against our interpretation regarding the neutral condition; specifically, that interference was not caused by nearby stimuli competing for response selection but simply because of the similarity of the items (i.e., all XXXs). This seems reasonable in light of findings that response times increase if target and distractors share common features, such as shape and color (Salo et al., 2001; Treisman & Gelade, 1980; but cf. Salo et al., 2001, for an opposite effect, with physical similarity speeding up responses). Moreover, it has been proposed that the best neutral condition to be compared with the corresponding incongruent is a condition where the only difference between them is the meaning of the word (Augustinova & Ferrand, 2012; Mead et al., 2002). To address this potential criticism, in the next experiment we added two conditions to examine the impact of similarity and word presence. If similarity and word presence of the items can partially account for the gradual decline in performance in the neutral condition, this decline will be attenuated as similarity decreases but will be inflated by word presence.

3. Experiment 2

In Experiment 1 we found within-task performance decrements using the multi-item version of the Stroop task, which we attributed to task complexity and parallel execution of multiple control demanding tasks. If this explanation is correct then no comparable performance decrement should be observed with the single-item version. Therefore, in Experiment 2 we examined within-task performance in the Stroop task in both the multi-item and the single-item version, keeping all other features of the task fixed. Moreover, to address concerns regarding the interpretation of our findings in the neutral condition of Experiment 1, two more neutral conditions were implemented in Experiment 2, namely one with different letters and one with animal names, consistent with neutral stimuli sometimes adopted in the Stroop literature (Augustinova & Ferrand, 2012; Mead et al., 2002; Salo et al., 2001).

3.1. *Method*

3.1.1. *Participants*

The total sample consisted of 43 adults (31 women) 21–36 years old ($M = 22.8$, $SD = 4.7$), mainly undergraduate and graduate students. All were native speakers of Greek.

3.1.2. *Material and apparatus*

Again, the Greek words for red, green, and yellow were used. The incongruent (INC) and the neutral condition (XXX) were the same as in Experiment 1. Two conditions were added:

The XBL condition was constructed by seven repetitions of the Greek capital letters chi, beta, and lambda (Χ, Β, and Λ) in the three different colors (i.e., red, green, and yellow). These letters were chosen because they do not share the same starting letter with the target colors and result in a similar visual extent as the XXX condition (5 × 20 mm). The ANI (animal) condition was constructed by the Greek words for whale (φάλαινα /falena/), gorilla (γορίλας /ɣorilas/), and roe deer (ζαρκάδι, /zarkaði/). These were chosen because (a) their initial letter differs from that of the target color, (b) they have the same number of letters and syllables as target colors, and (c) are of comparable printed frequency (0.24, 0.38, 0.16 per million, respectively, from the IPLR; Protopapas et al., 2012).[4]

The single-item task consisted of 60 stimuli in 20-pt Arial font displayed in random order on a black background at the center of the screen using DMDX (Forster & Forster, 2003). Each stimulus remained on the screen for 2 s. The interim period between stimuli was 166.67 ms (i.e., 10 frames with a 60 Hz screen refresh rate). Each single-item condition lasted approximately 3 min. All conditions in the multi-item version of the task satisfied the constraints of Experiment 1 related to sequential presentation of items. All stimuli used in both tasks had the same dimensions (5 × 20 mm).

3.1.3. *Procedure*

The procedure was the same as in Experiment 1 except that the four conditions were presented in random order (because of the two additional neutral conditions, to minimize confounds due to practice effects; e.g., Carter et al., 2000; Tzelgov et al., 1992). Half of the participants carried out the single-item task first, and the other half second. Presentation of the stimuli was blocked by condition in both the single-item and the multi-item version. There were no practice trials. As in Experiment 1, sample cards were shown to the participants prior to the multi-item version of the task, to verify production of the intended colors and understanding of the instructions. Similarly, three representative stimuli from each condition were presented to the participants prior to the single-item version of the task.

3.1.4. *Data preparation and dependent measures*

For the multi-item version of the task, data preparation and extraction was the same as in Experiment 1. For the single-item version, naming times were processed offline with CheckVocal (Protopapas, 2007) to mark response times and errors. Response times (i.e., onset latency, of correct responses only) were inverted and multiplied by 1000 to produce a scale comparable to the multi-item version (i.e., items per second). Mean response rates (per participant) were calculated for the three 20-trial blocks, to be compared to the 20-item columns of the multi-item version. Times less than 250 ms or greater than 1600 ms were excluded from the analysis (0.07%).

3.1.5. *Statistical analysis*

We performed a three-way repeated measures analysis of variance (ANOVA) with task version (multi-item, single-item), block/column (first, second, third), and condition (INC, XXX, ANI, XBL) as within-subjects factors in R, followed by post-hoc pairwise comparisons as in Experiment 1. One participant was excluded due to uncorrected-to-normal vision, leaving data from 42 participants for analysis.

3.2. *Results*

Table 3 shows the descriptive statistics for response rate and Table 5 for accuracy per block/column in each condition and task. Fig. 2 plots the means for each block/column in each condition and task.

3.2.1. *Response rate (items/s)*

3.2.1.1. Analysis of variance. In the analysis of response rates, the triple interaction of task version × column/block × condition was significant, $F(6, 246) = 3.86$, $\eta^2 = 0.003$, $p = .001$. All two-way interactions were also significant.

---

[4] It was not possible to obtain animal names with printed frequency closer to that of the color words without sharing the same initial letter.

Table 3
Response rate per block/column in each condition and task of Experiment 2.

Block/column	INC		XXX		ANI		XBL	
	M	SD	M	SD	M	SD	M	SD
Single-item								
1	1.59	0.26	1.97	0.27	1.78	0.23	1.87	0.24
2	1.62	0.24	1.94	0.31	1.79	0.27	1.86	0.28
3	1.64	0.29	1.92	0.25	1.79	0.27	1.83	0.27
Multi-item								
1	1.43	0.30	1.83	0.33	1.78	0.34	1.75	0.30
2	1.19	0.25	1.61	0.31	1.47	0.29	1.55	0.29
3	1.22	0.29	1.53	0.26	1.39	0.25	1.49	0.27

Note. INC, incongruent condition; XXX, neutral condition with repetition of Xs; ANI, condition with animal names; XBL, condition with repetition of different letters.

3.2.1.2. Post-hoc pairwise comparisons. Table 4 presents post-hoc analysis of response rates for the single-item version and multi-item version of the task. For the single-item version there was no difference between 20-item blocks in any condition.

In the multi-item version of the task there was a difference between the first and second and between the first and third column in the INC condition; there was no significant difference between the second and third column. The same pattern was observed in the XBL condition. A slightly different pattern of results was observed in the XXX and ANI condition, in which a statistically significant difference emerged between all consecutive columns.

3.2.2. Accuracy

3.2.2.1. Analysis of variance. In the analysis of errors, the triple interaction of task version × block/column × condition was not significant, $F(6, 246) = 0.76$, $\eta^2 = 0.003$, $p = .597$. The two-way interaction of task × block/column was significant, $F(2, 82) = 5.76$, $\eta^2 = 0.009$, $p = .004$. The interactions of block/column by condition and task by condition were only marginally significant, $F(6,246) = 2.08$, $\eta^2 = 0.10$, $p = .055$ and $F(3,123) = 2.46$, $\eta^2 = 0.006$, $p = .065$, respectively.

3.2.2.2. Post-hoc pairwise comparisons. Table 6 presents post-hoc analysis for errors for the single-item and multi-item versions of the task. For the single-item version no difference between columns in any condition of the single-item version of the task was observed.

In the multi-item version, there was no difference between columns in the INC condition (all $p > .999$) but there were differences in the other conditions. Specifically, in the XXX condition there was no difference between first and second column, but there was a difference between first and third column and between second and third column. In the ANI condition only the difference between first and third column was significant; the second column did not differ from the first or from the third. Finally, in the XBL condition the first column differed significantly from the second and the third column but the second and third column did not differ.

3.3. Discussion

As in Experiment 1, a steep performance decline was observed in the incongruent condition of the multi-item version of the Stroop task, as anticipated by multi-tasking. This performance decline was not evident in the single-item version of the task. More importantly, performance decline was again observed in all neutral conditions of the multi-item version, as in Experiment 1. This pattern of results eliminates the possibility that properties of the materials chosen (e.g., shape similarity) were responsible for performance decline in the neutral condition. Instead, we may conclude in favor of multiple sources of interference in the multi-item version, necessitating control allocation in incongruent and neutral conditions alike.

