Interference and cognitive control dynamics in the course of serial naming tasks

Laoura Ziaka

Doctoral Thesis

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)

Department of Special Needs Education

Faculty of Educational Sciences

UNIVERSITY OF OSLO

Spring 2022

Series of dissertations submitted to the
Faculty of Educational Sciences, University of Oslo
No. 357

ISSN 1501-8962

Print production: Graphics Center, University of Oslo.

Acknowledgements

It has been a journey with a lot of adventures. I have tried to put inside of me every moment of this journey and thanks to many people, I still have the urge to sail.

First of all, I would like to thank my Teacher and Supervisor, Athanassios Protopapas. I am not thanking him because he took the time to teach me so many and different things; to ground me when needed; to support me when needed. I thank him for always seeing and treating me as an equal partner and, importantly, for believing in me more than I did. Thank you, Thanassi.

Victor Kuperman is another person of my heart. The first person, who trained me in eye tracking, and the person, who has always been available for me. I am grateful for his support.

Vasiliki Diamanti, Argyro Vatakis, Kathrine Høegh-Omdal, and William Brixius thank you so much for all the help.

Special thanks to the persons, with whom I started this journey some years ago. Fotis Fotiadis, Lora Dimopoulou, and Angeliki Altani are invaluable companions, making everything seem doable.

Sangwon Yoon and Oleg Zacharov were the first people I met, when changing residence. I was lucky. We shared our office for five years and they became my family. Thanks for everything, guys.

My dear participants, thank you.

Lida, Myrsini, Mairi, Vangelis, Valantis, Koulara, and Nikolai no words for you; just my love. I am so proud of you.

Laoura.

Summary

This PhD project is grounded on the notion of automaticity and its aim is two-fold. First, to examine the impact of nearby-items interference and cognitive control dynamics in the course of single-item and multi-item Stroop tasks. Second, to investigate if the format of the Stroop task influences the reported relationship between Stroop interference and reading. In this context, nearby-items interference is defined as the impairment in target recognition (e.g., color word) by simultaneous presentation of items being in spatial proximity to the target, with target and nearby items requiring the concurrent execution of multiple processes, while cognitive control is defined as performance regulation and adaptation under challenging and typically conflicting conditions.

Multiple displays might be challenging in light of evidence suggesting that nearby items may elevate the processing demands by interfering with processing of the target item. More specifically, behavioral and eye-tracking evidence suggests that when items appear in the context of other items, interference emerges, cognitive control is required, and performance gets impaired when compared to single-item displays. However, despite this evidence, the issue of item presentation and its impact on performance remains insufficiently investigated, with the majority of studies either not making a distinction between multi-item and single-item tasks or implementing primarily single-item tasks in various fields and domains. In addition, in the majority of studies performance is averaged throughout the entire task thereby failing to identify the evolution of cognitive processes and changes in processing within a task that might affect performance.

Furthermore, multiple displays are found to correlate more strongly with reading ability than single-item (i.e., discrete) tasks. More specifically, the literature related to

Rapid Automatized Naming (RAN) tasks has shown that simultaneous presentation of items captures individual differences in reading ability more efficiently than isolated item presentation, allowing to distinguish between readers of different levels. Although these findings indicate that naming in multiple displays and reading might share common cognitive processes, which are absent in processing of individual items, the nature of these processes remains unclear.

The present dissertation aims to address these gaps by comparing within-task performance variations in single-item and multi-item tasks and by examining the format dependence of the relationship between processing in multiple displays and reading ability.

The first article focused on the impact of nearby-items interference on cognitive control implementation in two different developmental stages (childhood and adulthood). Within-task performance between single-item and multi-item Stroop tasks was compared. The results of this study showed performance decrements in the multi-item task only, suggesting that the presence of nearby items due to interference challenges the cognitive control system.

Similarly, the second study examined how control is applied under conditions that go beyond single-item presentation by using behavioral measures (i.e., response time and errors) and eye-tracking measures (i.e., pupil size and gaze duration) in order to explore in more depth the impact of format differences in the Stroop task on cognitive control. For the multi-item version, the results showed that performance declined, the pupil constricted, and dwell time increased, while for the single-item version performance remained stable, the pupil constricted, and dwell time decreased. Overall, these findings suggest practice effects in the course of the single-item task, and narrowing of attention due to capacity constraints on control in the multi-item task, justifying the proposal for the existence of nearby-items interference in simultaneous presentation of items.

The third article focused on the examination of the relationship between Stroop interference and reading by examining the emergence of this relationship based on the format of the task used. This work scrutinized the assumption of a direct link between interference and the speed of inhibition of the task-irrelevant dimension (i.e., word) based on reading ability. Data from six experiments using single-item and multi-item Stroop tasks and their relationship to reading measures were examined in a meta-analytic framework. The results indicated that reading performance is primarily related to the multi-item version of the Stroop task and not to the single-item version. This suggests an indirect link between interference and reading, determined by the efficiency in temporally overlapping processing of nearby items, rather than the previously posited direct link.

Taken together, the studies constituting this dissertation support the existence of nearby-items interference in multiple displays and indicate capacity constraints on control under conditions requiring parallel processing. The observed within-task performance decrements, as well as the format-dependent relationship between Stroop interference and reading ability, suggest dynamic attentional shifts in the course of carrying out multi-item tasks, with implications for educational and clinical research.

List of Articles

Article I

Ziaka, L., & Protopapas, A. (2022). Conflict monitoring or multi-tasking? Tracking within-task performance in single-item and multi-item Stroop tasks. *Acta Psychologica, 226*, 103583. https://doi.org/10.1016/j.actpsy.2022.103583

Article II

Ziaka, L., & Protopapas, A. (under review). Cognitive control beyond single-item tasks: Insights from pupillometry, gaze, and behavioral measures. *Journal of Experimental Psychology: Human Perception and Performance*.

Article III

Ziaka, L., Skoteinou, D., Protopapas, A. (2022). Task format modulates the relationship between reading ability and Stroop interference. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(4), 275–288. https://doi.org/10.1037/xhp0000964

Table of Contents

List of Tables

PART I

EXTENDED ABSTRACT

1 Introduction

The present thesis describes the results of three studies focusing on the role of interference and cognitive control dynamics in the course of serial naming tasks. My interest in these aspects originated from two main observations. First, that multi-item (i.e., serial) tasks, that is, tasks in which items are presented simultaneously on a screen or sheet of paper, are better predictors of reading fluency when compared to single-item (i.e., discrete) tasks, that is, tasks, in which each item is presented individually to be named (e.g., Altani, Georgiou, et al., 2017; Bowers & Swanson, 1991; Chiappe et al., 2002; de Jong, 2011; Protopapas et al., 2013, 2018; Stanovich et al., 1983) with the origin of this relationship remaining unclear (Georgiou & Parrila, 2012; Parilla & Protopapas, 2017).

Second, that reading has been traditionally considered as one of the most automatic processes (Cattell, 1886) with automatic processes being considered fast, unintentional, obligatory, and effortless due to practice, whereas controlled processes are slow, intentional, effortful, and capturing attentional resources (Cohen et al., 1990; Logan, 1997; Moors & De Houwer, 2006). Moreover, it is generally accepted that automaticity is not an all-or-none phenomenon, but rather a continuum ranging from fully controlled to fully automatized processes (Moors & De Houwer, 2006); even two fully automatized processes running in parallel might interfere with each other if they occupy the same response modalities (Shiffrin & Schneider, 1977). Hence, the involvement of automaticity in processing of multi-item displays and reading tasks inevitably raises the issue of control implementation for their successful execution.

Taken together, the goal of the present dissertation was an examination of the impact of interference and cognitive control on processing of multi-item tasks and their relationship to reading by using different dependent measures and research designs.

1.1 Aims and research questions

The main aim of this PhD project is to understand in more depth the underlying cognitive mechanisms of processing single-item and multi-item displays in order to decode the critical aspects that allow to distinguish between readers of different levels (Parilla & Protopapas, 2017), giving emphasis on the comparison between tasks posing differential demands on the individuals and on decomposing them for examining within-task processing requirements and variations.

Three overarching research questions have guided the present dissertation. They should be seen in relation to each article presented in more detail in Chapter 5 and Table 1. The overarching questions are as follows:

1. *How can nearby-items interference affect performance when processing multiple displays?*
2. *How is control applied in tasks posing different processing demands?*
3. *What processes distinguish between readers of different levels?*

To explore these research questions, I conducted three separate studies which are presented in three different articles. Article I and article II implemented an experimental design in order to investigate the impact of nearby-items interference and, hence, control implementation on within-task performance. For Article I, behavioral measures were used, that is, response times and errors, by using within-task performance evaluation, while for Article II, I combined behavioral and eye-tracking measures (i.e., pupil size and gaze duration) using again within-task analysis in order to flesh out the origin of within-task performance variations. Finally, Article III is a correlational study, in which I analyzed previously collected data, in order to investigate if, how, and why different presentation of material might distinguish between readers of different levels.

1.2 Positioning of the study

In this dissertation, nearby-items interference is defined as the impairment in task performance by simultaneous presentation of items in spatial proximity to the target, with target and nearby items requiring the concurrent execution of multiple processes. Nearby-items interference becomes the focus of this project based on the fact that in commonly used multiple displays (e.g., reading) similar-in-nature items are presented in the context of other visually distinctive items requiring the same processes and sub-processes. Hence, the automaticity level of the currently processed item is in alignment with the automaticity level of adjacent items. If so, additional demands are posed on the individuals, requiring cognitive control implementation, that is, identifying the conflict and adapt accordingly (Botvinick et al., 2004; Muraven & Baumeister, 2000).

Although many studies have investigated interference and control implementation in conflicting tasks in order to investigate how and when control is adapted, the majority of the studies focus on single-item tasks (e.g., Stroop task, Eriksen flanker task, Simon task; Draheim et al., 2021), that is, tasks in which each item 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

(MacLeod, 2005). By doing so, tasks in which items are simultaneously present, as in the case of multi-item tasks, and presumably require different cognitive processes for their accomplishment, are systematically neglected. Furthermore, even in the case of the single-item tasks, performance is averaged throughout the course of the whole task (Rouder & Haaf, 2019), making it impossible to observe within-task performance variations, which could be suggestive of different cognitive manifestations based on task requirements.

Interference tasks are used in experimental and correlational research to serve different research goals. On the one hand, in experimental studies, interference tasks are employed in order to approach the cognitive processes required to reach a goal by manipulating experimental conditions. As experimental research is based on the assumption that cognitive processes are reflected in average responses, the experimental approach aims to keep between-subjects variability low in order for within-subject variance to emerge (Draheim et al., 2021; Hedge et al., 2018). Yet on the other hand, interference tasks are also used in correlational research in order to identify the relationship between different individual traits (within individuals) which would allow to detect individual differences in reference to these traits. Thus, in correlational studies the emergence of between-subjects variability is the desired outcome in order to identify potential individual differences (Draheim et al., 2019, 2021).

A typical example of the use of interference tasks in experimental and correlational research is the Stroop task. The task consists of two conditions, namely the incongruent and the neutral[1]. In the incongruent condition of the task color words are printed in a different color (e.g., the word "red" in green-colored letters) and the participants' task is to name the color and ignore the word, whereas in the control condition participants have to name the color of neutral stimuli (such as color patches or arrays of Xs). In the typical Stroop task the two conflicting dimensions (i.e., color and word) are integrated, that is, they spatially overlap. In addition, the Stroop task has two formats of administration, that is, the single-item (i.e., discrete) format and the multi-item (i.e., serial) format, which are used in different domains and for different purposes.

As in the case with the other interference paradigms, the Stroop task is used in experimental research to examine group differences in attentional control and is also used in correlational research to examine individual traits which may distinguish participants who vary in their ability to implement control (Draheim et al., 2021; MacLeod, 1991). Because word reading is the crucial dimension that results in interference due to its automatic nature, the Stroop task is further employed in the study of reading at both levels, namely experimental and correlational. For example, Protopapas et al. (2007) investigated the relationship between Stroop interference and reading by adopting a correlational approach. However, to further explore the origin of the Stroop interference-reading relationship, some years later an experimental approach was used examining the effect of different types of practice (i.e., color naming vs word reading) on Stroop interference (Protopapas et al., 2014). In other words, the Stroop task has attracted a lot of attention by researchers aiming to understand the cognitive mechanisms related to reading. Moreover, because in the

[1] Although a third condition exists, that is, the congruent condition, in which color word and color match (e.g., the word "red" in red-colored letters), this condition is not presented here because it is not relevant to the studies conducted in the present dissertation.

Stroop task automatic (i.e., word reading) and controlled (i.e., color naming) dimensions coexist, its examination has a lot to offer to our understanding of the transition from controlled to automatic processing and the factors that affect this transition.

This dissertation is grounded on the concept of automaticity. As described in the introduction, automatic processes are thought to be fast, unintentional, obligatory, and effortless, whereas controlled processes are slow, intentional, effortful, and capturing attentional resources (Cohen et al., 1990; Moors & De Houwer, 2006). Moreover, it is considered that a continuum of processing automaticity exists ranging from highly controlled to highly automatic processes, with the level of automaticity altering the attentional resources needed to accomplish a goal (Moors & De Houwer, 2006). In this context, skilled word reading, as a highly practiced process, is considered to be fast, unavoidable, and effortless; that is, an automatic process requiring fewer attentional resources. In contrast, other processes like color or object naming are considered more controlled processes due to the absence of extensive practice (MacLeod, 1991; Roelofs, 2003, 2006). This point is further highlighted by the relationship of reading and naming based on the level of automaticity of the material (alphanumeric vs non-alphanumeric; Roelofs, 2006) and the format used, that is, single-item or multi-item (Altani, Protopapas, & Georgiou, 2017; Protopapas et al., 2013, 2018). Hence, one assumption that can be made is that in tasks in which competing responses are simultaneously present, cognitive control is required to adjust processing. The degree of cognitive control that is required depends on the automaticity of the material used.