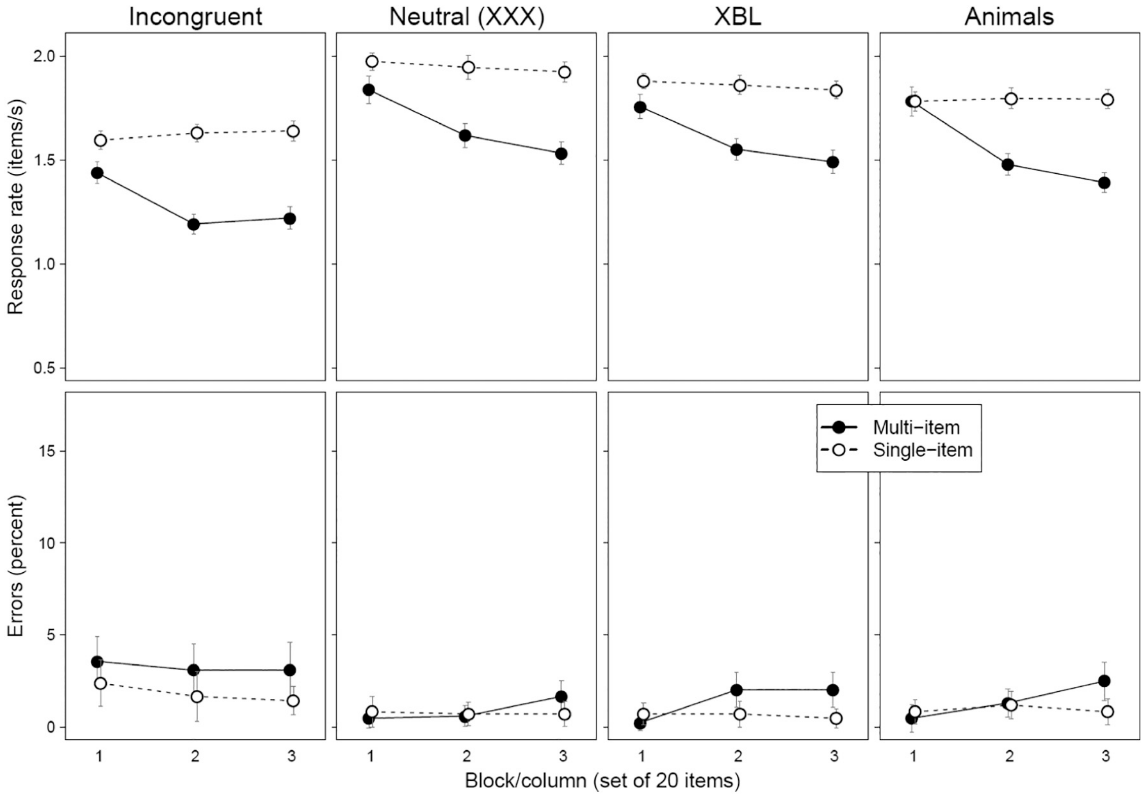

Fig. 2. Results of Experiment 2 (adults). Response rate (items per second) and accuracy (percentage of errors) in each block/column, condition, and task. Error bars show within-participant 95% confidence intervals.

Table 4
Post-hoc pairwise comparisons with Bonferroni adjustment for response rate in each condition and task of Experiment 2.

Block/column comparisons	Single-item			Multi-item		
	β	z	*p*	*β*	z	*p*
INC						
2–1	0.03	1.65	0.295	**−0.24**	**−8.66**	**<0.001**
3–1	0.04	2.16	0.090	**−0.21**	**−7.65**	**<0.001**
3–2	0.01	0.51	>0.999	0.02	1.00	0.938
XXX						
2–1	−0.02	−1.27	0.608	**−0.21**	**−8.01**	**<0.001**
3–1	−0.04	−2.20	0.082	**−0.30**	**−11.14**	**<0.001**
3–2	−0.02	−0.93	>0.999	**−0.08**	**−3.13**	**0.005**
ANI						
2–1	0.01	0.69	>0.999	**−0.30**	**−9.69**	**<0.001**
3–1	0.01	0.50	>0.999	**−0.39**	**−12.48**	**<0.001**
3–2	< 0.01	−0.19	>0.999	**−0.08**	**−2.78**	**0.015**
XBL						
2–1	−0.01	−0.82	>0.999	**−0.20**	**−7.30**	**<0.001**
3–1	−0.04	−1.86	0.187	**−0.26**	**−9.45**	**<0.001**
3–2	−0.02	−1.03	0.896	−0.06	−2.15	0.094

Note. INC, incongruent condition; XXX, neutral condition with repetition of Xs; ANI, condition with animal names; XBL, condition with repetition of different letters. Statistically significant comparisons are marked with bold.

Table 5
Accuracy per block/column in each condition and task of Experiment 2.

Block/column	INC		XXX		ANI		XBL	
	M	SD	M	SD	M	SD	M	SD
Single-item								
1	0.02	0.04	<0.01	0.02	<0.01	0.01	<0.01	0.01
2	0.01	0.04	<0.01	0.02	0.01	0.02	<0.01	0.02
3	0.01	0.02	<0.01	0.02	<0.01	0.02	<0.01	0.01
Multi-item								
1	0.03	0.04	<0.01	0.01	<0.01	0.02	<0.01	0.01
2	0.03	0.04	<0.01	0.01	0.01	0.02	0.02	0.03
3	0.03	0.05	0.01	0.02	0.02	0.03	0.02	0.03

Note. INC, incongruent condition; XXX, neutral condition with repetition of Xs; ANI, condition with animal names; XBL, condition with repetition of different letters.

Table 6
Post-hoc pairwise comparisons with Bonferroni adjustment for accuracy in each condition and task of Experiment 2.

Block/column comparisons	Single-item			Multi-item		
	β	z	*p*	*β*	z	*p*
INC						
2–1	<0.01	−0.86	>0.999	<0.01	−0.56	>0.999
3–1	<0.01	−1.15	0.739	<0.01	−0.56	>0.999
3–2	<0.01	−0.29	>0.999	<0.01	<0.01	>0.999
XXX						
2–1	<0.01	−0.27	>0.999	<0.01	0.27	>0.999
3–1	<0.01	−0.27	>0.999	**0.01**	**2.75**	**0.017**
3–2	<0.01	<0.01	>0.999	**0.01**	**2.47**	**0.039**
ANI						
2–1	<0.01	0.77	>0.999	<0.01	1.37	0.512
3–1	<0.01	<0.01	>0.999	**0.02**	**3.32**	**0.002**
3–2	<0.01	−0.77	>0.999	0.01	1.95	0.151
XBL						
2–1	<0.01	<0.01	>0.999	**0.01**	**3.33**	**0.002**
3–1	<0.01	−0.60	>0.999	**0.01**	**3.33**	**0.002**
3–2	<0.01	−0.60	>0.999	<0.01	<0.01	0.999

Note. INC, incongruent condition; XXX, neutral condition with repetition of Xs; ANI, condition with animal names; XBL, condition with repetition of different letters. Statistically significant comparisons are marked in bold.

Table 7
Response rate per block/column in each condition and task of Experiment 3.

Block/column	INC		XXX		ANI		XBL	
	M	SD	M	SD	M	SD	M	SD
Single-item								
1	1.29	0.20	1.58	0.19	1.42	0.20	1.41	0.20
2	1.26	0.20	1.49	0.23	1.36	0.20	1.38	0.22
3	1.24	0.19	1.49	0.24	1.32	0.19	1.36	0.21
Multi-item								
1	0.95	0.22	1.40	0.26	1.28	0.25	1.25	0.24
2	0.70	0.18	1.01	0.25	0.87	0.23	0.95	0.22
3	0.70	0.18	0.97	0.20	0.87	0.19	0.91	0.17

Note. INC, incongruent condition; XXX, neutral condition with repetition of Xs; ANI, condition with animal names; XBL, condition with repetition of different letters.

Table 8
Post-hoc pairwise comparisons with Bonferroni adjustment for response rate in each condition and task of Experiment 3.

Block/column comparisons	Single-item			Multi-item		
	β	z	*p*	*β*	z	*p*
INC						
2–1	−0.02	−1.14	0.752	**−0.24**	**−11.23**	**<0.001**
3–1	−0.05	−2.30	0.063	**−0.24**	**−11.31**	**<0.001**
3–2	−0.02	−1.15	0.744	< 0.01	−0.07	>0.999
XXX						
2–1	**−0.09**	**−4.35**	**<0.001**	**−0.39**	**−12.67**	**<0.001**
3–1	**−0.09**	**−4.36**	**<0.001**	**−0.43**	**−13.83**	**<0.001**
3–2	<0.01	0.01	>0.999	−0.03	−1.16	0.737
ANI						
2–1	**−0.06**	**−3.16**	**0.004**	**−0.40**	**−14.03**	**<0.001**
3–1	**−0.09**	**−5.22**	**<0.001**	**−0.40**	**−14.00**	**<0.001**
3–2	−0.03	−2.06	0.117	<0.01	0.03	>0.999
XBL						
2–1	−0.03	−1.68	0.274	**−0.29**	**−11.21**	**<0.001**
3–1	**−0.05**	**−2.65**	**0.023**	**−0.34**	**−12.90**	**<0.001**
3–2	−0.01	−0.96	>0.999	−0.04	−1.69	0.271

Note. INC, incongruent condition; XXX, neutral condition with repetition of Xs; ANI, condition with animal names; XBL, condition with repetition of different letters. Statistically significant comparisons are marked in bold.

Table 9
Accuracy per block/column in each condition and task of Experiment 3.

Block/column	INC		XXX		ANI		XBL	
	M	SD	M	SD	M	SD	M	SD
Single-item								
1	0.12	0.11	0.02	0.05	0.04	0.06	0.03	0.04
2	0.08	0.07	0.02	0.04	0.03	0.05	0.03	0.04
3	0.09	0.11	0.02	0.05	0.04	0.05	0.04	0.06
Multi-item								
1	0.10	0.08	0.02	0.03	0.02	0.03	0.02	0.04
2	0.13	0.09	0.07	0.05	0.06	0.05	0.05	0.05
3	0.14	0.10	0.06	0.06	0.04	0.05	0.05	0.06

Note. INC, incongruent condition; XXX, neutral condition with repetition of Xs; ANI, condition with animal names; XBL, condition with repetition of different letters.

Our findings are consistent with the idea that task complexity due to co-activation of multiple control demanding tasks is a critical factor leading to performance decline. If this explanation is correct then more spectacular effects on performance may be expected from the relatively immature control systems of children, because of their higher vulnerability to the demands of continuous effortful tasks. This possibility was tested in the final experiment.

Table 10
Post-hoc pairwise comparisons with Bonferroni adjustment for accuracy in each condition and task of Experiment 3.

Block/column comparisons	Single-item			Multi-item		
	β	z	*p*	*β*	z	*p*
INC						
2–1	**−0.03**	**−2.46**	**0.041**	**0.03**	**2.47**	**0.039**
3–1	−0.02	−1.82	0.203	**0.04**	**3.23**	**0.003**
3–2	<0.01	0.63	>0.999	<0.01	0.76	>0.999
XXX						
2–1	<0.01	<0.01	>0.999	**0.05**	**4.81**	**<0.001**
3–1	<0.01	−0.17	>0.999	**0.03**	**3.60**	**<0.001**
3–2	<0.01	−0.17	>0.999	−0.01	−1.20	0.687
ANI						
2–1	−0.01	−1.17	0.722	**0.04**	**4.31**	**<0.001**
3–1	<0.01	0.11	>0.999	0.02	2.15	0.092
3–2	0.01	1.29	0.590	−0.02	−2.15	0.092
XBL						
2–1	<0.01	0.26	>0.999	0.02	2.17	0.089
3–1	<0.01	0.91	>0.999	**0.03**	**2.97**	**0.008**
3–2	<0.01	0.65	>0.999	<0.01	0.80	>0.999

Note. INC, incongruent condition; XXX, neutral condition with repetition of Xs; ANI, condition with animal names; XBL, condition with repetition of different letters. Statistical significant comparisons are marked in bold.

4. Experiment 3

The aim of the current experiment was twofold. First, to replicate the findings of Experiment 2 at a different developmental stage, namely childhood, acknowledging that replicability contributes to the robustness of an effect. And second, to further expand the scope of our findings by examining the potential modulating role of control system maturity on the observed effects. Specifically, in studies of Stroop interference, a U-shaped development has been observed during childhood: Interference first emerges as children learn to read, reaches its highest levels in Grades 2–3, and then decreases through adulthood, ultimately leading to the well-documented differences in interference between adults and children (MacLeod, 1991).

This finding was attributed by Roelofs (2003) to strengthening of control structures from childhood in adulthood and it is in line with evidence of neurodevelopmental changes in cognitive control, such as poorer performance of children in response override, compared to adults (Bunge et al., 2002; Bunge & Wright, 2007; Schroeter et al., 2004). In addition, multi-tasking also develops from childhood to adulthood (Kliegel et al., 2008; Yang et al., 2017).

Taken together, control structure immaturity and multi-tasking skill level lead us to the following hypothesis: If the immaturity of the control system has a negative impact on control allocation and implementation it will contribute to performance costs by virtue of its higher vulnerability to capacity constraints. If that is the case, then the pattern observed in adults (i.e., gradual performance decrements in the neutral conditions) should manifest itself even more dramatically in children. This is in line with findings of vigilance decrements which show that novice participants invest greater effort compared to experienced ones (Shaw et al., 2013). The aim of the following experiment was to test these suggestions using the same materials and methods as in Experiment 2 and replicate its findings at a different developmental stage.

4.1. Method

4.1.1. Participants

The sample consisted of 45 children attending Grades 4–5. Participants' age range was 9.5 to 11.6 years, as estimated by elementary school starting age and the time of testing. Written informed consent was obtained from their parents for their participation.

4.1.2. Material and apparatus

The material and apparatus were the same as in Experiment 2.

4.1.3. Procedure

The procedure was the same as in Experiment 2.

4.1.4. Data preparation and dependent measures

As in Experiment 2. The proportion of response times excluded from the analysis (i.e., less than 250 ms or greater than 1600 ms) was 0.97%.

4.1.5. Statistical analysis

As in Experiment 2.

4.2. Results

Three participants were excluded (two due to software failure and one due to high nonresponse rate), leaving 42 for analysis. Table 7 presents the descriptive statistics for response rate and Table 9 for accuracy per block/column in each condition and task. Fig. 3 plots the means for each column in each condition in both tasks.

4.2.1. Response rate (items/s)

4.2.1.1. Analysis of variance. In 2 (task version) × 3 (block/column) × 4 (condition) repeated-measures analysis of variance (ANOVA) the three-way interaction was significant, $F(6, 246) = 2.32$, $\eta^2 = 0.003$, $p = .033$. Subsequent analyses showed that block/column × condition and block/column × task interacted significantly, $F(6, 246) = 7.81$, $\eta^2 = 0.009$, $p < .001$ and $F(2, 82) = 140.35$, $\eta^2 = 0.087$, $p < .001$, respectively. The same was the case for the interaction of task version × condition, $F(3, 123) = 9.18$, $\eta^2 = 0.013$, $p < .001$.

4.2.1.2. Post-hoc pairwise comparisons. Post-hoc analysis is presented in Table 8. For the single-item version and the INC condition, the difference between first and third column was only marginally significant (i.e., $p = .063$).The second column did not differ significantly from either the first or the third column. For the XXX condition, the first column differed significantly from the second and third column, but the second did not differ from the third. The same pattern of results was obtained for the ANI condition. For the XBL condition only first and third column differed significantly; the other comparisons did not reach significance.

For the multi-item version and the INC condition, the first column differed from the second and third but second and third column did not differ. The same pattern of results was also observed in all neutral conditions.

4.2.2. Accuracy

4.2.2.1. Analysis of variance. For errors, the triple interaction of task version × block/column × condition was not significant, $F(6, 246) = 1.69$, $\eta^2 = 0.005$, $p = .124$. Subsequent analyses revealed that the interaction of block/column × task was significant, $F(2, 82) = 20.80$, $\eta^2 = 0.024$, $p < .001$. The interaction of task × condition was only marginally significant $F(3, 123) = 2.67$, $\eta^2 = 0.007$, $p = .050$ and block/column × condition did not interact significantly, $F(6, 246) = 1.46$, $\eta^2 = 0.004$, $p = .192$.

4.2.2.2. Post-hoc pairwise comparisons. Table 10 presents post-hoc analysis for errors in both versions of the task. In the INC condition of the single-item version only first and second column differed significantly indicating a decrease. The other comparisons did not reach. For the neutral conditions, no comparison reached significance.

For the multi-item version and the INC condition, the first column differed significantly from the second and third, indicating an increase. Second and third column did not differ. The same results were obtained