What, however, has not come under scrutiny yet in this context is the existence of nearby-items interference in multi-item tasks and, consequently, the need for control implementation, even in simple tasks without a prevalent competing response (e.g., the neutral condition of the Stroop task). However, there is evidence that items in near proximity to the target cause interference. More specifically, research comparing single-item and multi-item Stroop tasks suggests that in the multi-item format nearby items could act as distractors and increase the difficulty of the task (Ludwig et al., 2010). Moreover, it has been proposed that one of the basic differences between the single-item and multi-item format is that items in the single-item version are presented at central fixation and, consequently, the generation of eye-movements is not required, while in the multi-item format the perception of flanking items may influence and impair performance (Salo et al., 2001). Importantly, the influence of nearby items on target recognition seems to affect not only the incongruent condition, involving the spatially integrated and competing reading response, but also the neutral condition, which consists of non-conflicting items. The proposal that the presence of nearby items might affect performance in the neutral condition of the task is supported by studies examining interference in modified Stroop tasks. More specifically, in a practice study, MacLeod (1998) compared the typical single-item task, in which the two competing dimensions are integrated (integrated task), with a modified Stroop task, in which color and word were spatially separated (i.e., colored asterisks to be named appeared above the task-irrelevant word; separated task). The results showed that color naming training led to a steep decrease in interference in the integrated task, while in the separated task the decrease in interference was more gradual. Thus, the problem stemming from the simultaneous presence of the two conflicting dimensions seems to be more persistent than the integration problem (i.e., that the two conflicting dimensions spatially overlap; MacLeod, 1998). These findings indicate that one of the main problems in the multi-item version of the Stroop task is that competing items are simultaneously present. In fact, this is the case for both conditions of the task, that is, incongruent and neutral. Glaser and Glaser (1982, 1989) showed that interference occurs even when modally pure stimuli (e.g., color-color or word-word) are used, further supporting the idea of competing responses in both conditions of the Stroop task. Taken together, these findings indicate that both conditions of the Stroop task pose cognitive control demands on the participants due to the presence of nearby items and, consequently, control implementation is required. The view of cognitive control implementation in multiple displays could be further expanded to "simple" naming tasks, as the serial RAN and reading tasks, when taking into account two aspects: a) the similarity between Stroop and RAN tasks and b) the simultaneous presentation of items in both serial RAN and reading, presumably causing between-item competition.

To conclude, studies examining automatic and controlled processes and their relationship to reading and cognitive control in group (i.e., experimental research) and individual (i.e., correlational research) level use the Stroop task, because of its double nature, that is, being a rapid naming task and an interference task. Based on these observations, the Stroop task becomes and ideal candidate to examine interference and control dynamics within its course and apply any conclusions drawn in tasks adopting different material, that is, sequential word processing and text reading.

1.3 Outline of the extended abstract

This thesis is divided into six chapters aiming to contextualize, exemplify, and discuss the current project. I have already positioned the study in the field of cognitive and experimental psychology. Chapter 2 positions the research even further with the presentation of state of the research field. In Chapter 3, I outline the theoretical framework related to cognitive control. Chapter 4 presents the research design and the methodology used. In addition, ethical considerations are discussed. Chapter 5 summarizes the three articles reported in this thesis, including their main findings and interpretation. In Chapter 6, the overall study findings, contributions, possible implications and limitations are discussed.

2 State of the field

This chapter offers an overview of international research related to a) processing requirements in single-item and multi-item displays, b) evidence supporting the existence of nearby-items interference and the consequent need for control implementation in multi-item tasks. The studies I have reviewed stem from two different research fields in psychology, that is, from experimental and educational psychology. By combining different types of evidence I strove for a detailed, comprehensive, and multi-factorial approach of the phenomena under study. The studies reported in this chapter, published in international peer-reviewed journals, include both seminal works that may today be considered “classics” as well as more contemporary views found in the aforementioned fields.

2.1 Literature review

A literature review was undertaken to summarize the research conducted and identify possible gaps (Grant & Booth, 2009). The focus of the literature review is on research methods and research outcomes. The coverage was exhaustive with selective citations (Cooper, 1988) based on: 1) studies’ innovative nature at the time of publishing, 2) establishment of the replicability/robustness of the effects under consideration, and 3) contradictory results raising concerns.

For the literature review I focused on published research articles as they appear in electronic databases (Google Scholar, Pubmed, PsycARTICLES). In the first step, search terms were used. Specifically, a combination of “serial”, “discrete”, “multi-item”, “single-item”, “Stroop task”, “RAN task”, “interference”, “cognitive load”, “cognitive overload” “cognitive control”, “reading”, “eye-movements”, “perceptual span”, and “parafoveal processing” was used. In the second step, references from the retrieved articles were located and screened to identify their relevance. There were no restrictions in reference to date of publication, but due to practical reasons, the focus was on English-language articles. Inclusion criteria were the relevance to the topic under study. In the final selected sample of studies the criteria of innovative nature, replicability, and contradictory results were taken into account.

2.2 Processing in single-item and multi-item displays

Serial Stroop and Rapid Automatized Naming (RAN) tasks are generally thought to resemble each other (Norton & Wolf, 2012) with some studies treating the control condition of the Stroop task as a color RAN task (e.g., Helland & Morken, 2016; Stringer et al., 2004). It has also been found that RAN and Stroop tasks load on the same factor as they are similar naming tasks (Di Filippo & Zoccolotti, 2011). For these reasons, I will first present the literature related to RAN tasks and move on to briefly describe single-item and multi-item Stroop tasks and review the literature for evidence proposing nearby-items interference in multi-item Stroop tasks.

2.2.1 Discrete and serial RAN

Rapid Automatized Naming (RAN) tasks require rapid naming of repeated familiar items, and may be composed of letters and numbers (i.e., alphanumeric RAN) or colors and objects (i.e., non-alphanumeric RAN; for a review see Wolf & Bowers, 1999, and Norton & Wolf, 2012). There are two kinds of rapid naming tasks. The first is the discrete (i.e., single-item) format, in which each item appears individually on the screen to be named. The second is the serial (i.e., multi-item) format, in which all items appear simultaneously on the screen or a sheet of paper to be named. For the serial format usually 50 items of familiar objects are arranged in five rows of ten items and presented pseudorandomly, with no successive identical items (Norton & Wolf, 2012).

What is of importance in the context of the present dissertation is that reading fluency correlates more strongly with serial naming tasks than with discrete naming tasks (Bowers & Swanson, 1991; Chiappe et al., 2002; de Jong, 2011; Protopapas et al., 2013, 2018; Stanovich et al., 1983). Researchers have argued that the critical aspect for the emergence of this relationship is the multiple presentation of items (Protopapas et al., 2013; 2018). The higher correlations between reading and serial naming in comparison to reading and discrete naming is summarized under the term "serial superiority effect" and refers to the view that reading and serial naming share cognitive processes due to the serial nature of both tasks, which is not encountered in the discrete format, where items appear in isolation (Altani, Protopapas, & Georgiou, 2017).

Additional format-specific associations are also present in the RAN–reading relationship. Specifically, it has been found that discrete word reading correlates more strongly to discrete naming than to serial naming and, accordingly, serial (i.e., word list) reading to serial naming than to discrete naming (Altani et al., 2020). Importantly, the differential association is modulated by reading experience, that is, the "position" of reading in the continuum of automatic processing. Specifically, de Jong et al. (2011) found that serial RAN in beginning Dutch readers correlated with both discrete and serial reading tasks. This was explained based on the assumption that, if sight-word reading is insufficiently developed, individual words are read serially (i.e., letter by letter or syllable by syllable), leading to the stronger relationship of serial RAN with discrete and serial reading tasks. These findings were replicated by later studies in Greek (Protopapas et al., 2013; 2018). Sight word reading as a concept amounts to automatic word recognition, meaning that sight words are recognized and read without the intention of the reader (Ehri, 2005). Therefore, findings supporting the impact of sight word reading on the format-specific associations reported in the RAN–reading relationship inevitably highlight the crucial role of automaticity of the material and the transition from controlled to automatic processing in the emergence of the described associations.

The impact of the level of automaticity of the material on RAN performance is also highlighted by findings showing that alphanumeric RAN tasks (i.e., consisting of letters and digits) are better predictors of reading when compared to non-alphanumeric equivalents (i.e., consisting of colors, objects, or recently dice; e.g., Altani et al., 2020) with most of the studies indicating that alphanumeric items are more efficiently processed than non-alphanumeric due to extensive practice (Roelofs, 2006). The role of automaticity of the material in performance is further supported by the production model of Roelofs (2006) suggesting that less familiar material, as in the case of dice, require additional processing steps and stages—defined in the model as conceptual identification and lemma retrieval—while alphanumeric items due their automatic nature bypass these steps by directly mapping to word-form encoding.

One additional difference between serial and discrete RAN tasks is that the former is performed faster than the latter. Specifically, it has been found that typically-developing readers produce shorter naming times in multiple displays when compared to the isolated presentation of items (Zoccolotti et al., 2013), a finding that has been termed “serial advantage” (Altani, Georgiou, et al., 2017; Altani, et al., 2020; Zoccolotti et al., 2013).

Recently, a hypothesis has been postulated to explain the strong relationship between serial RAN and reading ability and account for both the serial superiority effect and the serial advantage, namely, the "cascaded processing" hypothesis. According to this hypothesis, the common process between reading and serial RAN is the temporally overlapped sequential processing of successive items (termed "cascading"), that is, one item is processed while the previous one is articulated and the next one is viewed (and, possibly, an item further down is previewed; Protopapas et al., 2013, 2018).

Although the "cascaded processing" hypothesis is still under testing with more studies needed in order to be conclusive, findings related to eye movements in multiple displays support this interpretation. Specifically, the importance of parafoveal processing, that is, "the extraction of partial-word information from the parafovea" (Rayner, 1998, p. 382), is already known not only for reading, but also for RAN tasks (Henry et al., 2018; Yan et al., 2013; Kuperman et al., 2016). Similarly, studies have shown that, when naming aloud, the eyes are ahead of the voice (i.e., "eye-voice span"; Gordon & Hoedemaker, 2016; Huang, 2018; Pan et al., 2013; Silva et al., 2016). This line of evidence highlights temporal processing overlap between successive items, which is a crucial component of the "cascaded processing" hypothesis, and, thus, indirectly support it.

Finally, studies examining parafoveal processing and eye-voice span in readers with learning disabilities (here, dyslexia) indicate the role of automaticity in the parallel processing of items. More specifically, fluent reading is defined as the ability to process a passage fast, accurately, and with proper expression (Hudson et al., 2009; Kuhn et al., 2010; Wolf & Katzir-Cohen, 2001). Thus, poor readers and readers with dyslexia, whose reading is not fluent, are characterized by definition by a lack of reading automaticity (Protopapas et al., 2007) and the same seems to hold for automaticity in naming (Norton & Wolf, 2012). If so, the comparison between typical readers and readers with dyslexia in RAN tasks related to parafoveal processing and eye-voice span might reveal the modulating role of automaticity in sequential processing. Indeed, the studies of Yan et al. (2013) and Pan et al. (2013), in which parafoveal processing and eye-voice span were compared between participants with and without dyslexia, showed that parafoveal processing and eye-voice span were impaired in participants with dyslexia when compared to the corresponding control groups, suggesting that the level of automaticity has an effect on sequential processing.

2.2.2 Single-item and multi-item Stroop task

The Stroop task is one of the best known and most studied tasks in the domain of experimental and cognitive psychology. Despite being a very simple task, its contribution to our understanding of the concepts of attention, automaticity, and cognitive control is highly important.

In the initial implementation of the task by Stroop (1935), words and colors were combined in such a way that the two dimensions were incompatible (e.g., the word "green" printed in red ink). Participants were asked to name the color as quickly as possible, and to correct errors. This experimental condition was compared with a control condition consisting of colored rectangles (e.g., a rectangle appearing in red ink). The items of the two conditions appeared in two separate cards of 100 stimuli (10 rows and 10 columns) and participants named the items in a left-to-right fashion. Five colors were used (red, blue, green, brown, purple) and total naming time was measured for each card. What was found was that it took much longer to respond to the incongruent stimuli than to the neutral stimuli. The difference in response times between the two conditions (incongruent minus neutral) is now termed “Stroop interference”. For the interpretation of his findings, Stroop followed the practice account and the notion of automaticity of Cattell (1886). Since the first introduction of the Stroop task, many variants have been developed, for example the counting Stroop task, the emotional Stroop task, and the picture-word task (MacLeod, 2005).

More importantly, an alternative to the original card version of the color word Stroop task has been introduced, namely the single-item version of the task, which is now broadly used due to specific advantages. One of these advantages is that the presentation of isolated items allows researchers to mix congruent, incongruent, and neutral items as also to investigate different conditions of presentation (MacLeod, 1991, 2005; Salo et al., 2001). Additional advantages are related to treatment of errors. More specifically, errors in the single-item version can be identified and removed from analyses, a problem difficult to address in the multi-item version, which can result to inflated response times in this version (MacLeod, 2005; Salo et al., 2001). In addition, errors can be counted in order to obtain a measure of error proportion, which subsequently can be analyzed along with mean response times in order to reveal speed-accuracy trade-offs (MacLeod, 2005). Although it is considered that trade-offs between speed and accuracy are rare in the Stroop task, as indicated by positive correlations between the two measures in single-item tasks (MacLeod, 2005), studies related to individual differences research suggest that speed-accuracy trade-offs are a common confound in interference tasks—including the Stroop task—which hinders the distinction between individuals based on specific factors of interest (i.e., between-subjects variability; Draheim et al., 2021; Hedge et al., 2018). Finally, it should be noted that producing a measure of error proportion is also feasible for the multi-item version of the task, despite the fact that error recording in this version varies between studies (Salo et al., 2001).

Although single-item and multi-item Stroop tasks are used interchangeably in the literature, some studies have shown that the choice of administration and presentation might change the observed interference effect, which appears to be larger in the multi-item version when compared to its single-item counterpart (MacLeod, 2005; Salo et al., 2001). In addition, the estimated interference in the multi-item version can distinguish between different populations (for example, people with schizophrenia, older adults, head injured patients; Buchanan et al., 1994; Henik & Salo, 2004; Ludwig et al., 2010; Vakil et al., 1995), whereas less consistent results are obtained from the single-item version (for a review see Salo et al., 2001; Henik & Salo, 2004).

It has long been proposed that the single-item version fails to retain the initial "Stroop-like" (i.e., highly conflicting) nature of the multi-item version (Penner et al., 2012), with some authors arguing that the conflict is further raised in the multi-item task because of the presence of nearby items, which act as distractors, increasing overall the difficulty of the task (Ludwig et al., 2010). Boucart et al. (1999) reached a similar conclusion in reference to the abnormal interference observed in people with schizophrenia when compared to controls, evident primarily in the multi-item version. The authors proposed that the disproportional slowing of people with schizophrenia in the multi-item version stems from the presence of distracting items in near proximity to the target.

Finally, as in the case of RAN tasks, the relationship of the Stroop task with reading ability is well established. Everatt et al. (1997) found that children with dyslexia exhibit more interference than age-matched controls, a finding confirmed across languages and ages (Di Filippo & Zoccolotti, 2011; Faccioli et al., 2008; Helland & Asbjørnsen, 2000;

Kapoula et al. 2010; Kelly et al., 1989; Protopapas et al., 2007; Reiter et al., 2005; Wang & Gathercole, 2015). The group differences in interference have been attributed to general factors such as inhibition impairments (Reiter et al., 2005; van der Schoot et al., 2000; but cf. Wang & Gathercole 2015) or to general slowness of readers with dyslexia (Di Filippo & Zoccolotti, 2011). In contrast, Protopapas et al. (2007) proposed a direct link between interference and the speed of inhibition of the task-irrelevant dimension (i.e., word), that is, reading ability affects the time course of suppression of the task-irrelevant response. Specifically, it has been proposed that skilled readers are faster in reading the word, resulting in rapid activation of the task-irrelevant response, which therefore is suppressed faster when compared to poor readers. The faster suppression of the word amounts to less interference, compared to poor readers, whose slow word reading delays suppression of the task-irrelevant dimension and leads to slower task-relevant response, that is, greater interference. This proposal seems to be further supported by findings showing that reading practice reduces Stroop interference, at least in children (for whom reading is less automatic; Protopapas et al., 2014).

Taken together, the studies reviewed in sections 2.2.1 and 2.2.2 indicate the complexity of the multi-item Stroop task and the difference from its single-item counterpart. In addition, they show the similarity of Stroop and RAN tasks as rapid naming tasks and their relationship to reading.

2.3 Evidence for nearby-items interference in multiple displays

As previously described, nearby-items interference is defined here as the impairment in performance due to the simultaneous presentation of items in spatial proximity to the target, with target and nearby items requiring the concurrent execution of multiple processes. The evidence supporting the notion of nearby-items interference in multiple displays comes from behavioral and eye-tracking research, and is reviewed next.

2.3.1 Behavioral evidence

The first piece of evidence indicating nearby-items interference due to simultaneous presentation of items comes from flanker tasks and response competition paradigms. The letter flanker task is a simple task originally used to examine selective attention processes (Eriksen & Eriksen, 1974). In this 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), compatible (e.g., S flanked by H, where H is associated with the same key response), or incompatible (e.g., S flanked by L, where L is associated with the opposite key response). The main and robust finding is that response-incompatible flankers produce increased response times compared to all other conditions. This finding is well-documented in different studies and for different material and response mappings (i.e., arrows and letters, same-different tasks; e.g., Eriksen, 1995; Paap & Sawi, 2019; Ridderinkhof et al., 2021; Salthouse, 2010). Crucially, the distance between the target and flankers seems to additionally modulate this effect with evidence showing that, although flankers cause interference if present, the interference effect reaches its maximum when distance is decreased, with Eriksen & Eriksen (1974) arguing that if a flanker appears at a distance less than 1 degree of visual angle, it is inevitably processed and has to be inhibited.

Furthermore, the flanker task inspired modifications to the Stroop task. Specifically, Gatti and Egeth (1978) modified the Stroop task in such a way that combined the Stroop task with the flanker task. Participants were asked to name the color of a patch in the presence of an incompatible word response appearing above and below the patch at a distance of 1, 3, or 5 degrees of visual angle. The results showed that conflicting distractors interfere at all distances, although their impact seems reduced as a function of distance. Interference from spatially distinct, task-irrelevant stimuli has been replicated in other Stroop studies (Kahneman & Chajczyk, 1983; Kahneman & Henik, 1981; MacLeod, 1998) justifying the claim of Eriksen (1995) that the "response compatibility effect is much more general than had been assumed from experimentation with the Stroop task. The effect is not limited to conflict inherent within the stimulus itself; it can come from other objects in the visual field in proximity to the attended object" (p. 105).

Although these findings suggest that nearby items might interfere with the identification of the target item, some studies implementing the lexical decision task and using reading-like material (i.e., words) suggest the opposite, that is, facilitation, thereby challenging the notion of nearby-items interference in multiple displays. More specifically, it has been found that, when words are flanked by orthographically related words (e.g., rock *rock* rock), word recognition is faster, compared to orthographically unrelated flankers (e.g., step *rock* step; Snell & Grainger, 2018; Snell et al., 2021). The difference in recognition time has been interpreted as evidence for orthographic facilitation due to parafoveal processing.

However, the methodology adopted in these studies leaves room for alternative interpretations and even suggests interference rather than facilitation. More specifically, the conclusions drawn by the authors are based on the comparison between an orthographically related condition (e.g., rock *rock* rock) and an orthographically unrelated condition (e.g., step *rock* step; Snell & Grainger, 2018; Snell et al., 2021). There is no comparison between conditions with and without flankers. However, it is already known from the Eriksen flanker task that flankers that are identical to the target do not cause interference and lead to faster responses compared to dissimilar flanker conditions (Eriksen & Eriksen, 1974; Eriksen, 1995). In fact, the interference effect in the Eriksen flanker task is measured by subtracting the identical or compatible condition from the incompatible one (Draheim et al., 2021). Thus, it does not seem surprising that the orthographically related conditions used in the previous studies proved to be faster than the unrelated condition. This interpretation of the reported findings as a methodological artifact is further supported by the study of Snell and Grainger (2018), in which a non-flanking condition was included. In that study, orthographically related targets were found to cause facilitation when presented *alone* and *only* at the right side of the target (e.g., *rock* rock). In contrast, no facilitation was found, in comparison to the non-flanking condition, when related flankers were presented only at the left side or at both sides (i.e., left and right of the target). Importantly, the leftward repetition flanker —despite being identical to the target— produced longer responses times when compared to the non-flanking condition. In other words, it caused interference.

Altogether it seems that orthographically unrelated flankers cause interference irrespective of the condition. Snell and Grainger (2018) acknowledged the existence of the interference effect in multiple-words presentation referring to parafoveal-on-foveal effects. Typically parafoveal-on-foveal effects refer to the possibility that the word to the right of fixation (i.e., the parafoveal word) may influence processing of the currently fixated word (i.e., the foveal word; Rayner et al., 2005). In the Snell and

Grainger study parafoveal-on-foveal effects also refer to the word on the left of the fixation, with the authors suggesting that "orthographic parafoveal-on-foveal effects from word $n - 1$ on word n should be observable in sentence reading, even if influences from word $n + 1$ should nonetheless be stronger" (p. 1518), indicating interference and processing costs.

2.3.2 Eye-tracking evidence

Processing costs in the presence of other words have recently come into attention in sentence reading and in the investigation of the preview benefit. The preview benefit refers to the observation that readers start processing a word before fixating it, leading to its faster recognition when compared to conditions in which the preview of the word is made unavailable or invalid (Rayner, 1998, 2009; Rayner et al., 2005). The preview benefit is related to parafoveal processing and the perceptual span (Rayner et al., 2005). It is typically investigated with the use of the boundary technique, in which an invisible boundary is placed before the parafoveal word and display changes happen (i.e., in invalid conditions) or do not happen (i.e., in valid conditions) when participants' gaze crosses this boundary. The typical manipulation for investigating the preview benefit is to compare a condition with valid (i.e., same) preview to conditions with invalid previews. The existence of the preview benefit and its size is subsequently derived by subtraction of the fixation durations of the different conditions (Rayner, 1975).

Although the existence of a preview benefit seems reliable and robust (see Schotter et al., 2012 for a review), recent evidence shows that the preview benefit comes with an intrinsic processing cost. Specifically, the baseline conditions implemented have recently come under scrutiny based on evidence showing that the nature of the parafoveal masks used and their relationship to the target (i.e., related or unrelated) in invalid conditions lead to differential processing costs (Vasilev & Angele, 2017). In addition, readers' awareness about the display change modulates the size of the preview benefit, as aware participants—who presumably have larger perceptual span—become more conservative in processing the upcoming item (e.g., White et al., 2005). Moreover, Kliegl et al. (2013) examined how the difference in fixation durations between random letter and identical parafoveal masks (i.e., the preview benefit) depends on processing of the pretarget word and concluded that "we need to keep in mind that the term 'preview benefit' is really a combination of benefit and costs … Relative to this baseline, are we not forced to interpret the increasing gap between the two preview conditions with increasing preview space as preview cost due to interference from the parafoveal random-letter string rather than as preview benefit due to the correct preview of the later target word?" (p. 14).

Interestingly, at about the same time Hutzler et al. (2013) acknowledged that the preview benefit was based on the assumption that the baseline condition (e.g., XXX masks) suppresses parafoveal processing without interfering with foveal processing, taking the neutrality of the parafoveal masks for granted. By analyzing fixation-related brain potentials, Hutzler et al. showed that unrelated parafoveal masks interfere with foveal word recognition. In a follow-up study, Hutzler et al. (2019) adopted an alternative methodology, in which they manipulated the salience of the parafoveal masks to account for baseline considerations, and concluded that in the classical boundary paradigms mask preview costs are introduced by the baseline that lead to overestimation of the preview benefit.

Although the search for a "pure" baseline for the preview benefit is still ongoing, the meta-analysis of Vasilev and Angele (2017), including 93 experiments that used the boundary paradigm examining preview benefit effects, showed that facilitation and interference are related to the properties of the material used and the information that the material carries. Irrelevant material causes interference while relevant material leads to facilitation, again suggestive of parafoveal-on-foveal effects (Vasilev & Angele, 2017).

Parafoveal-on-foveal effects remain highly controversial, as the effect appears either small or inconsistent (Kennedy & Pynte, 2005; Rayner et al., 2005; Drieghe, 2011). Much of the confusion related to their existence is because of the blending of reading-like and pure reading paradigms, with the former providing evidence for its existence and the latter producing less robust results (Drieghe, 2011). This observation led Rayner and Juhasz (2004) to argue that studies providing evidence for parafoveal-on-foveal effects use visual search or pattern matching tasks that resemble but are not reading.

However, corpus studies investigating parafoveal-on-foveal effects under normal reading conditions provide convincing evidence for the existence of these effects. Specifically, Kennedy and Pynte (2005) analyzed eye-movement data derived by 10 English and 10 French speaking participants while reading newspaper articles (i.e., approx. 50000 words per participant). What they found was that the features of the parafoveal word had an effect on the inspection time of short (not long) foveal words, indicating parafoveal-on-foveal effects when the properties of the foveal word allow it. Similar results were obtained by the study of Kliegl et al. (2006) who examined parafoveal-on-foveal effects in a grand corpus (222 participants reading 144 sentences) and replicated the previous findings of Kennedy and Pynte for the parafoveal frequency effect with the authors concluding that processing of words is distributed across fixations, that is, processing of current and nearby words runs in parallel.

Studies examining the role of interword spacing in word recognition further support the notion of nearby-items interference in reading tasks. More specifically, increasing the distance between the word to be recognized and the words around it has been found to increase reading speed and facilitate word recognition (Drieghe et al., 2005; Rayner et al., 2013; Slattery & Rayner, 2013; Slattery et al., 2016; Spragins et al., 1976). Similarly, it has been observed that inserting interword spacing in languages in which interword spacing is not used (e.g., Chinese or Thai) facilitates word recognition (Hsu & Huang, 2000; Winskel et al., 2009). In addition, second language learners seem to benefit by the introduction of extra space in these languages (i.e., Chinese; Shen et al., 2012). This last findings indicates that, if the material is less familiar and requires intensive processing, increasing the distance from adjacent words reduces nearby-items interference in word identification, resulting in improved performance. Finally, the facilitative effect of increased interword spacing has also been obtained in the study of Drieghe et al. (2005). In this study, texts in which predefined target words were followed by one or two blank spaces were compared. In addition, in Experiment 2 a "z-reading" task was used, in which participants were asked to scan strings of z symbols (e.g., zzzzz) and pretend to read. What was found was that double space facilitated recognition of words when compared to single space, a pattern not observed in z-reading, indicating that facilitation due to increased spacing is specific to linguistic materials.

Finally, the foveal load hypothesis is consistent with the idea of nearby-items interference in multiple displays. The foveal load hypothesis refers to the notion that the complexity of the currently fixated word alters the amount of information acquired from the parafovea. Henderson and Ferreira (1990) were the first who described this effect by manipulating the difficulty of the foveal word in a gaze-contingent display in order to investigate dynamic adjustments of the perceptual span during reading. In Experiment 1, the foveal word was either a high or low frequency word. Difficulty in word recognition and, consequently, foveal load was taken to vary as a function of word frequency, that is, less load for high frequency words. Assuming that the perceptual span is dynamically adjusted based on the difficulty of the words, the authors proposed that the preview benefit should be smaller for low frequency words than for high frequency words. The results confirmed their hypothesis by showing a 10–20 ms preview effect on word $n + 1$ when word n was of high frequency. No such preview effect was observed for low frequency words. Similar results were obtained in their second experiment, in which the manipulation of difficulty for word n was at the syntactic level. The authors concluded that when foveal load is high, less information is acquired from the parafovea and the perceptual span is dynamically adjusted.

Although moment-to-moment alterations in eye movements and the perceptual span are widely accepted (Rayner, 2009, 2015; Veldre & Andrews, 2018), the foveal load hypothesis remains controversial because of the baselines used to identify the effect. Based on the fact that this hypothesis is grounded on alterations of the preview benefit effect, it inevitably inherits all criticisms described earlier in reference to the subtraction between valid and invalid conditions and the baseline used (Veldre & Andrews, 2018; Vasilev et al., 2021).