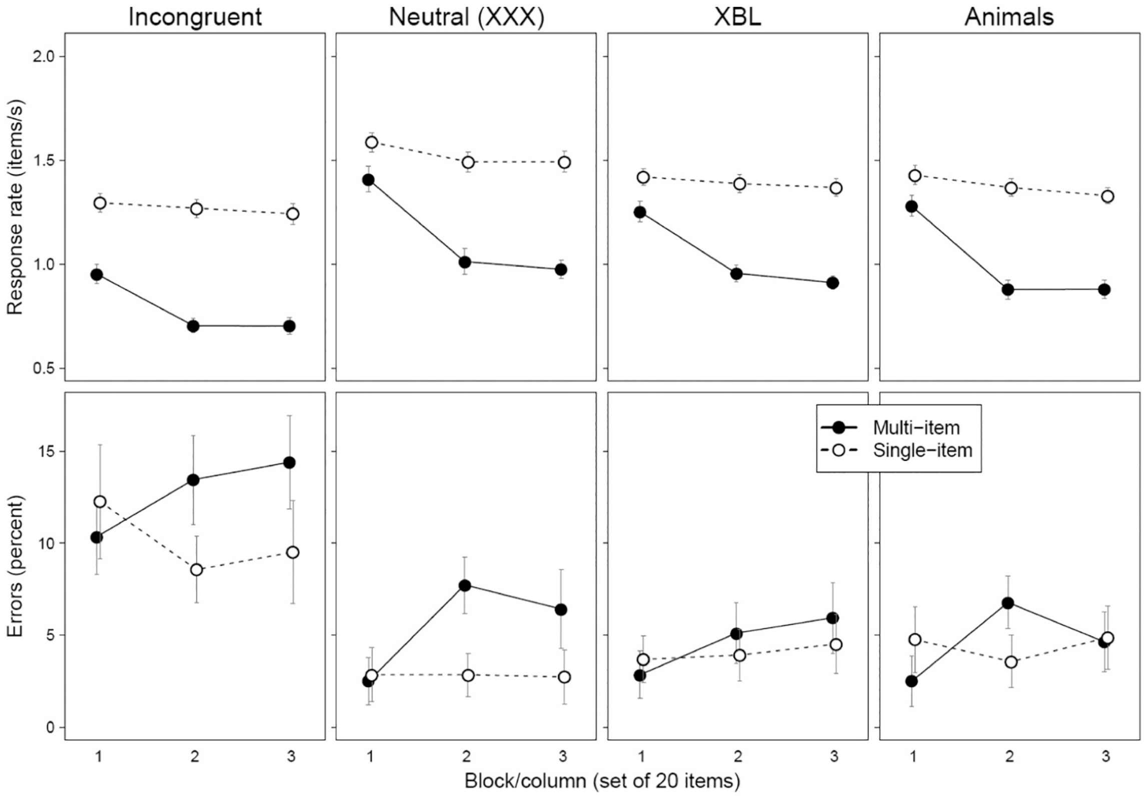

Fig. 3. Results of Experiment 3 (children). Response rate (items per second) and accuracy (percentage of errors) in each block/column, condition, and task. Error bars show within-participant 95% confidence intervals.

for the XXX condition. For the ANI condition, only the comparison between first and second column reached significance. For the XBL condition, only first and third column differed.

4.2.3. *Comparison of Experiments 2 and 3*

Finally, to compare within-task performance changes between adults and children, a mixed analysis of variance (ANOVA) for each condition and task was performed with group (adults, children) as a between-subjects factor and block/column (first, second, third) as within-subjects factor.

We first analyzed response rate. In the single-item version, the interaction of group × block/column was significant in the INC condition, $F(2, 164) = 5.09$, $\eta^2 = 0.007$, $p = .007$, and the ANI condition, $F(2, 164) = 7.89$, $\eta^2 = 0.009$, $p < .001$. The interaction did not reach significance in the XXX and the XBL condition, $F(2, 164) = 2.33$, $\eta^2 = 0.002$, $p = .100$ and $F(2, 164) = 0.10$, $\eta^2 < 0.001$, $p = .897$, respectively.

In the multi-item version, group × block/column did not interact significantly in the INC condition, $F(2, 164) = 0.51$, $\eta^2 < 0.001$, $p = .596$. In contrast, the interaction was significant in all neutral conditions [XXX: $F(2, 164) = 9.67$, $\eta^2 = 0.018$, $p < .001$; ANI: $F(2, 164) = 3.24$, $\eta^2 = 0.006$, $p = .041$; XBL: $F(2, 164) = 3.15$, $\eta^2 = 0.005$, $p = .045$].

We then analyzed errors. The interaction of group × block/column was not significant in any condition of the single-item version. In contrast, in the multi-item version the interaction reached significance in all conditions except XBL [INC: $F(2, 164) = 4.97$, $\eta^2 = 0.016$, $p = .007$; XXX: $F(2, 164) = 9.52$, $\eta^2 = 0.068$, $p < .001$; ANI: $F(2, 164) = 5.64$, $\eta^2 = 0.038$, $p = .004$; XBL: $F(2, 164) = 0.64$, $\eta^2 = 0.004$, $p = .528$].

4.3. *Discussion*

In this experiment, we replicated the findings of the previous two experiments in a different developmental stage, namely childhood. As was the case with adults, children showed a steep increase in response time as the incongruent condition of the task progressed, but only in the multi-item version, in accordance with the prediction related to the multi-tasking nature of the multi-item Stroop task.

More importantly, and in contrast to adults, in all control conditions the increase in color-naming time was steep rather than gradual. This finding supports the notion that the rate at which performance drops may depend on the strength of the control system, that is, its maturity.

5. General discussion

The conflict monitoring hypothesis and EVC theory suggest that in cognitively demanding tasks control is gradually adjusted, via conflict monitoring, leading to an overall stable performance, as previously observed in the single-item Stroop task. As the majority of studies examining control implementation have focused on single-task contexts (Schuch et al., 2019), we were interested in examining the applicability of the existing theories to tasks posing different demands on the cognitive control system. In this study we approached the multi-item version of the Stroop task as a multi-task, questioning the presumed equivalence of the different versions. We adopted within-task performance evaluation as an alternative methodology to shed more light on cognitive control processes, taking into account that control is dynamically adjusted with response delays and errors within a task acting as informants for control allocation and adjustments (Botvinick et al., 2001; Shenhav et al., 2013).

The results of our experiments demonstrated that within-task performance decrements were evident only in the multi-item version of the Stroop task. Notably, this was the case in both the incongruent and neutral condition and in both developmental stages, namely children and adults. Children were found to be more prone to performance decrements, consistent with the suggestion that the immaturity of the control system has a negative impact on control allocation and implementation. Furthermore, the fact that the interaction of group and block/column was significant for all neutral conditions of the multi-item task, and more importantly for the conditions involving a task-irrelevant word (i.e., incongruent, neutral with animal names) in the single-item version, suggests that for children a single source of interference is sufficient to cause performance decrements. Similar results have been previously reported by Klein et al. (1997), who examined the effect of test duration on Stroop task performance in different developmental stages by dividing a 100-item (i.e., 10 by 10 lines) version of the Stroop task in two parts (i.e., part I, lines 1–4 and part II, lines 5–10), and by Amtmann et al. (2007) in row-by-row analysis of multi-item naming tasks. Overall, the results justify the conceptualization of the multi-item Stroop task as a multi-task and suggest that the underlying control mechanisms may be qualitatively different between the two Stroop task versions.

In contrast to the multi-item Stroop task, performance in the single-item Stroop task was stable, consistent with the predictions of conflict monitoring and EVC theory. In particular, stable performance is consistent with the idea that in *simple* control demanding tasks control signal identity and intensity is gradually adjusted, in accordance with studies examining control in single-item tasks (e.g., Carter et al., 1995; Egner & Hirsch, 2005a; Egner & Hirsch, 2005b; Hinault et al., 2019; Kerns et al., 2004; Pardo et al., 1990; Schulte et al., 2019; Teubner-Rhodes et al., 2019). However, our findings indicate that, when similar control demanding tasks are active in parallel, as in the case of the multi-item Stroop task, specifying more than one identity-intensity pairing (i.e., more than one task) may exceed the processing capacity of the control system, as expected based on the multi-tasking literature.

5.1. Multi-item Stroop task as a multi-task: implications for cognitive control

EVC theory stresses the role of efficacy in cognitive control. Efficacy is “the likelihood that a goal will be reached with a given investment of control”; and “differences in efficacy (holding expected reward and difficulty constant) should itself be sufficient to drive changes in behavioral and neural signatures of control allocation” (Frömer et al., 2021, p. 2). In the context of the present study this implies that efficacy in the first column of the task can drive adjustment of control over the following columns. Performance in the first column was in fact quite high and should therefore have indicated that the level of control investment was appropriate for the goal to be reached. As all three columns were parts of one and the same task, there was no increase in difficulty to cause a change in expected efficacy and thereby lead to control intensity adaptations. Yet participants did slow down in subsequent columns while retaining their overall successful performance, evidenced in high accuracy. What could drive participants to change their behavior?

Cost-benefit analysis is relevant to consider in this situation. If the cost of maintaining the current level of control is too high relative to the expected payoff, this can lead participants to disengage from the task. This does not seem to be a very likely explanation, for the following reasons: First, feelings of self-efficacy must have provided participants with positive feedback during this time, given the rewarding nature of efficacy (e.g., Kool et al., 2010) and the interaction between efficacy and reward in control allocation (Frömer et al., 2021). Second, the presumed disengagement would seem to have a stable effect on response rate only, and not on error rate. This is most puzzling in the incongruent condition, where cost should be highest yet no difference in proportions of errors was observed between columns, in either experiment with adults (Figs. 1 and 2, bottom left). In fact error rates showed great variability in contrast to the systematic decline observed in response rate.