2.4 Summary and contribution to the existing research

With the literature review related to single-item (i.e., discrete) and multi-item (i.e., serial) displays I aimed to highlight the different processing demands underlying successful performance. Evidence related to RAN and Stroop tasks suggests that multiple displays require parallel processing of successive items, thus allowing participants to speed up their responses depending on the automaticity of the material used (i.e., alphanumeric vs non-alphanumeric). Based on that, I moved a step further by suggesting that parallel processing of items might carry an intrinsic cost due to nearby-items interference and referred to the relevant literature that supports my claim. The reviewed evidence from various research fields implementing different paradigms suggests that nearby items may interfere with processing of the target item and, hence, impair its identification. Flanker tasks using different material clearly show that flankers interfere with the target when incompatible. Preview costs, parafoveal-on-foveal effects, and the foveal load hypothesis further support this notion. Facilitation of word recognition by increased interword spacing further strengthens the proposal of nearby-items interference due to parafoveal and parallel word processing. If so, it seems reasonable to assume that during multi-item tasks control implementation is required, making performance prone to control failures. This leads me to the adoption of cognitive control theories in approaching performance in multiple displays. Two currently popular theories of cognitive control are introduced in Chapter 3.

3 Cognitive control implementation

Cognitive control is implemented in challenging and typically conflicting tasks in order to adapt to them and perform them successfully. One critical question for the literature of cognitive control is how the cognitive system is able to determine how much control is needed in order for a task to be accomplished. Detection of situations that require cognitive control is a central aspect of human behavior (Botvinick et al., 2001; Botvinick et al., 2004).

To address this question, Botvinick et al. (2001) proposed the conflict monitoring hypothesis, which was subsequently extended by Shenhav et al. (2013) with the introduction of the Expected Value of Control theory.

3.1 The conflict monitoring hypothesis[2]

According to Botvinick et al. (2001), a function exists, namely conflict monitoring, which is responsible for detecting and evaluating a potential conflicting situation. After conflict is confirmed, this system "informs" the control centers and gives rise to behavioral adjustments and adequate information processing to prevent performance decrements. 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 specifically, Botvinick et al. (2001) developed two computational models in order to support the conflict monitoring hypothesis and explain neuroscientific and behavioral data related to the Stroop and Eriksen flanker task, which are also the focus of this project.

---

[2] This section is based on Ziaka, L. (2014). Study of intratask components of the Stroop task: Column Time and Pause Time. [Master's Thesis, Athens University]. Pergamos.

To simulate the Stroop task, Botvinick et al. (2001) adapted the model of Cohen and Huston (1994) and added a conflict monitoring unit, which primarily corresponds to the role of ACC. The model has two input layers, one for the color of the stimulus and one for the word, which are connected to an output layer for potential responses. The task demand (or control) units correspond to color naming and word reading, and their activation is responsible for the response. The conflict monitoring unit is connected to the rest of the network and receives input from it. Conflict is defined as "the simultaneous activation of mutually inhibiting units" (p. 630). When incompatible units are inactive the energy of the model is equal to zero and no conflict occurs. The same is true in the case of only one active unit. In contrast, when both incompatible units are active the energy increases, indicating conflict. The level of conflict depends on the level of activation of both units: if maximal, the conflict is strong. There is also a feedback loop (connection) between the conflict monitoring unit and the task demand units. As a result of the model's architecture, in frequent incongruent trials continuous activation of the conflict monitoring unit raises the activation of the control units (i.e., strengthens control and improves performance), while the opposite is true for infrequent incongruent trials (i.e., activation of the conflict monitoring unit declines and activation of the control units drops, resulting in lax control).

An analogous model was constructed for the Eriksen Flanker Task, where it is supposed that interference is caused due to the nearby stimuli. The only difference between the two models is that input units were assigned to the spatial location of the stimuli and the task demand units were replaced by attentional units. Botvinick et al. (2001) managed to successfully simulate all the critical findings regarding the Stroop and the Eriksen Flanker Task, that is, the sequential adjustments in the Eriksen

Flanker Task and the trial type frequency effects and improvement in performance in the single-item Stroop task. Consequently, this work supported the notion of conflict monitoring.

3.2 The Expected Value of Control theory

More recently, Shenhav and colleagues (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 is a decision-making function which refers to the decision about the task goal (identity; here, color naming) and how intensively this goal must be pursued (intensity). Specification is based on future rewards by trying to maximize them and is expressed in the model by the expected value of control for each control signal. For any given control signal rewards and costs are taken into account as well as the cost of control implementation itself. In other words, the expected value of control is determined by task-related and environmental factors, as for example motivation. Different control signals are determined and the most optimal is selected and applied until an undesired change is detected via monitoring and a new optimal control signal is selected. 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. As indicated by the expected value of control, for EVC theory cost and 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 and colleagues (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).

Furthermore, based on evidence showing that control implementation carries a cost and requires mental effort, which is aversive, incentive components are taken into account in this model. More specifically, it is argued that under conflicting situations a cost-benefit analysis is active and control implementation is a balance between these costs and benefits. "Critically, these benefits are a function of both the expected outcomes for reaching one's goal (reward, e.g., money or praise) and the likelihood that this goal will be reached with a given investment of control (efficacy). The amount of control invested is predicted to increase monotonically with a combination of these two incentive components" (Frömer et al., 2021, p. 2).

Finally, it should be noted that EVC specification is not restricted to one intensity-identity pairing, but multiple pairings can be concurrently active, meaning that cognitive control can be theoretically applied in more than one tasks at the same time. Although Shenhav et al. acknowledge that in reality cognitive control has constraints, they do not elaborate on this point, arguing that simple control-demanding tasks are the most common circumstance (for a different view see Schuch et al., 2019).

Summary— 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. I aimed to evaluate this prediction in single-item and multi-item Stroop tasks. The methodology and data are presented in detail in Chapter 4.

4 Methodology and data

Depending on the research question, I combined an experimental approach—by using experimental paradigms which are helpful for isolating specific cognitive processes (Draheim et al., 2021)—and the individual differences approach—which allows to identify the contribution of individual traits in performance. More specifically, for addressing the research question *How can nearby-items interference affect performance when processing multiple displays?* an experimental paradigm was used comparing within-task performance between single-item and multi-item Stroop tasks. A similar paradigm was used for the research question *How is control applied in tasks posing different processing demands?* In this study, however, I additionally included pupillometry and gaze measures to shed more light on the underlying cognitive processes and disentangle alternative interpretations. Finally, for the research question *What processes distinguish between readers of different levels?* a correlational approach was adopted focusing on individual differences. Table 1 provides an overview of the research questions, empirical data, and main findings of the three articles.

Table 1

Overview of the three articles

Studies	Research question	Sample	Focus and tasks	Main findings
Article I Conflict monitoring or multi-tasking? Tracking within-task performance in single-item and multi-item Stroop tasks	*How can nearby-items interference affect performance when processing multiple-displays?*	Adults and children	• Multi-item and single-item Stroop tasks • Within-task performance evaluation of behavioral measures • Between-tasks and between-groups comparisons	Within-task performance decline only in the multi-item task for both conditions (i.e., incongruent and neutral) Different pattern of results for the neutral conditions depending on developmental stage
Article II Cognitive control beyond single-item tasks: Insights from pupillometry, gaze, and behavioral measures	*How is control applied in tasks posing different processing demands?*	Adults	• Multi-item and single-item Stroop tasks • Within-task performance evaluation of behavioral, pupillary, and gaze measures • Between-tasks comparison	*Multi-item format* • Within-task performance decline • Pupil constriction • Dwell time increase in both conditions *Single-item format* • Stable within-task performance • Pupil constriction • Dwell time decrease in both conditions
Article III Task format modulates the relationship between reading ability and Stroop interference	*What processes distinguish between readers of different levels?*	Adults and children	• Multi-item and single-item Stroop tasks • Serial and discrete reading measures • Correlational analysis • Residualized scores analysis	Reading performance was primarily related to the multi-item version of the Stroop task and not to the single-item version

4.1 Study 1: How can nearby-items interference affect performance when processing multiple-displays?

For investigating how nearby items might affect processing in multi-item tasks, I used an experimental design which allowed isolating the underlying cognitive processes, while adopting a within-participant design permitted controlling for variability at the individual level. I chose to use the Stroop task as it allows straightforward comparisons between its single-item and multi-item format. For the single-item format each item appeared on the screen individually to be named, while in the multi-item Stroop task a card was displayed on a computer screen in which all items were simultaneously present and participants were asked to name the color of the ink as fast as possible without errors from top-to-bottom.

The analysis included within-task evaluation performance by dividing the tasks into three equal parts (i.e., 20 items divided into three blocks) and estimating mean response time and errors for each part. By doing so, I was able to observe within-task performance variations as a function of task format, that is, single-item or multi-item.

4.1.1 Power considerations

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. The decision to include approximately 40 participants in the three Experiments was based on the common sample size used to examine the study of Stroop tasks (e.g., Dulaney & Rogers, 1994; Ellis & Dulaney, 1991; Salo et al., 2001).

Post-hoc power analysis using the effect sizes and SDs observed in the experiments would not have much to offer as it could not distinguish between true and false negatives, and would actually lead to a repetition of the observed p-values in a different way (Lakens, 2022).

Treating, however, the first experiment as a pilot study, a prior power analysis indicates that a sample size of 25 would be sufficient to achieve a power of 0.95 to observe the effect of interest for, e.g., the incongruent condition in all subsequent experiments. The sample size used in all experiments far exceeded this estimate.

4.2 Study 2: How is control applied in tasks posing different processing demands?

The aim of Study 2 was to expand on the findings of the previous study by adopting the same experimental design, additionally including pupillometry and gaze measures which would allow to disentangle between possible alternative interpretations, that is, performance variations due to capacity constraints or due to incentives (i.e., withdrawal from the task via a cost-benefit analysis).

For this study, three different variants of the multi-item Stroop task were used, that is, the variant used in Study 1 plus a variant with increased between-item spacing and a third one with a different naming direction (i.e., left-to-right). The inclusion of the additional Stroop task variants aimed to ensure that any results obtained were not related to specific features of the task, thereby strengthening their robustness.

In this study, each condition and task was divided into 12 blocks of 5 items. Within-task performance evaluation for each task and condition was conducted in order to observe variations in all dependent measures, that is, behavioral (i.e., response rate and errors) and eye-tracking (i.e., pupil size and dwell time). Linear trend analysis was additionally performed in order to detect specific patterns of within-task variations in the dependent measures.

4.2.1 Methodological considerations for pupillometry

Pupil is very sensitive to a variety of properties of items including color (Markwell et al., 2010; Mathôt, 2022). For that reason, in Article II the ink for the colors used corresponded to the ones frequently used in pupillometry Stroop studies (e.g., Hasshim & Parris, 2015; Hershman et al., 2020; Laeng et al., 2011). Crucially, and in contrast to some previously published studies investigating pupillary responses in single-item Stroop tasks (e.g., Hershman & Henik, 2019; Hershman et al., 2020; Laeng et al., 2011), no blue colored items were used, as blue color is known to produce sustained pupil constriction when compared to red items (Markwell et al., 2010; Mathôt, 2022).

Furthermore, pupil size was averaged for 5 consecutive items consisting of different combinations of the three colors, and, hence, the impact of luminance on pupil size for the different blocks was kept relatively constant. Importantly, our main hypotheses were focusing on within-task variations in pupil size and not to direct comparisons between tasks. Even in the cases in which between-task comparisons were performed to test for the impact of between-items spacing and naming direction on pupillary responses, the analysis focused on the effect of time-on-task (i.e., block) on the linear trends observed. As such, luminance differences between tasks could not affect the results related to the hypothesis of interest.

Finally, as also noted in the article, position artifacts can be introduced in the study of pupil size if different parts of the screen have to be fixated, resulting in a decrease or increase in pupil size depending on the angle of the eye with respect to the eye-tracker (Mathôt, 2018). For that reason, in Article II I added a “scanning” condition before each experimental condition to serve as a reference baseline (Gagl et al., 2011; Mathôt, 2018). Scanning conditions in the multi-item tasks matched the corresponding experimental conditions in stimulus position and extent, with "items" consisting in strings of letters displayed in black color. Participants were asked to simply scan the items without any additional processing. Pupil size in the scanning conditions was included in the main analysis in order to account for position artifacts.

4.3 Study 3: What processes are distinguishing readers with different level of fluency?

For Study 3, the approach was correlational focusing on individual differences. Because I was interested to investigate if Stroop interference is directly related to reading ability, as previously proposed (Protopapas et al., 2007), and irrespective of the format used, that is, single-item or multi-item, I revisited six experiments and analyzed their relationship to discrete and serial reading in adults and children.

4.3.1 Power considerations

Sufficient power of the individual studies included in Article III was not determined by power analysis because the main interest was not the probability in observing a significant relationship but, rather, its size estimation, for which power analysis has little to offer. For that reason, a meta-analytical approach was applied to clearly depict the dispersion between individual studies and the reliability of the final estimate. As such, although sample size for individual studies could be considered small, the adopted meta-analytic approach provides confidence and precision by its post hoc reliability estimation.

4.4 Ethical aspects

The studies described in this dissertation were conducted in Greece and an ethics approval was not required as per applicable institutional and national guidelines and regulations for both the adult and children studies. The children studies were 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.

For all studies, the data were anonymized and participants could not be identifiable.

In addition, for the studies including children, except the consent required by parents/legal guardians, I provided the children with sufficient information about their involvement in the study in order to additionally obtain their consent and ensure that their participation was voluntary.

5 Summary of the articles

The aim of the present PhD project was to examine in depth the underlying cognitive mechanisms of processing single-item and multi-item tasks within their course, taking into account evidence suggesting the presence of nearby-items interference in simultaneous presentation of items. For serving this goal, I conducted three studies which are presented in three different articles. A brief description of the articles follows, including main questions and findings. In Chapter 6 the findings and their implications are discussed in detail.

5.1 Article I

Ziaka, L., & Protopapas, A. (2022). Conflict monitoring or multi-tasking? Tracking within-task performance in single-item and multi-item Stroop tasks. *Acta Psychologica, 226*, 103583. https://doi.org/10.1016/j.actpsy.2022.103583

The research questions that guided this article was if simultaneous presentation of items elevates the processing demands in multiple displays because of nearby-items interference and if the cognitive control system is challenged by these demands. Having as starting point the assumption of the conflict monitoring hypothesis and the Expected Value of Control (EVC) theory that a control system responsible for detecting conflicting occasions and adapting to them dynamically, I aimed to compare single-item and multi-item Stroop tasks to test this prediction. Performance was estimated within the tasks to allow tracking of within-task performance. I hypothesized that if nearby items create interference, multi-tasking requirements might be embedded in the multi-item version, challenging the cognitive control system and leading to performance impairments. To serve my goal I 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. Furthermore, this effect was modulated by the developmental stage, with children showing steep performance decrements in both conditions of the multi-item Stroop task, suggesting an influence of the presumed maturity of the control system in processing multiple displays. Altogether, the findings of the present study supported the existence of nearby-items interference in multiple displays and justified my conceptualization of the multi-item Stroop task as a multi-task, suggesting capacity constraints on control implementation and allocation under conditions requiring parallel execution of multiple cognitive tasks.

5.2 Article II

Ziaka, L., & Protopapas, A. (under review). Cognitive control beyond single-item tasks: Insights from pupillometry, gaze, and behavioral measures. *Journal of Experimental Psychology: Human Perception and Performance*.

In this study, my starting point was the observation that laboratory tasks used to examine cognitive control implementation and allocation are usually single-item tasks which might pose reduced processing demands on the individuals. This point has implications for the generalizability of theories of control implementation, especially when taking into account that real-life situations often require multi-tasking. Specifically, the research question was how control is applied under conditions that go beyond single-item presentation. In the present study, behavioral (i.e., response time) and eye-tracking measures (i.e., pupil size and gaze) were combined aiming to explore the implications of format differences for cognitive control. Within-task tracking of all dependent measures in single-item and multi-item tasks was adopted. The results indicated within-task performance decline in the multi-item version of the Stroop task, accompanied by pupil constriction and dwell time increase, in both the incongruent and the neutral condition. In contrast, no performance decline or dwell time increase was observed in the course of the single-item version of the task. However, pupil constriction was also evident in the single-item version, which in combination with behavioral and gaze measures is suggestive of practice effects. In contrast, for the multi-item Stroop tasks the pattern of results points to capacity constraints and dynamic narrowing of attention. This finding is beyond the explanatory range of current theories of cognitive control and has implications for cognitive control research. In total, this study highlighted the need for better understanding the cognitive demands of multi-item tasks and raises substantial concerns for the use of primarily single-item tasks in the study of cognitive control implementation.

5.3 Article III

Ziaka, L., Skoteinou, D., Protopapas, A. (2022). Task format modulates the relationship between reading ability and Stroop interference. *Journal of Experimental Psychology: Human Perception and Performance, 48*(4)*,* 275–288. https://doi.org/10.1037/xhp0000964

The aim of this study was to examine the relationship between Stroop interference and reading ability and possibly identify which processes are critical for distinguishing readers of different levels. More specifically, the motivation for conducting this study was based on previous research showing that Stroop interference and reading ability are negatively related, with higher reading skills associated with less interference, and the proposal for a direct link between interference and the speed of inhibition of the task-irrelevant dimension (i.e., word). However, a limitation was identified in previous studies, as the majority of them used the multi-item version of the task, leaving open the question of possible format dependence of this relationship. I hypothesized that, if a direct link between interference and the speed of inhibition exists, it should apply regardless of the format of the Stroop task, that is, whether stimuli are presented simultaneously (multi-item version) or individually (single-item version). For this goal, I examined data from six experiments using single-item and multi-item Stroop tasks and their relationship to serial and discrete reading measures in different developmental stages (i.e., childhood and adulthood). The results indicated that reading performance was primarily related to the multi-item version of the Stroop task and not to the single-item version, thus questioning the direct link between inhibition and interference as an interpretation of the reading–interference relationship. I have argued that cascaded processing of successive items, and the ability to monitor and control this process via attentional adaptations, is the cognitive mechanism regulating the relationship between reading and interference. As such, the link between Stroop interference and reading seems indirect and their relationship is determined by the efficiency in temporally overlapping processing of nearby items.

6 Discussion

The aim of the present dissertation was to identify the underlying cognitive mechanisms governing performance in single-item and multi-item presentation of the material, focusing on the dynamics of interference and cognitive control for the execution of tasks. In the following sections we will discuss in detail the main conclusions drawn and their implications.

6.1 Nearby-items interference in processing of multiple displays

In the context of the present dissertation, it was shown that the presence of nearby items in multiple displays elevates the processing demands posed to the individuals and might result in performance deterioration. Overall the results suggest that multiple displays require the parallel execution of different tasks and subtasks resembling multi-task functioning (Koch et al., 2018; Meyer & Kieras, 1997; Monsell, 1996) and resulting in task interference (Wickens, 2002) and crosstalk (Fischer & Plessow, 2015).

By showing that, the present dissertation contributes to our understanding about the complexity of multiple displays and highlights the need to approach single-item and multi-item tasks as distinct entities that require different processes and sub-processes. In addition, the fact that performance within the tasks showed a decrease suggests that averaging response time throughout the entire task might be misleading and lead to overly narrow conclusions by overlooking critical aspects contributing to the observed behaviors.

As noted in Article I and II, this is not the first time that performance decrements have been observed in the context of multi-item (i.e., serial) tasks. Similar results have been previously reported by Amtmann et al. (2007) in row-by-row analysis of multi-item naming tasks and by Klein et al. (1997), who examined the effect of time 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). In fact, the study of Klein et al. supports standardized Stroop tasks for examination of attention and inhibition deficits that are brief and are not affected by time on task (e.g., Victoria Stroop Test; Strauss et al., 2006) and indicates that within-task performance variations should be seriously taken into account.

6.2 Lockout-scheduling and dynamic narrowing of attention as a response to nearby-items interference

The cognitive control theories examined in the present dissertation, that is, the conflict monitoring hypothesis and EVC theory, cannot account for the pattern of results in the described studies. For both approaches cognitive control is dynamically adjusted based on the requirements of the task, leading to stable or improved performance (Botvinick et al., 2001; Shenhav et al., 2013). Although the EVC theory could approach our data based on the assumption that participants withdraw from the task due to low motivation, this proposal is weakened by the results of Article II including pupillary responses and gaze measures.

The studies of the present dissertation speak in favor of a shift from a more parallel to a more serial processing via "lockout scheduling" (Meyer & Kieras, 1997, p. 20) and the narrowing of attention in order to be able to deal more efficiently with task requirements. More specifically, I have argued that in order for the current response to be protected by nearby-items interference so that successful performance can be secured, a serial strategy is adopted during the multi-item tasks, whereby perception of the next item is delayed until planning of the response to the current item has been completed

(Roelofs, 2007). This strategy effectively blocks the high degree of parallelism and, as a result, although serial processing of each item through successive stages and serial processing of successive items are preserved, parallel processing of multiple items in the various stages becomes impaired.

Furthermore, it appears that individual characteristics, as for example reading ability, determine the moment of the shift from one item to the next. More specifically, the results of Article III suggest that reading skills affect the shift criterion for the transition between items, especially in the incongruent condition of the Stroop task. Because word-form encoding of the irrelevant word is more automatic for skilled readers, compared to poor readers, the level of complexity in processing the current item is lower for skilled readers. Consequently, the shift criterion for unlocking attention and initiating processing of the next item is positioned earlier in time, allowing skilled readers to move faster to the next item and increase the degree of parallel processing.

Finally, as stated in Article II, my argumentation of the adoption of a more serial processing during the task, and the consequent narrowing of attention, gets further support by studies examining attentional breadth in visual research (Brocher et al., 2018) and its impact on pupil size. More specifically, attentional breadth is based on the focus of attention, which is narrowed when the focal point of attention is centered, and broadens when attention is peripherally expanded (Mathôt, 2020). “This terminology implicitly characterizes attention as a zoomlight that changes size while remaining centered on central vision, rather than as a spotlight that moves around in space” (Mathôt, 2020, p. 436). Pupil seems to reflect these attentional shifts of covert attention, as it constricts when attention is narrowed and dilates when attention is broadened (Brocher et al., 2018; Daniels et al., 2012; Mathôt, 2020). Furthermore, there seems to be a relationship between exploration, exploitation, and attentional breadth. Exploitation is considered to emerge when the focus is on a single task, which suggests a smaller attentional breadth, whereas exploration emerges when the focus of attention is expanded (e.g., in task switching), leading consequently to a larger attentional breadth and causing behavior to be prone to distraction (Mathôt, 2018). What is of importance here is that exploitation and focus on a single task results in smaller pupil, while the opposite is true for exploration (Jepma & Nieuwenhuis, 2011; Mathôt, 2018).

In my opinion the relevance of these studies to the present dissertation is straightforward. Perceptual span and parafoveal processing are all about expansion of attention and can be considered as a special case of attentional breadth in reading and naming tasks. As such, it seems reasonable to assume that when participants, because of nearby-items interference, "decide" to narrow their attention via lockout scheduling and to shift from a more parallel to a more serial processing, this should be reflected on their pupil size as constriction. Indeed, this what was observed in Article II. Pupil constricted during the task, while performance dropped and dwell time increased. Altogether, these findings corroborate our proposal of changes in attentional shifts during the multi-item tasks. Moreover, these findings contribute to the ongoing debate about the source of pupil constriction under dual-task and excessive load conditions by suggesting that pupil reflects an overload state because of the narrowing of attention.

6.3 Is interference spatial, temporal, or both?

In Chapter 2, I provided evidence to support the proposal of nearby-items interference in multiple displays. This evidence is primarily based on studies that examine the impact of adjacent items on target recognition and might erroneously lead to the conclusion that the nature of nearby-items interference is exclusively spatial. However, this is not the case.

To elaborate on this point, I will turn to my definition of nearby-items interference, namely that "nearby-items interference is defined as impairment in task performance by simultaneous presentation of items in spatial proximity to the target, with target and nearby items requiring the concurrent execution of multiple processes". Indeed, this definition has a spatial component as it suggests that in order for nearby-items interference to emerge the target item should be in spatial proximity to adjacent items. Spatial proximity sets the prerequisite for nearby-items interference to emerge. However, this is not enough. The definition goes on by suggesting that target and adjacent items require the concurrent execution of multiple processes. This claim suggests a temporal component, meaning that these processes must overlap in time (i.e., run in parallel) in order for performance to be affected. Moreover, the temporal dimension of nearby-items interference has been emphasized in Article I by the conceptualization of the multi-item Stroop task as a "multi-task" based on the proposal that "its successful execution entails simultaneous and parallel activation of more than one task sets" (p. 3) and "requires *time sharing* among concurrent tasks and subtasks" (p. 3). Taken together, all these arguments converge to the proposal that simultaneous presentation of items in multiple displays leads to this temporal overlap.

In fact, this is exactly what the proposed shift from a more parallel to a more serial strategy via lockout scheduling aims to mitigate: It separates processing of simultaneously presented items by *postponing* processing of the next item until processing of the current item is secured from nearby-items interference; crucially, the separation concerns time, not space. Hence, nearby-items interference is neither spatial nor temporal; it is both.

Based on the study of Roelofs (2007), in the articles of the present dissertation I proposed that response planning must be completed in order to move on to the next item, implying that the previous stages of processing (i.e., perception, conceptual identification for color naming, lemma retrieval, and word-form encoding) are more prone to nearby-items interference than subsequent stages (articulation and vocal response; Roelofs, 2003, 2007). Although this claim seems plausible, further research, including eye-voice span measures, is needed in order to determine the “when” dimension of nearby-items interference with high confidence.

6.4 Implications

6.4.1 Implications for educational research and reading acquisition

Although the present dissertation and the paradigms used are based on experimental psychology, the findings reported here have significant implications for educational research.

As presented in Chapters 1 and 2, the starting point of this project was the observation that reading is primarily related to serial rapid naming, while the underlying mechanisms determining this relationship remain unclear. Moreover, based on evidence showing that alphanumeric rapid naming predicts reading more strongly when compared to non-alphanumeric rapid naming, I have tried to highlight the role of the automaticity level of the material in the emergence of this relationship, and argued about the need for cognitive control allocation during these tasks depending highly on the format (i.e., serial or discrete).

The results of the described studies seem to support this assumption by showing that during serial (multi-item) tasks control is not only required, but might modulate within-task performance by virtue of cognitive control failures. More importantly, this seemed to be the case not only for the incongruent condition, but also for the neutral condition, which consists of non-conflicting stimuli. Based on the similarity of the neutral condition with color rapid naming tasks—and with all similar tasks irrespective of the material used—the data suggest that the same processes should also be involved in the latter and might affect performance. By extension, this should be also the case for text reading, where multiple words are simultaneously present. As a result, cognitive control processes and within-task performance variations should be seriously taken into account in rapid naming and reading research as well as in individual differences research examining these topics.

Furthermore, the proposal of nearby-items interference in multiple displays is something we should take seriously into account in reading research. If multiple words are competing for selection when presented simultaneously, as researchers we have to identify the factors that determine the magnitude of this competition and its impact on performance and on the observed behavior. This dissertation represents an initial attempt at systematic investigation of these aspects in serial naming tasks. It remains to be established whether these processes are also active in reading, how they influence reading assessment and reading acquisition, and to which degree.

More specifically, the present dissertation focused on the impact of nearby-items interference in naming tasks; not reading tasks. Consequently, it remains unclear whether nearby-items interference could also have an impact on performance and impair reading at the level of individual letters or whether the word superiority effect suffices to forestall interference. Word superiority refers to faster recognition of letters when they are embedded within a word than when presented alone or in the context of meaningless strings (e.g., Reicher, 1969; Wheeler, 1970), an effect that could conceivably protect from nearby-items interference. However, previous evidence suggests that nearby-items interference might emerge at the level of individual elements (i.e., letters) depending on reading ability. For example, Spinelli et al. (2005) examined word length effect in fluent readers and readers with dyslexia and found that increased length produced a linear increase in response time for readers with dyslexia, while the performance of fluent readers was affected only for words consisting of more than five letters. Interestingly, the authors proposed that fluent readers switch between parallel and sequential processing based on word length, that is, words with a length less than five letters promote parallel processing of individual letters while words consisting of 5–8 letters promote sequential processing. In the case of readers with dyslexia sequential processing seemed to be the default strategy. This interpretation chimes with Article III of the present dissertation in which a shift from parallel to serial processing was proposed for multi-item Stroop tasks based on the reading level of the participants and the automaticity of the material. Furthermore, nearby-items interference within individual words is also supported by evidence showing that beginner readers process individual words in a serial way (de Jong et al., 2011) and that increased inter-letter spacing might facilitate recognition of foveally presented words in readers with dyslexia (Spinelli et al., 2002). As a result, it remains to be determined whether within-words nearby-items interference indeed exists and, if it does, how it affects performance across levels of reading skill.