An alternative explanation would point to capacity limitations (as proposed by Wickens, 2002, 2008, and earlier by Kahneman, 1973) resulting “from the depletion of resources as time on task and task demands increase” (Neigel et al., 2019, p. 3). The resources referred to are not abstract in nature but related to task-specific information processing (Wickens, 2002) or attentional processes (Kahneman, 1973). If capacity limits are reached in the course of carrying out a task, cognitive overload results in within-task performance deterioration, as was evident in the multi-item version of the Stroop task. This approach is in line with an explanation of vigilance decrements based on automatic and controlled processes, in which controlled processes pose additional resource demands (Fisk & Scerbo, 1987; Fisk & Schneider, 1981). In our case, automatic and controlled processes are involved in both versions of the Stroop task. However, as noted in the introduction, the single-item version is simpler, because items appear individually, for a fixed duration, and—more importantly—separated by interstimulus intervals. These intervals provide an opportunity for rest and can therefore be crucial, as it is known from the vigilance literature and the control failure literature that rest provision benefits performance (Helton & Russell, 2015; Helton & Russell, 2017; Muraven & Baumeister, 2000). This approach is also in agreement with the multi-tasking literature showing that simultaneous processing implemented through short stimulus-onset-asynchronies (SOAs) has a detrimental effect on performance, with greater performance decrements associated with increased temporal overlap (Fischer & Plessow, 2015). Finally, the possibility of cognitive overload is consistent with neuroimaging studies of control. For example, Inzlicht and Gutsell (2007) found that Error Related Negativity (a pattern associated with ACC activity) was attenuated when participants were depleted, compared to non-depleted participants. Other studies have observed reductions in ACC activity during attentional lapses and after depletion (Persson et al., 2013; Weissman et al., 2006).

So, how can the observed pattern of within-task performance in the multi-item Stroop task be accounted for by conflict monitoring and EVC theory? As a starting point, we follow Shenhav et al. (2013) in assuming that when competing cognitive control demanding tasks are co-active, the control system estimates which task is most worth specifying. At the very beginning of the task participants focus on both current and nearby items, pursuing parallel processing not only for the purpose of speeding up responses, but primarily because of its default nature. As the task proceeds, conflict—as an internal index of task difficulty and control allocation—indicates via monitoring the need to re-specify the identity-intensity pairings of the control signal. As a result, current-item processing, and consequently serial processing (i.e., item-by-item processing; Fischer & Plessow, 2015), is prioritized against nearby-items (i.e., parallel) processing. Thus, we speculate that control allocation is accordingly adjusted by “lockout scheduling” (Meyer & Kieras, 1997, p. 20), meaning that subsequent items are excluded from processing until response planning of the current item has been completed (Roelofs, 2007).

Focusing primarily on current-item color naming has the inevitable implication of slowing down task execution, leading to the observed pattern of results, namely within-task performance decrement early on. This interpretation is in accordance with shielding of the most prioritized task (Berger et al., 2019; Fischer & Hommel, 2012; Fischer & Plessow, 2015) and shifting from more parallel to more serial processing (Miller et al., 2009) as an adaptive and flexible behavior. The unequal rates of performance drops between conditions in adults suggests that task shielding runs stronger in the incongruent condition and more moderate in the neutral condition. A similar distinction has been proposed in the context of sequential control adaptations in the Simon task under multi-task conditions in adults (Berger et al., 2019). Selection of an identity-intensity pairing to be prioritized, that is, the decision to prioritize the color naming task (overriding the word reading default) over the concurrent processing of nearby items, can be attributed to the explicit task instructions, consistent with a variety of studies showing the modulating role of instructions (Fischer & Hommel, 2012; Lehle et al., 2009; Lehle & Hübner, 2009).

In short, the within-task performance decrement observed in multi-item Stroop tasks is attributed to a capacity limitation, which cannot be accounted for by poor initial performance (via efficacy) and goes beyond the "normal" control requirements of single-item tasks (e.g., in the incongruent condition). The effect of this limitation on performance is quite dramatic. Whether this can be incorporated into EVC theory through cost estimation or might require a distinct mechanism is at present unclear. Our findings cannot conclusively determine the nature of the observed capacity constraints, that is, whether they are structural or functional. Still, they contribute to this central and still ongoing debate by highlighting the need for within-task performance evaluation of simple and more complicated cognitive control tasks in combination by adopting different methodologies, instructions, and material, especially when taken into account that "multitasking in itself constitutes a prime control dilemma" (Fischer & Plessow, 2015, p. 7).

Furthermore, the need for eye movements and oculomotor control should not be neglected. Eye movements are an inherent part of the multi-item task, making the visual input more complex and dynamic (Salo et al., 2001; Snell et al., 2018), as in other multiple-item displays (e.g., Henry et al., 2018; Kuperman et al., 2016; Pan et al., 2013). Additional research is needed to flesh out the origins of within-task performance decline in the multi-item version and the potential role of oculomotor control demands in it. In addition, the artificial nature of laboratory tasks should be also taken seriously into consideration in future paradigms examining control and the origin of multi-tasking costs within a task, in light of findings highlighting the need for naturalistic and ecological valid contexts to reveal their underlying cognitive processes in real-life situations (e.g., Blanco-Elorrieta & Pylkkänen, 2018).

5.2. Interference estimation in single-item and multi-item Stroop tasks

Beyond the implications for theories of control, our findings are also informative with regard to the estimation of interference, which is typically calculated as a difference in response time between the incongruent and the neutral condition. In theory, a "pure" baseline measure is necessary to demonstrate the impact of an automatic process (such as word reading) on a controlled process (such as color naming). MacLeod (1991) discussed in detail what kinds of stimuli could appropriately constitute a neutral condition and until today it is generally thought that a neutral condition that shares critical features with the incongruent condition (e.g., general slowing, word presence, response channel, etc.) constitutes a valid baseline for comparison (Hanauer & Brooks, 2005; Henik, 1996; MacLeod, 1991; Wright, 2017). However, Lindsay and Jacoby (1994) have argued that there is no such thing as "pure control stimuli" stating that "if the control items themselves cause some degree of interference with color-naming processes, then the interfering effect of word-reading processes on incongruent items will be systematically underestimated" (p. 219). They showed that degrading stimulus colors (i.e., bright vs. dull colors) affected only color-naming time performance, causing interference to appear greater in the bright-colors condition (129.4 ms) than in the dull-colors condition (91.5 ms), as a side-effect of subtraction.

The appropriateness of different versions of the Stroop task in specific contexts has come under scrutiny in recent years. For example, the multi-item version is widely used in clinical settings to assess deficits in attention and inhibition (Salo et al., 2001). However, the clinical literature is replete with inconsistencies, as some studies report abnormal interference in various patient groups whereas others fail to observe differences from the corresponding control groups, depending on the version used (Buchanan et al., 1994; Carter et al., 1992; Henik & Salo, 2004). The same pattern has emerged in comparisons of older to younger adults (Ludwig et al., 2010). Acknowledging this state of affairs, Salo et al. aimed to disentangle the different task components between the multi-item and the single-item version of the Stroop task. They observed that the neutral condition was faster in the multi-item version of the task, compared to the single-item version (i.e., a serial advantage; cf. Altani et al., 2019), contributing in their view to the differences observed. They therefore concluded that subtraction may not be the appropriate operation to derive an index of interference.

Although arising from a different starting point, our findings corroborate and extend these reservations, indicating that even identical neutral stimuli may act differently in various versions of the same task, or even at different times during the course of a single task, thereby suggesting that there is no such thing as a neutral stimulus irrespective of implementation. Our findings even suggest that arguing in favor of specific versions of a task may be misleading because different versions may pose substantially distinct task requirements. Different versions of the same task, despite sharing the same experimental stimuli, do not guarantee that the same underlying cognitive processes are recruited to accomplish them. Instead, different underlying processes may be involved, which may be unstable (Kindt et al., 1997), so that the nature of the phenomenon under investigation may vary as a function of administration and responding (Penner et al., 2012).

Moreover, our findings highlight additional factors such as the need for control and variations of performance not only within the incongruent but also within the neutral condition. These hitherto overlooked—but systematic—variations may hold the key to understanding the disparate and, often apparently contradictory, findings and conclusions regarding a variety of disorders and populations. This idea is also supported by the different patterns of results by adults and children that were observed in the neutral conditions of the multi-item version in our study.

5.3. Additional considerations

Some possible alternatives as plausible explanations of our findings should be addressed. First, withdrawal from the current task due to prolonged mental effort and intensive control demands have been proposed as interpretations for previously observed control failures (Kool & Botvinick, 2014). However, participants must be engaged in the control-demanding task over a prolonged period of time for a state of mental fatigue to arise, because duration is a precondition of fatigue. In fact the need for prolonged cognitive activity is what defines a state of mental fatigue (Boksem & Tops, 2008; Kato et al., 2009). Therefore, the mental fatigue approach is effectively countered by the fact that the multi-item version of the Stroop task is a very brief task and thus by definition unlikely to cause mental fatigue.

A potential alternative approach to our interpretation of between-column differences in the multi-item version might invoke the phenomenon of post-error slowing. Post-error slowing refers to the tendency of participants to be more "conservative" by slowing down after errors (Carter & van Veen, 2007). This might account for the observed increase in color naming times between columns to the extent that increased naming times were systematically accompanied by increased numbers of errors across conditions and populations. However, this was not the case. Instead, differences between columns in the number of errors varied widely among conditions and populations, in contrast to the robust increase of color-naming time.

Another possibility might be to attribute our findings to a speed-accuracy tradeoff, defined as "the complex relationship between an individual's willingness to respond slowly and make relatively fewer errors compared to their willingness to respond quickly and make relatively more errors" (Zimmerman, 2011, p. 2344). However, there was no systematic decrease in error rate associated with increased naming time, therefore our findings cannot be attributed to a speed-accuracy tradeoff. In fact the dissociation is highlighted by the fact that color naming times showed a decrease even when error rates were not significantly affected (e.g., incongruent conditions of Experiments 1 and 2).

Finally, sequential effects such as negative priming might also have affected the results. Negative priming refers to the increase of response times in the incongruent condition when the incorrect word-response of the preceding item matches the correct color-response of the currently named item (e.g., the word "green" printed in red followed by an item printed in green color; Dalrymple-Alford & Budayr, 1966; Neill, 1977). Negative priming items were indeed present in our multi-item version

(27 out of 60 items; 8, 10, 9 in the three columns, respectively). However, negative priming loses its strength as an alternative explanation due to two observations. First, although negative priming items were present in all three columns of the multi-item version, it was only during the second and third column where performance dropped. More importantly, performance decrements emerged also in the neutral condition, in which negative priming is absent. Hence, there is no reason to believe that the observed pattern of results should be attributed to sequential effects.

6. Conclusion

In conclusion, in this study we were interested in examining the course of control within a task in order to examine how current cognitive control theories could apply on superficially similar tasks that may in fact be posing substantially different demands on the cognitive control system. The starting point was the observation that most studies examining control implementation have focused on simple control-demanding tasks. Our findings suggest that single-item and multi-item Stroop tasks engage qualitatively different control processes. Our conceptualization of the multi-item Stroop task as a multi-task highlights the need for caution in comparing studies using different versions of the Stroop task and in measuring and interpreting interference. Our findings are consistent with capacity constraints under concurrent control demanding tasks. Future studies using different variants of instructions, methodologies, and tasks can shed more light on whether control is strategically adapted to meet task requirements or if the control system has a limit after which it can no longer be effective.

Ethics statement

An ethics approval was not required as per applicable institutional and national guidelines and regulations for both the adult and children studies. The children study was approved by the Institute of Educational Policy of the Greek Ministry of Education, Research and Religious Affairs as per applicable regulations and requirements. Written informed consent was obtained from all adult participants and from the parents/legal guardians of non-adult participants.

Data accessibility

OSF repository for this manuscript: https://osf.io/geuxn/

Declaration of competing interest

We have no conflicts of interest in the conduct and reporting of this research.

Acknowledgements

We thank Angeliki Altani for advice on data collection and audio processing, and Fotis Fotiadis, Argiro Vatakis, and Petros Roussos for help with recruiting participants.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2022.103583.

References

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology, 24*(2), 225–235. https://doi.org/10.1080/00335557243000102

Altani, A., Protopapas, A., Katopodi, K., & Georgiou. (2019). Tracking the serial advantage in the naming rate of multiple over isolated stimulus displays. *Reading and Writing, 33*, 349–375. https://doi.org/10.1007/s11145-019-09962-7

Amtmann, D., Abbott, R. D., & Berninger, V. W. (2007). Mixture growth models of RAN and RAS row by row: Insight into the reading system at work over time. *Reading and Writing: An Interdisciplinary Journal, 20*(8), 785–813. https://doi.org/10.1007/s11145-006-9041-y

Arán Filippetti, V., Richaud, M. C., Krumm, G., & Raimondi, W. (2021). Cognitive and socioeconomic predictors of Stroop performance in children and developmental patterns according to socioeconomic status and ADHD subtype. *Psychology & Neuroscience, 14*(2), 183–206. https://doi.org/10.1037/pne0000224

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., 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*, 5462. https://doi.org/10.1016/j.actpsy.2017.03.009

Berger, A., Fischer, R., & Dreisbach, G. (2019). It's more than just conflict: The functional role of congruency in the sequential control adaptation. *Acta Psychologica, 197*, 64–72. https://doi.org/10.1016/j.actpsy.2019.04.016

Bezdicek, O., Lukavsky, J., Stepankova, H., Nikolai, T., Axelrod, B. N., Michalec, J., Růžička, E., & Kopecek, M. (2015). The Prague Stroop test: Normative standards in older Czech adults and discriminative validity for mild cognitive impairment in Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology, 37*(8), 794–807. https://doi.org/10.1080/13803395.2015.1057106

Björngrim, S., van den Hurk, W., Betancort, M., Machado, A., & Lindau, M. (2019). Comparing traditional and digitized cognitive tests used in standard clinical evaluation - a study of the digital application minnemera. *Frontiers in Psychology, 10*, 2327. https://doi.org/10.3389/fpsyg.2019.02327

Blanco-Elorrieta, E., & Pylkkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences, 22*(12), 1117–1126. https://doi.org/10.1016/j.tics.2018.10.001

Boersma, P., & Weenink, D. (2012). *Praat: Doing phonetics by computer (Version 5.3.17) [Computer program]*. Retrieved from http://www.praat.org.

Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews, 59*(1), 125–139. https://doi.org/10.1016/j.brainresrev.2008.07.001

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8*, 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Bray, S. R., Martin Ginis, K. A., Hicks, A. L., & Woddgate, J. (2008). Effects of self-regulatory strength depletion on muscular performance and EMG activation. *Psychophysiology, 45*(2), 337–343. https://doi.org/10.1111/j.1469-8986.2007.00625.x

Buchanan, R. W., Strauss, M. E., Kirkpatrick, B., Holstein, C., Breier, A., & Carpenter, W. T., Jr. (1994). Neuropsychological impairments in deficit vs nondeficit forms of schizophrenia. *Archives of General Psychiatry, 51*(10), 804–811. https://doi.org/10.1001/archpsyc.1994.03950100052005

Bunge, S. A., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Current Opinion in Neurobiology, 17*(2), 243–250. https://doi.org/10.1016/j.conb.2007.02.005

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. (2002). Immature frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron, 33*, 301–311. https://doi.org/10.1016/S0896-6273(01)00583-9

Burger, L., Fay, S., Angel, L., Borella, E., Noiret, N., Plusquellec, P., & Taconnat, L. (2020). Benefit of practice of the Stroop test in young and older adults: Pattern of gain and impact of educational level. *Experimental Aging Research, 46*(1), 52–67. https://doi.org/10.1080/0361073X.2019.1693013

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 367–379. https://doi.org/10.3758/CABN.7.4.367

Carter, C. S., Robertson, L. C., & Nordhal, T. E. (1992). Abnormal processing of irrelevant information in chronic schizophrenia: Selective enhancement of stroop facilitation. *Psychiatry Research, 41*(2), 137–146. https://doi.org/10.1016/0165-1781(92)90105-C

Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An H2150 PET study of stroop task performance. *NeuroImage, 2*(4), 264–272. https://doi.org/10.1006/nimg.1995.1034

Carter, C. S., MacDonald, A. M., Botvinick, M., Ross, L. L., Stenger, A., Noll, D., & Cohen, J. D. (2000). Parsing executive processes: Strategic versus evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Sciences, 97*(4), 1944–1948. https://doi.org/10.1073/pnas.97.4.1944

Chen, K., Huang, L., Lin, B., Zhou, Y., Zhao, Q., & Guo, Q. (2019). The number of items on each stroop test card is unrelated to its sensitivity. *Neuropsychobiology, 77*, 38–44. https://doi.org/10.1159/000493553

Clemmensen, F. K., Hoffmann, K., Siersma, V., Sobol, N., Beyer, N., Andersen, B. B., Vogel, A., Lolk, A., Gottrup, H., Høgh, P., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S. (2020). The role of physical and cognitive function in performance of activities of daily living in patients with mild-to-moderate Alzheimer's disease - A cross-sectional study. *BMC Geriatrics, 20*(1), 513. https://doi.org/10.1186/s12877-020-01926-9

Comalli, P. E., Jr., Wapner, S., & Werner, H. (1962). Interference effects of stroop color-word test in childhood, adulthood, and aging. *The Journal of Genetic Psychology: Research and Theory on Human Development, 100*, 47–53. https://doi.org/10.1080/00221325.1962.10533572

Dalrymple-Alford, E. C., & Budayr, B. (1966). Examination of some aspects of the Stroop Color-Word Test. *Perceptual and Motor Skills, 23*(3, PT. 2), 1211–1214. https://doi.org/10.2466/pms.1966.23.3f.1211

DeWall, N. C., Baumeister, R. F., Stillman, T. F., & Gailliot, M. T. (2007). Violence restrained: Effects of self-regulation and its depletion on aggression. *Journal of Experimental Social Psychology, 43*(1), 62–76. https://doi.org/10.1016/j.jesp.2005.12.005

Di Filippo, G., & Zoccolotti, P. (2011). Separating global and specific factors in developmental dyslexia. *Child Neuropsychology, 18*(4), 391–396. https://doi.org/10.1080/09297049.2011.613809

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(2), 470–484. https://doi.org/10.1037/0278-7393.20.2.470