In addition, within-task performance evaluation, as an alternative methodology to average response times, has a lot to offer to our understanding of reading processes by allowing us to investigate transient phenomena within a task and to study how individual traits might affect these phenomena. Such transient phenomena are completely missed if response times are averaged throughout the entire task. In contrast, by tracking performance within multiple displays it can be determined—for example—if reading ability modulates readers' resistance to cognitive control failures, causing performance decrements to emerge at an earlier or at a later time during the task.

Woods et al. (2005) have aptly remarked that "because a large majority of children's reading material (i.e., textbooks, standardized tests, literature) is found in printed form, it is particularly important to examine readability of printed text for children. Publishing companies have guidelines, but these are often based on font types and sizes most frequently used by other publishing companies rather than on empirical data investigating legibility and readability" (p. 86). Hence, the notion of nearby-items interference and the need for cognitive control implementation also has significant practical implications for reading acquisition, as the findings of the present dissertation suggest that a) item recognition is not immune to the presence of other items in spatial proximity to the target item and b) not all readers are impacted to the same degree by simultaneous item presentation, which seems to depend on their reading ability.

Taking these two points into account, my findings indicate that beginning and less proficient readers might benefit by an increased distance between words. More specifically, usually a one-space distance between words is implemented in most of the reading material, which consequently means that nearby-words are less than 1 degree of visual angle away from the target word, resulting presumably in processing costs. A more efficient strategy would be to increase the interword distance in order to minimize processing costs, but not so much as to deny the facilitative effects of parafoveal processing and perceptual span. Similarly, increased intraword (i.e., inter-letter) distance may be also beneficial as it would free up resources from nearby-letters interference (for a similar view Perea & Gomez, 2012; Perea et al., 2012; Spinelli et al., 2002) in beginning or struggling readers. In addition, font characteristics should be taken into account as it has been found that the effect of inter-word and inter-letter distance on word recognition seems to be font dependent (Slattery & Rayner, 2013; Slattery et al., 2016).

Finally, the attentional processes engaged in reading should also be taken into consideration when planning reading interventions. More specifically, it has been found that the quality of lexical representations, as indicated by the combination of reading and spelling ability, modulates the extent of perceptual span and parafoveal processing. Specifically, studies have showed that individuals who excelled in both of these measures were more affected by the denial of parafoveal information than individuals who scored low in both reading and spelling tasks (Veldre & Andrews, 2014, 2015a, 2015b). These findings, taken together with the findings of the present dissertation showing that attentional shifts happen more rapidly in skilled readers than in poor readers, imply that targeting the quality of lexical representations by focusing on the refinement not only of reading but also of spelling might make readers less prone to distraction from nearby items and thereby reduce the processing costs induced by the simultaneous presentation of items.

6.4.2 Implications for clinical research and assessment

The multi-item version of the Stroop task is widely used in clinical settings to assess deficits in attention and inhibition (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) mainly because of its easy and effortless administration. However, multi-item and single-item versions do not produce the same results when applied in clinical population as in some studies patient groups show abnormal interference, whereas in other studies no such pattern is observed, leading to inconsistencies (Buchanan et al., 1994; Carter et al., 1992; Henik & Salo, 2004). A similar pattern has been reported in aging research and in the comparison between older and younger adults (Ludwig et al., 2010).

As noted in Articles I and II of the present dissertation, it seems that two main factors contribute to the inconsistencies observed, that is, (a) within-task variations in the neutral condition, and (b) the common practice for interference estimation, that is, subtraction between the two conditions.

More specifically, my studies showed that performance in the neutral condition is not stable within the multi-item Stroop task. In addition, the findings of Article I and Article II indicate that even if tasks share the same experimental stimuli, this does not ensure that different versions of the same task require the same underlying cognitive processes for goal-oriented behavior to emerge. Task administration modulates performance. Because the Stroop effect (i.e., the interference effect) is a product of subtraction, the findings of the present dissertation suggest that the impact of additional processes (i.e., cognitive control implementation) on the neutral condition depending on the population under examination (here, adults and children) may help us understand the origin of the inconsistencies observed in the literature regarding a variety of disorders and populations. As Salo et al. (2001) puts it, "attention to methodological differences when interpreting studies in patient populations is essential if we are to better understand patterns of cognition in patients and healthy populations" (p. 470).

Similarly, in Article II I discussed the previous suggestion about speeding up of responses in the neutral condition of the multi-item Stroop task, making this condition to appear faster in this version than in the single-item version, as a possible source of differences in interference estimates between the two versions (Salo et al., 2001). However, the results of Article II showed that in the beginning of the task not only the neutral but also the incongruent condition is faster compared to the single-item version.

However, this advantage is only temporary, as it is abolished further into the task, especially in the incongruent condition, thereby weakening the proposal of Salo et al.

Finally, by adopting different multi-item variants, in Article II I pointed to a presumably significant role of naming direction, because left-to-right naming appeared more demanding than naming top-to-bottom. This finding runs counter to the proposal of McCown and Arnoult (1981) who suggested that the form of the Stroop task does not alter its conflicting nature. Further studies are needed in order to replicate the results of Article II, as it will be of relevance for clinical practice, in which different standardized Stroop tasks are used varying among other things in naming direction (e.g., left-to-right naming in Victoria version vs top-to-bottom naming in Golden version; for a review see Strauss et al., 2006), all aiming to assess the same types of deficits.

6.5 Additional considerations and limitations

As noted in Article III of the present dissertation, the impact of sequential effects on within-task performance variations cannot be precluded. Specifically, I referred to negative priming effects, that is, the increase of response times in the incongruent condition when the incorrect word-response of the preceding item corresponds to the correct color-response of the current item (e.g., the word “green” printed in red followed by an item printed in green color). Although I addressed the reasons for which negative priming cannot account for the pattern of results obtained, I acknowledged the usefulness of studies examining its effect on the relationship between Stroop interference and reading.

Furthermore, as per common practice (MacLeod, 2005), the presentation of all multi-item Stroop tasks used in the present dissertation was blocked, that is, each card consisted of one kind of stimuli, either only neutral or only incongruent. However, a frequently used approach to examine cognitive control processes is to observe response time variation as a function of the required response in the previous trial (i.e., congruent or incongruent; Gratton et al., 1992). The common finding is that interference is larger after a congruent trial than after an incongruent one (termed the Gratton effect). The Gratton effect could not have played a role in the present study because of the blocked presentation of tasks. Moreover (and importantly), control adaptations were not completely effective, as was evident in the performance decrements of Article I and II. Still, it would be interesting to examine the impact of the Gratton effect on within-task performance in a modified version of the multi-item Stroop task consisting of both neutral and incongruent items. By taking advantage of eye-tracking measures it could be ascertained whether the Gratton effect would introduce additional processing costs, because of the need to dynamically adapt control based on the nature of the previous trial; or, in contrast, whether it might benefit performance, because of the interruption of rapid, sequential processing of only incongruent items; or whether the proposed switch to a serial strategy runs so strong once adopted that no impact of mixing the items would be observed.

In an effort to have tasks that are comparable with the previously published Stroop literature, I made the methodological choice in all studies to instruct participants to name the color “as fast as possible and try to avoid errors”, putting, hence, emphasis on speed. As acknowledged in Article II, by emphasizing speed I may have indirectly promoted parallel processing for speeding up responses and, consequently, nearby-items interference. Future studies can shed more light on possible effects of instructions on within-task performance.

Furthermore, as noted in Article II, within-task variations were investigated in only one of the tasks used in the cognitive control literature (i.e., the Stroop task), so it remains unclear if the same pattern of results would be obtained in other cognitive control tasks. The same holds for the population included in the presented study, as participants were mostly young adults. However, age may modulate the effect of nearby-items interference in the multi-item Stroop task. The difference between adults and children reported in Article I speaks to this possibility and might be only one of various factors affecting within-task performance, indicating the need for studies in other populations adopting within-task performance evaluation.

In reference to reading, as previously discussed, Stroop tasks were used in the present dissertation rather than reading tasks. Therefore, any conclusions related to the impact of nearby-items interference on reading performance and its modulation by reading ability are indirect. Although the results of Article III strengthen the claim for interference of nearby items in regular reading tasks, more research is needed to be confident about its presence.

Finally, all studies were conducted with Greek-speaking participants. However, Greek is a language with relatively high orthographic transparency, which may modulate some of the effects under consideration. It remains unclear how nearby items might affect processing of multiple displays in other languages having a lower degree of orthographic transparency, as for example in English and French.

6.6 Conclusion

The focus of this dissertation was on the differences between single-item and multi-item tasks and the impact of nearby-items interference on processing multiple displays. With the three studies presented here I tried to highlight the importance of task version when examining cognitive control allocation, modulating factors, and the possible influence of differential processing requirements in more ecological tasks as reading.

Although the introduction of nearby-items interference in multiple displays is new and further investigation is required to be conclusive, altogether the studies of the present dissertation show this proposal must be taken seriously into account when comparing tasks requiring different presentation of the material. Furthermore, I showed that within-task performance evaluation, as an alternative to the common practice of averaging performance throughout the entire task, might have a lot to offer to our understanding of the underlying cognitive mechanisms required for the accomplishment of a task goal.

References

Altani, A., Georgiou, G. K., Deng, C., Cho, J. R., Katopodi, K., Wei, W., & Protopapas, A. (2017). Is processing of symbols and words influenced by writing system? Evidence from Chinese, Korean, English, and Greek. *Journal of Experimental Child Psychology*, *164*, 117–135. https://doi.org/10.1016/j.jecp.2017.07.006

Altani, A., Protopapas, A., & Georgiou, G. K. (2017). The contribution of executive functions to naming digits, objects, and words. *Reading and Writing, 30*(1), 121–141. https://doi.org/10.1007/s11145-016-9666-4

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). Tracking the serial advantage in the naming rate of multiple over isolated stimulus displays. *Reading and Writing*, *33*(2), 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*, *20*(8), 785–813. https://doi.org/10.1007/s11145-006-9041-y

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

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

Boucart, M., Mobarek, N., Cuervo, C., & Danion, J. M. (1999). What is the nature of increased Stroop interference in schizophrenia? *Acta Psychologica*, *101*(1), 3–25. https://doi.org/10.1016/s0001-6918(98)00037-7

Bowers, P. G., & Swanson, L. B. (1991). Naming speed deficits in reading disability: Multiple measures of a singular process. *Journal of Experimental Child Psychology*, *51*(2), 195–219. https://doi.org/10.1016/0022-0965(91)90032-n

Brocher, A., Harbecke, R., Graf, T., Memmert, D., & Hüttermann, S. (2018). Using task effort and pupil size to track covert shifts of visual attention independently of a pupillary light reflex. *Behavior Research Methods*, *50(6)*, 2551–2567. https://doi.org/10.3758/s13428-018-1033-8

Buchanan, R. W., Strauss, M. E., Kirkpatrick, B., Holstein, C., Breier, A., Carpenter Jr, W. T. (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

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

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*(41), 63–65. https://doi.org/10.1093/mind/os-XI.41.63

Chiappe, P., Stringer, R., Siegel, L. S., & Stanovich, K. E. (2002). Why the timing deficit hypothesis does not explain reading disability in adults. *Reading and Writing*, *15*(1-2), 73–107. https://doi.org/10.1023/A:1013868304361

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*(3), 332–361. https://doi.org/10.1037/0033-295X.97.3.332

Cooper, H.M. (1988). Organizing knowledge syntheses: A taxonomy of literature. *Knowledge in Society*, *1*, 104–126. https://doi.org/10.1007/BF03177550

Daniels, L. B., Nichols, D. F., Seifert, M. S., & Hock, H. S. (2012). Changes in pupil diameter entrained by cortically initiated changes in attention. *Visual Neuroscience*, *29*(2), 131–142. https://doi.org/10.1017/S0952523812000077

De Jong, P. F. (2011). What discrete and serial rapid automatized naming can reveal about reading. *Scientific Studies of Reading*, *15*(4), 314–337. https://doi.org/10.1080/10888438.2010.485624

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

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, *150*(2), 242– 275. https://doi.org/10.1037/xge0000783

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, *145*(5), 508–535. https://doi.org/10.1037/bul0000192

Drieghe, D. (2011). Parafoveal-on-foveal effects in eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford handbook on eye movements* (pp. 839–855). Oxford: Oxford University Press.

Drieghe, D., Brysbaert, M., & Desmet, T. (2005). Parafoveal-on-foveal effects on eye movements in text reading: does an extra space make a difference? *Vision Research*, *45*(13), 1693–1706. https://doi.org/10.1016/j.visres.2005.01.010

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

Ehri, L. C. (2005). Development of sight word reading: Phases and findings. In M. J. Snowling & C. Hulmes (Eds.), *The Science of reading: A handbook* (pp. 135–154). Blackwell Publishing Ltd.

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

Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, *2*, 101–118.

https://doi.org/10.1080/13506289508401726

Everatt, J., Warner, J., Miles, T. R., & Thomson, M. E. (1997). The incidence of Stroop interference in dyslexia. *Dyslexia*, *3*(4), 222–228. https://doi.org/10.1002/(sici)1099-0909(199712)3:4<222::aid-dys12>3.0.co;2-p

Faccioli, C., Peru, A., Rubini, E., & Tassinari, G. (2008). Poor readers but compelled to read: Stroop effects in developmental dyslexia. *Child Neuropsychology*, *14*(3), 277–283. https://doi.org/10.1080/09297040701290040

Fischer, R., & Plessow, F. (2015). Efficient multitasking: Parallel versus serial processing of multiple tasks. *Frontiers in Psychology*, *6*, 1366. https://doi.org/10.3389/fpsyg.2015.01366

Frömer, R., Lin, H., Dean Wolf, C. K., Inzlicht, M., & Shenhav, A. (2021). Expectations of reward and efficacy guide cognitive control allocation. *Nature Communications*, *12*(1), 1030. https://doi.org/10.1038/s41467-021-21315-z

Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: analysis and correction. *Behavior Research Methods*, *43*(4), 1171–1181. https://doi.org/10.3758/s13428-011-0109-5

Gatti, S. V., & Egeth, H. E. (1978). Failure of spatial selectivity in vision. *Bulletin of the Psychonomic Society, 11*(3), 181–184. https://doi.org/10.3758/BF03336801

Georgiou, G. & Parrila, R. (2012). Rapid naming and reading. In L. Swanson, K. Harris & S. Graham (Eds.), *Handbook of learning disabilities* (2nd ed., pp. 169– 185). New York: Guilford.

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

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

Grant, M. J. & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, *26*, 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480–506. https://doi.org/10.1037//0096-3445.121.4.480

Hasshim, N., & Parris, B. A. (2015). Assessing stimulus-stimulus (semantic) conflict in the Stroop task using saccadic two-to-one color response mapping and preresponse pupillary measures. *Attention, Perception, & Psychophysics*, *77*(8),2601–2610. https://doi.org/10.3758/s13414-015-0971-9

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Helland, T., & Asbjørnsen, A. (2000). Executive functions in dyslexia. *Child Neuropsychology*, *6*(1), 37–48.

https://doi.org/10.1076/0929-7049(200003)6:1;1-b;ft037

Helland, T., & Morken, F. (2016). Neurocognitive development and predictors of L1 and L2 literacy skills in dyslexia: A longitudinal study of children 5–11 years old. *Dyslexia*, *22*(1), 3–26. https://doi.org/10.1002/dys.1515

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 417–429. https://doi.org/10.1037//0278-7393.16.3.417

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

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: An Interdisciplinary Journal, 31*(7), 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

Hsu, S. H., & Huang, K. C. (2000). Interword spacing in Chinese text layout. *Perceptual and Motor Skills*, *91*(2), 355–365. https://doi.org/10.2466/pms.2000.91.2.355

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

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2009). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 25*(1), 4–32. https://doi.org/10.1080/10573560802491208

Hutzler, F., Fuchs, I., Gagl, B., Schuster, S., Richlan, F., Braun, M., & Hawelka, S. (2013). Parafoveal X-masks interfere with foveal word recognition: Evidence from fixation-related brain potentials. *Frontiers in Systems Neuroscience*, *7*, 33. https://doi.org/10.3389/fnsys.2013.00033

Hutzler, F., Schuster, S., Marx, C., & Hawelka, S. (2019). An investigation of parafoveal masks with the incremental boundary paradigm. *PLoS One*, *14*(2), e0203013. https://doi.org/10.1371/journal.pone.0203013

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, *23*(7), 1587–1596. https://doi.org/10.1162/jocn.2010.21548

Kahneman, D, & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Erlbaum.

Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology. Human Perception and Performance*, *9*(4), 497–509. https://doi.org/10.1037//0096-1523.9.4.497

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

Kelly, M. S., Best, C. T., & Kirk, U. (1989). Cognitive processing deficits in reading disabilities: A prefrontal cortical hypothesis. *Brain and Cognition*, *11*(2), 275–293. https://doi.org/10.1016/0278-2626(89)90022-5

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*(2), 153–168. https://doi.org/10.1016/j.visres.2004.07.037

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

Kliegl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2013). How preview space/time translates into preview cost/benefit for fixation durations during reading. *Quarterly Journal of Experimental Psychology*, *66*(3), 581–600. https://doi.org/10.1080/17470218.2012.658073

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

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

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, *45*(2), 230–251. https://doi.org/10.1598/RRQ.45.2.4

Kuperman, V., Van Dyke, J. A. and 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

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing*, *12*(1), 13–21. https://doi.org/10.1007/s10339-010-0370-z

Lakens, D. (2022). *Sample size justification*. Collabra: Psychology.

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 13*(2), 123–146. https://doi.org/10.1080/1057356970130203

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*(2), 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.

Markwell, E. L., Feigl, B., & Zele, A. J. (2010). Intrinsically photosensitive melanopsin retinal ganglion cell contributions to the pupillary light reflex and circadian rhythm. *Clinical and Experimental Optometry*, *93*(3), 137–149. https://doi.org/10.1111/j.1444-0938.2010.00479.x

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1), 16. https://doi.org/10.5334/joc.18

Mathôt S. (2020). Tuning the senses: How the pupil shapes vision at the earliest stage. *Annual Review of Vision Science*, *6*(1), 433–451. https://doi.org/10.1146/annurev-vision-030320-062352

Mathôt, S., & Vilotijević , A. (2022). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods.* https://doi.org/10.3758/s13428-022-01957-7

McCown, D. A. & Arnoult, M.D. (1981). Interference produced by modified Stroop stimuli. *Bulletin of Psychonomic Society*, *17*, 5–7. https://doi.org/10.3758/BF03333649

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

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

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326.

https://doi.org/10.1037/0033-2909.132.2.297

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, June). *A computational model of control allocation based on the expected value of control* [Paper presentation]. 2nd Multidisciplinary Conference on Reinforcement Learning and Decision Making, Edmonton, Canada.

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

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, *63*(1), 427–452. https://doi.org/10.1146/annurev-psych-120710-100431

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, *5*, 962. https://doi.org/10.3389/fpsyg.2014.00962

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

Parilla, R., & Protopapas, A. (2017). Dyslexia and word reading problems. In K. Cain, D. Compton, & R. Parrila (Eds.), *Theories of reading development* (pp. 333–358). John Benjamins.

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

Perea, M., & Gomez, P. (2012). Increasing interletter spacing facilitates encoding of words. *Psychonomic Bulletin and Review*, *19*(2), 332–338. https://doi.org/10.3758/s13423-011-0214-6

Perea, M., Panadero, V., Moret-Tatay, C., & Gómez, P. (2012). The effects of inter-letter spacing in visual-word recognition: Evidence with young normal readers and developmental dyslexics. *Learning and Instruction, 22*(6), 420–430. https://doi.org/10.1016/j.learninstruc.2012.04.001

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

Protopapas, A., Altani, A., & Georgiou, G. K. (2013). Development of serial processing in reading and rapid naming. *Journal of Experimental Child Psychology*, *116*(4), 914–929. https://doi.org/10.1016/j.jecp.2013.08.004

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology*, *54*(3), 251–282. https://doi.org/10.1016/j.cogpsych.2006.07.003

Protopapas, A., Katopodi, K., Altani, A., & Georgiou, G. K. (2018). Word reading fluency as a serial naming task. *Scientific Studies of Reading*, *22*(3), 248–263. https://doi.org/10.1080/10888438.2018.1430804

Protopapas, A., Vlahou, E. L., Moirou, D., & Ziaka, L. (2014). Word reading practice reduces Stroop interference in children. *Acta Psychologica*, *148*, 204–208. https://doi.org/10.1016/j.actpsy.2014.02.006

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. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*(1), 65–81. https://doi.org/10.1016/0010-0285(75)90005-5

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. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8),1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., & Juhasz, B. J. (2004). Eye movements in reading: Old questions andnew directions. *European Journal of Cognitive Psychology, 16*(1-2), 340–352. https://doi.org/10.1080/09541440340000385

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

Rayner, K., Yang, J., Schuett, S., & Slattery, T. J. (2013). Eye movements of older and younger readers when reading unspaced text. *Experimental Psychology*, *60*(5), 354–361. https://doi.org/10.1027/1618-3169/a000207

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*(2), 275–280. https://doi.org/10.1037/h0027768

Reiter, A., Tucha, O., & Lange, K. W. (2005). Executive functions in children with dyslexia. *Dyslexia*, *11*(2), 116–131. https://doi.org/10.1002/dys.289

Ridderinkhof, K. R., Wylie, S. A., van den Wildenberg, W., Bashore, T. R., Jr, & van der Molen, M. W. (2021). The arrow of time: Advancing insights into action control from the arrow version of the Eriksen flanker task. *Attention, Perception, & Psychophysics*, *83*(2), 700–721. https://doi.org/10.3758/s13414-020-02167-z

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. (2006). Functional architecture of naming dice, digits, and number words. *Language and Cognitive Processes*, *21*(1-3), 78–111. https://doi.org/10.1080/01690960400001846

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

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467.

https://doi.org/10.3758/s13423-018-1558-y

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

Salthouse T. A. (2010). Is flanker-based inhibition related to age? Identifying specific influences of individual differences on neurocognitive variables. *Brain and Cognition*, *73*(1), 51–61. https://doi.org/10.1016/j.bandc.2010.02.003

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

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing inreading. *Attention, Perception, & Psychophysics*, *74*(1), 5–35. https://doi.org/10.3758/s13414-011-0219-2

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

Shen, D., Liversedge, S. P., Tian, J., Zang, C., Cui, L., Bai, X., Yan, G., & Rayner, K. (2012). Eye movements of second language learners when reading spaced and unspaced Chinese text. *Journal of Experimental Psychology: Applied*, *18*(2), 192–202. https://doi.org/10.1037/a0027485

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

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127–190.

https://doi.org/10.1037/0033-295X.84.2.127

Silva, S., Reis, A., Casaca, L., Petersson, K. M., & Faísca, L. (2016). When the eyes no longer lead: Familiarity and length effects on eye-voice span. *Frontiers in Psychology*, *7*, 1720. https://doi.org/10.3389/fpsyg.2016.01720

Slattery, T. J., & Rayner, K. (2013). Effects of intraword and interword spacing on eye movements during reading: Exploring the optimal use of space in a line of text. *Attention, Perception, & Psychophysics*, *75*(6), 1275–1292. https://doi.org/10.3758/s13414-013-0463-8

Slattery, T. J., Yates, M., & Angele, B. (2016). Interword and interletter spacing effects during reading revisited: Interactions with word and font characteristics. *Journal of Experimental Psychology: Applied, 22*(4), 406–422. https://doi.org/10.1037/xap0000104

Snell, J., Cauchi, C., Grainger, J., & Lété, B. (2021). Attention extends beyond single words in beginning readers. *Attention, Perception, & Psychophysics*, *83*(1), 238–246. https://doi.org/10.3758/s13414-020-02184-y

Snell, J., & Grainger, J. (2018). Parallel word processing in the flanker paradigm has a rightward bias. *Attention, Perception, & Psychophysics*, *80*(6), 1512–1519. https://doi.org/10.3758/s13414-018-1547-2

Spinelli, D., De Luca, M., Di Filippo, G., Mancini, M., Martelli, M., & Zoccolotti, P. (2005). Length effect in word naming in reading: Role of reading experience and reading deficit in Italian readers. *Developmental Neuropsychology*, *27*(2), 217–235. https://doi.org/10.1207/s15326942dn2702_2

Spinelli, D., De Luca, M., Judica, A., & Zoccolotti, P. (2002). Crowding effects on word identification in developmental dyslexia. *Cortex*, *38*(2), 179–200. https://doi.org/10.1016/s0010-9452(08)70649-x

Spragins, A. B., Lefton, L. A., & Fisher, D. F. (1976). Eye movements while reading and searching spatially transformed text: A developmental examination. *Memory & Cognition*, *4*(1), 36–42. https://doi.org/10.3758/BF03213252

Stanovich, K. E., Feeman, D. J., & Cunningham, A. E. (1983). The development of the relation between letter-naming speed and reading ability. *Bulletin of the Psychonomic Society*, *21*(3), 199–202. https://doi.org/10.3758/bf03334686

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.

Stringer, R. W., Toplak, M. E., & Stanovich, K. E. (2004). Differential relationships between RAN performance, behaviour ratings, and executive function measures: Searching for a double dissociation. *Reading and Writing*, *17*(9), 891–914. https://doi.org/10.1007/s11145-004-2770-x

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

Vakil, E., Weisz, H., Jedwab, L., Groswasser, Z., & Aberbuch, S. (1995). Stroop color-word task as a measure of selective attention: Efficiency in closed-head-injured patients. *Journal of Clinical and Experimental Neuropsychology*, *17*(3),335–342. https://doi.org/10.1080/01688639508405127

van der Schoot, M., Licht, R., Horsley, T. M., & de Sergeant, V. R. (2000). Inhibitory deficits in reading disability depend on subtype: Guessers but not spellers. *Child Neuropsychology*, *6*(4), 297–312. https://doi.org/10.1076/chin.6.4.297.3139

Vasilev, M. R., & Angele, B. (2017). Parafoveal preview effects from word N + 1 and word N + 2 during reading: A critical review and Bayesian meta-analysis.

Psychonomic Bulletin & Review, *24*(3), 666–689. https://doi.org/10.3758/s13423-016-1147-x

Vasilev, M. R., Yates, M., Prueitt, E., & Slattery, T. J. (2021). Parafoveal degradation during reading reduces preview costs only when it is not perceptually distinct. *Quarterly Journal of Experimental Psychology*, *74*(2), 254–276. https://doi.org/10.1177/1747021820959661

Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *Quarterly Journal of Experimental Psychology*, *67*(4), 703–727. https://doi.org/10.1080/17470218.2013.826258

Veldre, A., & Andrews, S. (2015a). Parafoveal lexical activation depends on skilled reading proficiency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 586–595. https://doi.org/10.1037/xlm0000039

Veldre, A., & Andrews, S. (2015b). Parafoveal preview benefit is modulated by the precision of skilled readers' lexical representations. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 219–232. https://doi.org/10.1037/xhp0000017

Veldre, A., & Andrews, S. (2018). How does foveal processing difficulty affect parafoveal processing during reading? *Journal of Memory and Language*, *103*,74–90. https://doi. org/10.1016/j.jml.2018.08.001

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

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*(1), 59–85. https://doi.org/10.1016/0010-0285(70)90005-8

White, R. W. & Kelly, D. (2006). A study on the effects of personalization and task information on implicit feedback performance. *Proceedings of the 15th ACM international conference on Information and knowledge management* (pp. 297–306). Arlington, Virginia, USA. https://doi.org/10.1145/1183614.1183659

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

Winskel, H., Radach, R., & Luksaneeyanawin, S. (2009). Eye movements when reading spaced and unspaced Thai and English: A comparison of Thai-English bilinguals and English monolinguals. *Journal of Memory and Language*, *61*(3), 339–351. https://doi.org/10.1016/j.jml.2009.07.002

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*(3), 415–438. https://doi.org/10.1037/0022-0663.91.3.415

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading, 5*(3), 211–239. https://doi.org/10.1207/S1532799XSSR0503_2

Woods, R. J., Davis, K., & Scharff, L. F. V. (2005). Effects of typeface and font size on legibility for children. *American Journal of Psychological Research*, *1*, 86e102.