Egner, T., & Hirsch, J. (2005a). The neural correlates and functional integration of cognitive control in a Stroop task. *NeuroImage, 24*(2), 539–547. https://doi.org/10.1016/j.neuroimage.2004.09.007

Egner, T., & Hirsch, J. (2005b). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience, 8*, 1784–1790. https://doi.org/10.1038/nn1594

Ellis, N. R., & Dulaney, C. L. (1991). Further evidence for cognitive inertia of persons with mental retardation. *American Journal of Mental Retardation, 95*(6), 613–621.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149. https://doi.org/10.3758/BF03203267

Fischer, R., & Hommel, B. (2012). Deep thinking increases task-set shielding and reduces shifting flexibility in dual-task performance. *Cognition, 123*(2), 303–307. https://doi.org/10.1016/j.cognition.2011.11.015

Fischer, R., & Plessow, F. (2015). Efficient multitasking: Parallel versus serial processing of multiple tasks. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01366

Fisk, A. D., & Scerbo, M. W. (1987). Automatic and control processing approach to interpreting vigilance performance: A review and reevaluation. *Human Factors, 29* (6), 6533–6660. https://doi.org/10.1177/001872088702900605

Fisk, A. D., & Schneider, W. (1981). Control and automatic processing during tasks requiring sustained attention: A new approach to vigilance. *Human Factors, 23*(6), 737–750. https://doi.org/10.1177/001872088102300610

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*, 116–124. https://doi.org/10.3758/BF03195503

Frömer, R., Lin, H., Dean Wolf, C. K., Inzlicht, M., & Shenhav, A. (2021). Expectations of reward and efficacy guide cognitive control allocation. *NatureCommunications, 12*(1). https://doi.org/10.1038/s41467-021-21315-z

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8*(6), 875–894. https://doi.org/10.1037/0096-1523.8.6.875

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118*(1), 13–42. https://doi.org/10.1037/0096-3445.118.1.13

Golden, C. J. (1978). *Stroop Color and Word Test: A Manual for Clinical and Experimental Uses*. Chicago, IL: Stoelting Co.

Golden, C. J., & Freshwater, S. M. (2002). *Stroop Color and Word Test: Revised examiner's manual*. Wood Dale, IL: Stoelting Co.

Gordon, P. C., & Hoedemaker, R. S. (2016). Effective scheduling of looking and talking during rapid automatized naming. *Journal of Experimental Psychology: Human Perception and Performance, 42*(5), 742–760. https://doi.org/10.1037/xhp0000171

Hanauer, J. B., & Brooks, P. J. (2005). Contributions of response set and semantic relatedness to crossmodal Stroop-like picture-word interference in children and adults. *Journal of Experimental Child Psychology, 90*(1), 21–47. https://doi.org/10.1016/j.jecp.2004.08.002

Helton, W. S., & Russell, P. N. (2015). Rest is best: The role of rest and task interruptions on vigilance. *Cognition, 134*, 165–173. https://doi.org/10.1016/j.cognition.2014.10.001

Helton, W. S., & Russell, P. N. (2017). Rest is still best: The role of the qualitative and quantitative load of interruptions on vigilance. *Human Factors, 59*(1), 91–100. https://doi.org/10.1177/0018720816683509

Henik, A. (1996). Paying attention to Stroop effect? *Journal of the International Neuropsychological Society, 2*, 467–470. https://doi.org/10.1017/S1355617700001557

Henik, A., & Salo, R. (2004). Schizophrenia and the Stroop effect. *Behavioral and Cognitive Neuroscience Reviews, 3*(1), 42–59. https://doi.org/10.1177/1534582304263252

Henik, A., Bibi, U., Yanai, M., & Tzelgov, J. (1997). The stroop effect is largest during first trials. In *, 2. Abstracts of the Psychonomic Society* (p. 57).

Henry, R., van Dyke, J. A., & Kuperman, V. (2018). Oculomotor planning in RAN and reading: A strong test of the visual scanning hypothesis. *Reading and Writing, 31*, 1619–1643. https://doi.org/10.1007/s11145-018-9856-3

Hershman, R., & Henik, A. (2019). Dissociation between reaction time and pupil dilation in the Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(10), 1899–1909. https://doi.org/10.1037/xlm0000690

Hershman, R., Levin, Y., Tzelgov, J., & Henik, A. (2020). Neutral stimuli and pupillometric task conflict. *Psychological Research, 85*(3), 1084–1092. https://doi.org/10.1007/s00426-020-01311-6

Hinault, T., Larcher, K., Zazubovits, N., Gotman, J., & Dagher, A. (2019). Spatio-temporal patterns of cognitive control revealed with simultaneous electroencephalography and functional magnetic resonance imaging. *Human Brain Mapping, 40*(1), 80–97. https://doi.org/10.1002/hbm.24356

Hirst, W., & Kalmar, D. (1987). Characterizing attentional resources. *Journal of Experimental Psychology: General, 116*(1), 68–81. https://doi.org/10.1037//0096-3445.116.1.68

Hothorn, T., Bretz, F., & Westfall, P. (2008). Multcomp: Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363.

Huang, Y. T. (2018). Real-time coordination of visual and linguistic processes in novice readers. *Journal of Experimental Child Psychology, 173*, 388–396. https://doi.org/10.1016/j.jecp.2018.02.010

Inzlicht, M., & Gutsell, J. (2007). Running on empty: Neural signals for self-control failure. *Psychological Science, 18*(11), 933–937. https://doi.org/10.1111/j.1467-9280.2007.02004.x

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kalanthroff, E., Goldfarb, L., Usher, M., & Henik, A. (2013). Stop interfering: Stroop task conflict independence from informational conflict and interference. *Quarterly Journal of Experimental Psychology, 66*(7), 1356–1367. https://doi.org/10.1080/17470218.2012.741606

Kapoula, Z., Lê, T., Bonnet, A., Bourtoire, P., Demule, E., Fauvel, C., Quilicci, C., & Yang, Q. (2010). Poor stroop performances in 15-year-old dyslexic teenagers. *Experimental Brain Research, 203*(2), 419–425. https://doi.org/10.1007/s00221-010-2247-x

Kato, Y., Endo, H., & Kizuka, T. (2009). Mental fatigue and impaired response processes: Event-related brain potentials in a Go/NoGo task. *International Journal of Psychophysiology, 72*(2), 204–211. https://doi.org/10.1016/j.ijpsycho.2008.12.008

Kerns, J. G., Cohen, J. D., MacDonalrd, A. W., III, Cho, R., Stenger, A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*(5660), 1023–1026. https://doi.org/10.1126/science.1089910

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136*(5), 849–874. https://doi.org/10.1037/a0019842

Kindt, M., Bierman, D., & Brosschot, J. F. (1997). Cognitive bias in spider fear and control children: Assessment of emotional interference by a card format and a single trial format of the Stroop task. *Journal of Experimental Child Psychology, 66*(2), 163–179. https://doi.org/10.1006/jecp.1997.2376

Klein, M., Ponds, R. W., Houx, P. J., & Jolles, J. (1997). Effect of test duration on age-related differences in stroop interference. *Journal of Clinical and Experimental Neuropsychology, 19*(1), 77–82. https://doi.org/10.1080/01688639708403838

Kliegel, M., Mackinlay, R., & Jäger, T. (2008). Complex prospective memory: Development across the lifespan and the role of task interruption. *Developmental Psychology, 44*(2), 612–617. https://doi.org/10.1037/0012-1649.44.2.612

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin, 144*(6), 557–583. https://doi.org/10.1037/bul0000144

Kool, W., & Botvinick, M. M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General, 143*(1), 131–141. https://doi.org/10.1037/a0031048

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General, 139* (4), 665–682. https://doi.org/10.1037/a0020198

Kuperman, V., Van Dyke, J. A., & Henry, R. (2016). Eye movement control in RAN and reading. *Scientific Studies of Reading, 20*(2), 173–188. https://doi.org/10.1080/10888438.2015.1128435

Lehle, C., & Hübner, R. (2009). Strategic capacity sharing between two tasks: Evidence from tasks with the same and with different task sets. *Psychological Research, 73*(5), 707–726. https://doi.org/10.1007/s00426-008-0162-6

Lehle, C., Steinhauser, M., & Hübner, R. (2009). Serial or parallel processing in dual tasks: What is more effortful? *Psychophysiology, 46*(3), 502–509. https://doi.org/10.1111/j.1469-8986.2009.00806.x

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

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 219–234. https://doi.org/10.1037/0096-1523.20.2.219

Ludwig, C., Borella, E., Tettamanti, M., & de Ribaupierre, A. (2010). Adult age differences in the color Stroop test: A comparison between an item-by-item and a blocked version. *Archives of Gerontology and Geriatrics, 51*(2), 135–142. https://doi.org/10.1016/j.archger.2009.09.040

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

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition, 26*, 201–211. https://doi.org/10.3758/BF03201133

MacLeod, C. M. (2005). The Stroop task in cognitive research. In A. Wenzel, & D. S. Rubin (Eds.), *Cognitive methods and their application to clinical research* (pp. 17–40). Washington, DC: American Psychological Association.