Yan, M., Pan, J., Laubrock, J., Kliegl, R., & Shu, H. (2013). Parafoveal processing efficiency in rapid automatized naming: A comparison between Chinese normal and dyslexic children. *Journal of Experimental Child Psychology*, *115*(3), 579–589. https://doi.org/10.1016/j.jecp.2013.01.007

Ziaka, L. (2014). *Study of intratask components of the Stroop task: Column time and pause time*. [Master's thesis, Athens University]. Pergamos. https://pergamos.lib.uoa.gr/uoa/dl/frontend/file/lib/default/data/1318876/theFile

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

PART II
THE ARTICLES

Article I

Ziaka, L., & Protopapas, A. (2022). Conflict monitoring or multi-tasking?

Tracking within-task performance in single-item and multi-item Stroop tasks.

Acta Psychologica, *226*, 103583.

https://doi.org/10.1016/j.actpsy.2022.103583

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
Available online 2 April 2022
0001-6918/

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.

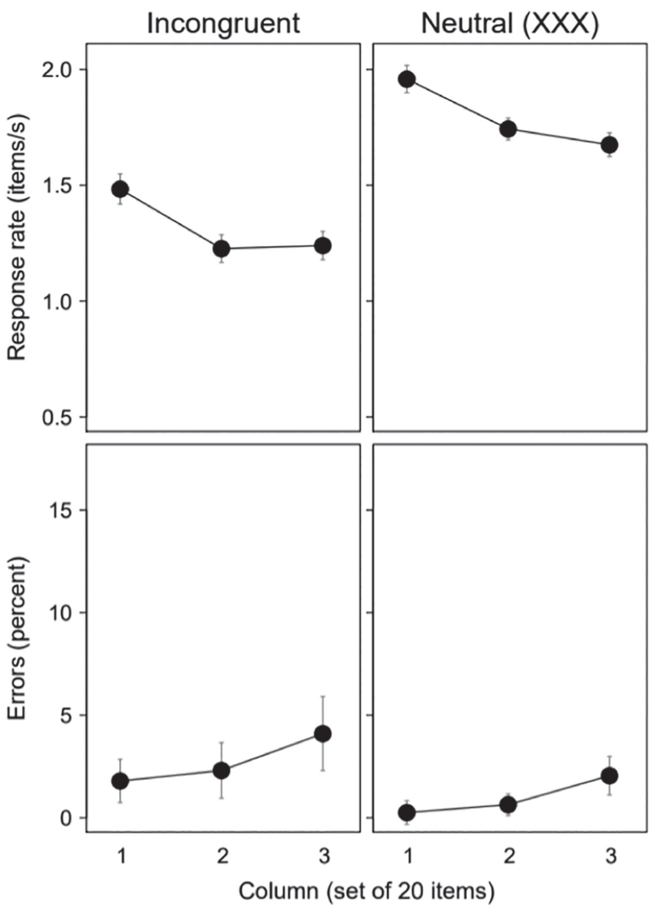

Fig. 1. Results of Experiment 1. Response rate (items per second) and accuracy (percent errors) in each column and condition. Error bars show within-participant 95% confidence intervals.

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, $F(1, 38) = 182.21$, $\eta^2 = 0.482$, $p < .001$.

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 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

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 (X, B, 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.

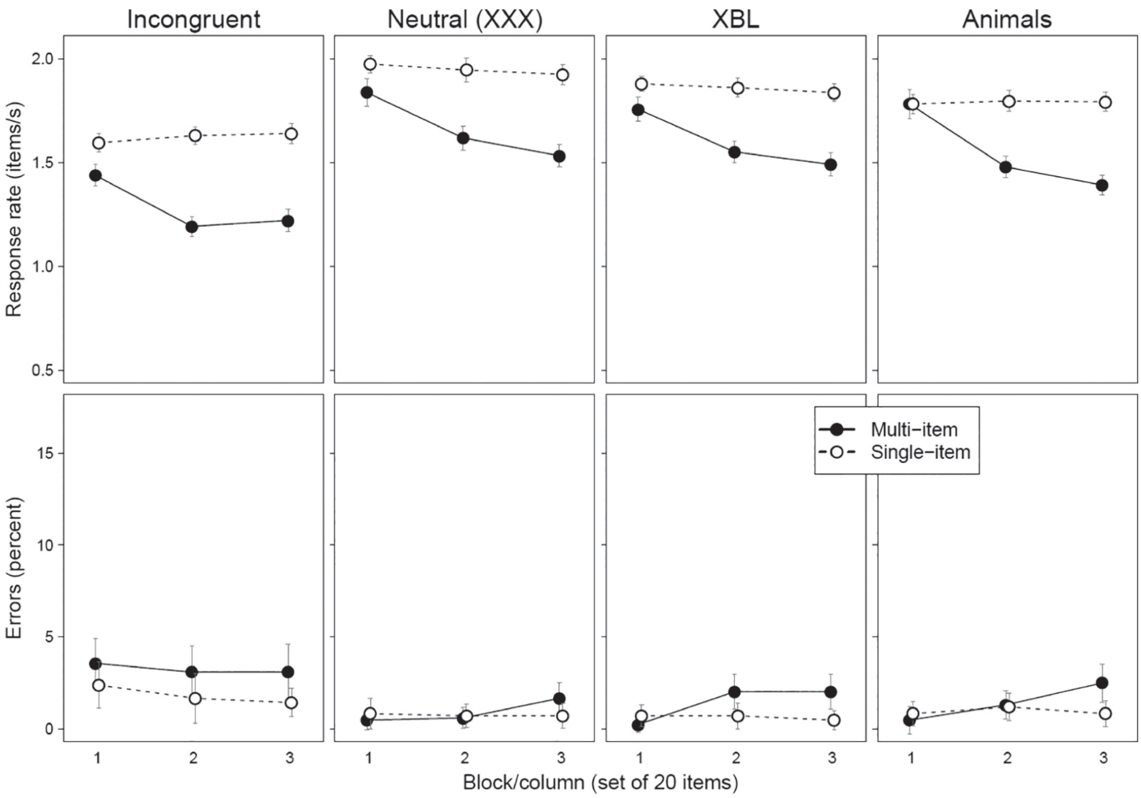

Fig. 2. Results of Experiment 2 (adults). Response rate (items per second) and accuracy (percentage of errors) in each block/column, condition, and task. Error bars show within-participant 95% confidence intervals.

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

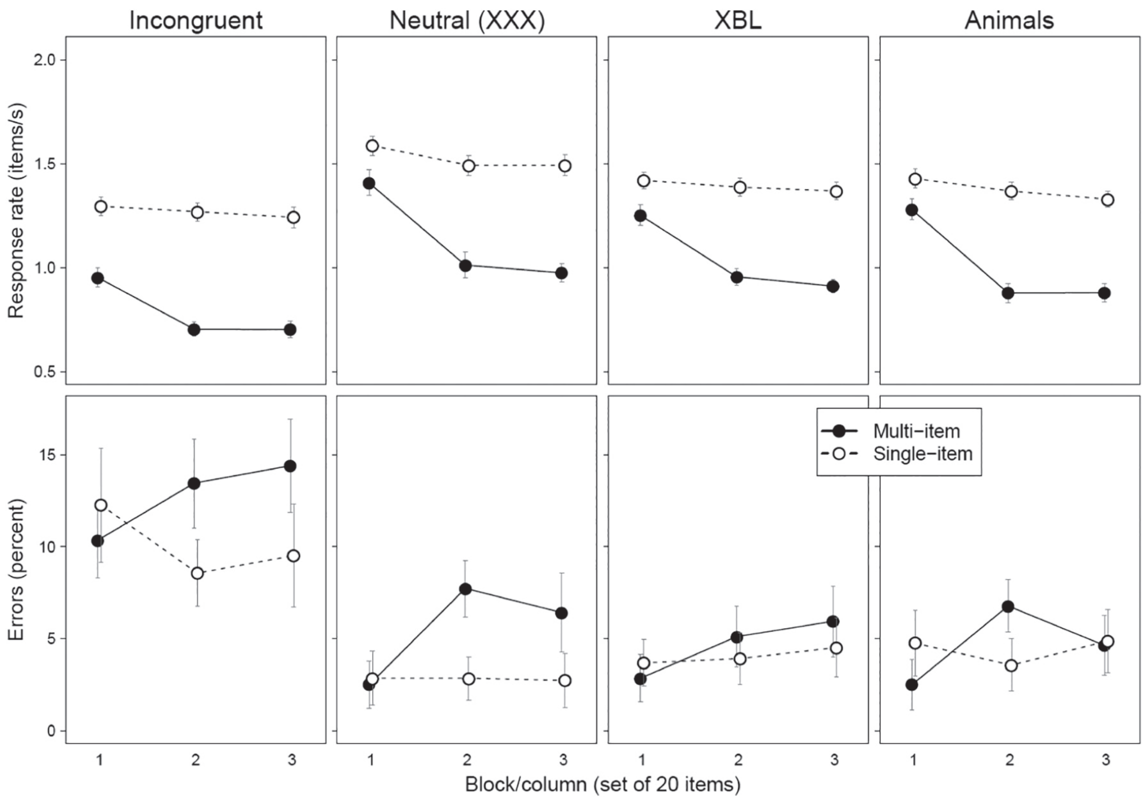

Fig. 3. Results of Experiment 3 (children). Response rate (items per second) and accuracy (percentage of errors) in each block/column, condition, and task. Error bars show within-participant 95% confidence intervals.

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

Article II

Ziaka, L., & Protopapas, A. (under review). Cognitive control beyond single-item tasks: Insights from pupillometry, gaze, and behavioral measures. *Journal of Experimental Psychology: Human Perception and Performance*.

Article III

Ziaka, L., Skoteinou, D., Protopapas, A. (2022). Task format modulates the relationship between reading ability and Stroop interference. *Journal of Experimental Psychology: Human Perception and Performance, 48(4),* 275–288. https://doi.org/10.1037/xhp0000964

Errata

Name of Candidate: Laoura Ziaka

Titel of Thesis: Interference and cognitive control dynamics in the course of serial naming tasks

Abbreviations for different types of corrections:

Cor – correction

Cpltf – change of page layout or text format

Page(s)	Line	Original Text	Type of correction	Corrected text
VII	4	in single-items and multi-items Stroop tasks	Cor	in single-item and multi-item Stroop tasks
VIII		Table of contents	Cpltf	Table of Contents updated
		Headings and subheadings	Cpltf	Headings and subheadings re-checked and corrected
13	8	Altani et al., 2017	Cor	Altani, Georgiou, et al., 2017
16	5	whole task (Rouder, 2019)	Cor	whole task (Rouder & Haaf, 2019)

18	11	a goal (Moors & De Houwer)	Cor	a goal (Moors & De Houwer, 2006)
18	19	(Altani, Georgiou, & Protopapas, 2017	Cor	(Altani, Protopapas, & Georgiou, 2017
26	16	“eye-voice span”; Gordon & Hoedemaker, 2016; Huang, 2018; Silva et al., 2016; Pan et al., 2013).	Cor	“eye-voice span”; Gordon & Hoedemaker, 2016; Huang, 2018; Pan et al., 2013; Silva et al., 2016)
26	24	with proper expression (Wolf & Katzir-Cohen, 2001; Hudson et al, 2009; Kuhn et al, 2010).	Cor	with proper expression (Hudson et al., 2009; Kuhn et al., 2010; Wolf & Katzir-Cohen, 2001).
29	10	Ludwig, 2010; Vakil et al., 1995), whereas less	Cor	Ludwig et al., 2010; Vakil et al., 1995), whereas less
30	1	Kelly et al., 1989; Reiter et al., 2005; Kapoula et al. 2010; Protopapas et al., 2007; Wang & Gathercole, 2015).	Cor	Kapoula et al. 2010; Kelly et al., 1989; Protopapas et al., 2007; Reiter et al., 2005; Wang & Gathercole, 2015).

30	3 & 4	inhibition impairments (van der Schoot et al., 2000; Reiter et al., 2005;	Cor	inhibition impairments (Reiter et al., 2005; van der Schoot et al., 2000;
30	17	the studies reviewed in sections 2.2.1 and 2.2.3	Cor	the studies reviewed in sections 2.2.1 and 2.2.2
32	20	Snell & Grainger 2018	Cor	Snell & Grainger, 2018
33	2	Snell & Grainger 2018	Cor	Snell & Grainger, 2018
33	8	(Draheim et al. 2021)	Cor	(Draheim et al., 2021)
33	12	the study of Snell & Grainger (2018)	Cor	the study of Snell and Grainger (2018)
33	25	(i.e., the foveal word; Rayner, 2015)	Cor	(i.e., the foveal word; Rayner et al., 2005)
34	11	(Rayner 1998, 2009, 2015).	Cor	(Rayner, 1998, 2009; Rayner et al., 2005).
34	13	perceptual span (Rayner, 2015).	Cor	perceptual span (Rayner et al., 2005).

35	25	the meta-analysis of Vasilev & Angele (2017)	Cor	the meta-analysis of Vasilev and Angele (2017)
37	22	indicating that facilitation due to increased spacing is language specific.	Cor	indicating that facilitation due to increased spacing is specific to linguistic materials.
43	24	Although Shenav et al.	Cor	Although Shenhav et al.
46		Conflict Monitoring or Multi-tasking? Tracking Within-Task Performance in Single-item and Multi-item Stroop Tasks	Cor	Conflict monitoring or multi-tasking? Tracking within-task performance in single-item and multi-item Stroop tasks
46		Cognitive Control Beyond Single-Item Tasks: Insights from Pupillometry, Eye-Tracking, and Behavioral Measures	Cor	Cognitive control beyond single-item tasks: Insights from pupillometry, gaze, and behavioral measures

47	5	while adopting a within-participant permitted	Cor	while adopting a within-participant design permitted
49	2 & 9	(Markwell, 2010;	Cor	(Markwell et al., 2010;
51	11	I provided the children with sufficiently information	Cor	I provided the children with sufficient information
52	11	in single-items and multi-items	Cor	in single-item and multi-item
58	25	Exploration is considered to emerge	Cor	Exploitation is considered to emerge
65	18	inhibition (Björngrim et al., 2019; Bezdicek et al., 2015;		inhibition (Bezdicek et al., 2015; Björngrim et al., 2019;
68	5	The Gratton effect could not be part of the present dissertation because	Cor	The Gratton effect could not have played a role in the present study because
71–91		Reference section	Cpltf	Reference section re-checked and corrected