Mead, L. A., Mayer, A. R., Bobholz, J. A., Woodley, S. J., Cunningham, J. M., Hammeke, T. A., & Rao, S. M. (2002). Neural basis of the Stroop interference task: Response competition or selective attention? *Journal of the International Neuropsychological Society, 8*(6), 735–742. https://doi.org/10.1017/s1355617702860015

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psychological Review, 104*(1), 3–65. https://doi.org/10.1037/0033-295x.104.1.3

Miller, J., Ulrich, R., & Rolke, B. (2009). On the optimality of serial and parallel processing in the psychological refractory period paradigm: Effects of the distribution of stimulus onset asynchronies. *Cognitive Psychology, 58*(3), 273–310. https://doi.org/10.1016/j.cogpsych.2006.08.003

Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved mysteries of the mind* (pp. 93–148). Hove, England: Erlbaum.

Morrey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*(2), 247–259. https://doi.org/10.1037/0033-2909.126.2.247

Musslick, S., Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2015). A computational model of control allocation based on the expected value of control. In *Paper presented at the 2nd multidisciplinary conference on reinforcement learning and decision making, Edmonton, Canada.* June.

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: Insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective, & Behavioral Neuroscience, 7*(1), 1–17. https://doi.org/10.3758/cabn.7.1.1

Neigel, A. R., Claypoole, V. L., Smith, S. L., Waldfogle, G. A., Fraulini, N. W., Hancock, G. M., Helton, W. S., & Szalma, J. L. (2019). Engaging the human operator: A review of the theoretical support for the vigilance decrement and a discussion of practical applications. *Theoretical Issues in Ergonomics Science, 40*(2), 239–258. https://doi.org/10.1080/1463922X.2019.1682712

Neill, W. T. (1977). Inhibitory and facilitatory processes in selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 3*(3), 444–450. https://doi.org/10.1037/0096-1523.3.3.444

Neshat-Doost, H. T., Dalgleish, T., & Golden, A.-M. J. (2008). Reduced specificity of emotional autobiographical memories following self-regulation depletion. *Emotion, 8*(5), 731–736. https://doi.org/10.1037/a0013507

Pan, J., Yan, M., Laubrock, J., Shu, H., & Kliegl, R. (2013). Eye–voice span during rapid automatized naming of digits and dice in Chinese normal and dyslexic children. *Developmental Science, 16*(6), 967–979. https://doi.org/10.1111/desc.12075

Pardo, J. V., Pardo, P., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the stroop attentional conflict paradigm. *Proceedings of the National Academy of Science, USA, 87*(1), 256–259. https://doi.org/10.1073/pnas.87.1.256

Penner, I., Kobel, M., Stocklin, M., Weber, P., Opwis, K., & Calabrese, P. (2012). The stroop task: Comparison between the original paradigm and computerized versions in children and adults. *The Clinical Neuropsychologist, 26*(7), 1142–1153. https://doi.org/10.1080/13854046.2012.713513

Periáñez, J. A., Lubrini, G., García-Gutiérrez, A., & Ríos-Lago, M. (2021). Construct validity of the Stroop color-word test: Influence of speed of visual search, verbal fluency, working memory, cognitive flexibility, and conflict monitoring. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists, 36*(1), 99–111. https://doi.org/10.1093/arclin/acaa034

Persson, J., Larsson, A., & Reuter-Lorenz, P. A. (2013). Imaging fatigue of interference control reveals the neural basis of executive resource depletion. *Journal of Cognitive Neuroscience, 25*(3), 338–351. https://doi.org/10.1162/jocn_a_00321

Pinheiro, J., Bates, D., Sarkar, D., Team, R. C., & DebRoy, S. (2014). *nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–117*. http://CRAN.R-project.org/package=nlme.

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39*, 859–862. https://doi.org/10.3758/BF03192979

Protopapas, A., Tzakosta, M., Chalamandaris, A., & Tsiakoulis, P. (2012). IPLR: An online resource for Greek word-level and sublexical information. *Language Resources & Evaluation, 46*, 449–459. https://doi.org/10.1007/s10579-010-9130-z

R Core Team. (2018). *R: A language and environment for statistical computing*. Retrieved from. Vienna, Austria: R Foundation for Statistical Computing https://www.Rproject.org/.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA division 40 members. *Archives of Clinical Neuropsychology, 20*(1), 33–65. https://doi.org/10.1016/j.acn.2004.02.005

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., Juhasz, B. J., & Pollatsek, A. (2005). Eye movements during reading. In M. J. Snowling, & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 79–97). Blackwell Publishing Ltd.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110*(1), 88–125. https://doi.org/10.1037/0033-295X.110.1.88

Roelofs, A. (2007). Attention and gaze control in picture naming, word reading, and word categorizing. *Journal of Memory and Language, 57*(2), 232–251. https://doi.org/10.1016/j.jml.2006.10.001

Salo, R., Henik, A., & Robertson, L. C. (2001). Interpreting Stroop interference: An analysis of differences between task versions. *Neuropsychology, 15*(4), 462–471. https://doi.org/10.1037/0894-4105.15.4.462

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 13*(2), 107–121. https://doi.org/10.1080/1057356970130202

Scarpina, F., & Tagini, S. (2017). The Stroop color and word test. *Frontiers in Psychology, 8*, 557. https://doi.org/10.3389/fpsyg.2017.00557

Schroeter, M. L., Zysset, S., Wahl, M., & von Cramon, D. Y. (2004). Prefrontal activation due to Stroop interference increases during development–An event-related fNIRS study. *NeuroImage, 23*(4), 1317–1325. https://doi.org/10.1016/j.neuroimage.2004.08.001

Schuch, S., Dignath, D., Steinhauser, M., & Janczyk, M. (2019). Monitoring and control in multitasking. *Psychonomic Bulletin & Review, 26*(1), 222–240. https://doi.org/10.3758/s13423-018-1512-z

Schulte, T., Hong, J. Y., Sullivan, E. V., et al. (2019). Effects of age, sex, and puberty on neural efficiency of cognitive and motor control in adolescents. *Brain Imaging and Behavior*. https://doi.org/10.1007/s11682-019-00075-x

Shaw, T. H., Satterfield, K., Ramirez, R., & Finomore, V. (2013). Using cerebral hemovelocity to measure workload during a spatialised auditory vigilance task in novice and experienced observers. *Ergonomics, 56*(8), 1251–1263. https://doi.org/10.1080/00140139.2013.809154

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron, 79*, 210–240. https://doi.org/10.1016/j.neuron.2013.07.007

Slattery, T. J., & Parker, A. J. (2019). Return sweeps in reading: Processing implications of undersweep-fixations. *Psychonomic Bulletin & Review, 26*(6), 1948–1957. https://doi.org/10.3758/s13423-019-01636-3

Snell, J., Mathôt, S., Mirault, J., & Grainger, J. (2018). Parallel graded attention in reading: A pupillometric study. *Scientific Reports, 8*(1), 3743. https://doi.org/10.1038/s41598-018-22138-7

Speech Analyzer (Version 3.0.1) [Computer Software], n.d.Speech analyzer (version 3.0.1) [computer software] n.d. Dallas: SIL International.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

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

Teubner-Rhodes, S., Bolger, D., & Novick, J. (2019). Conflict monitoring and detection in the bilingual brain. *Bilingualism: Language and Cognition, 22*(2), 228–252. https://doi.org/10.1017/S1366728917000670

Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling stroop effects by manipulating expectations for color words. *Memory & Cognition, 20*, 727–735. https://doi.org/10.3758/BF03202722

Wang, S., & Gathercole, S. E. (2015). Interference control in children with reading difficulties. *Child Neuropsychology, 21*(4), 418–431. https://doi.org/10.1080/09297049.2014.918594

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience, 9*, 971–978. https://doi.org/10.1038/nn1727

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*, 449–455. https://doi.org/10.1518/001872008X288394

Wright, B. C. (2017). What stroop tasks can tell us about selective attention from childhood to adulthood. *British Journal of Psychology, 108*(3), 583–607. https://doi.org/10.1111/bjop.12230

Yang, T., Xie, W., Chen, C., Altgassen, M., Wang, Y., Cheung, E. F., & Chan, R. C. (2017). The development of multitasking in children aged 7–12 years: Evidence from cross-sectional and longitudinal data. *Journal of Experimental Child Psychology, 161*, 63–80. https://doi.org/10.1016/j.jecp.2017.04.003

Zimmerman, M. E. (2011). Speed-accuracy tradeoff. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (p. (p. 2344).). New York, NY: Springer. https://doi.org/10.1007/978-0-387-79948-3_1247.

Zoccolotti, P., De Luca, M., Lami, L., Pizzoli, C., Pontillo, M., & Spinelli, D. (2013). Multiple stimulus presentation yields larger deficits in children with developmental dyslexia: A study with reading and RAN-type tasks. *Child Neuropsychology, 19*(6), 639–647. https://doi.org/10.1080/09297049.2012.718325