

DR. LUDOVIC FERRAND (Orcid ID : 0000-0003-1631-7849)

Article type : Paper

## **Stroop Interference is a Composite Phenomenon: Evidence from Distinct Developmental Trajectories of its Components**

Ludovic Ferrand<sup>1\*</sup>, Stéphanie Ducrot<sup>2</sup>, Pierre Chausse<sup>1</sup>, Norbert Maïonchi-Pino<sup>1</sup>, Richard J. O'Connor<sup>3</sup>, Benjamin A. Parris<sup>4</sup>, Patrick Perret<sup>5</sup>, Kevin J. Riggs<sup>3</sup>, and Maria Augustinova<sup>6\*</sup>

<sup>1</sup> *Université Clermont Auvergne, CNRS, LAPSCO, Clermont-Ferrand, France*

<sup>2</sup> *Aix-Marseille Univ, CNRS, LPL, Aix-en-Provence, France*

<sup>3</sup> *Department of Psychology, University of Hull, United-Kingdom*

<sup>4</sup> *Faculty of Science and Technology, Bournemouth University, Poole, United-Kingdom*

<sup>5</sup> *Aix-Marseille Univ, PSYCLE, Aix-en-Provence, France*

<sup>6</sup> *Normandie Université, UNIROUEN, CRFDP, Rouen, France*

\*Contributed equally. The first versus the last authorship was determined by a coin toss.

**Running title:** Stroop Interference is a Composite Phenomenon

**Correspondance:** Ludovic Ferrand, Université Clermont Auvergne, LAPSCO CNRS UMR 6024, 34 avenue Carnot, 63000 Clermont-Ferrand, France ; ludovic.ferrand@uca.fr ; Maria Augustinova, Normandie Université, UNIROUEN, CRFDP, 1 rue Thomas Becket, 76821 Mont Saint-Aignan, France ; maria.augustinova@univ-rouen.fr

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/desc.12899

This article is protected by copyright. All rights reserved.

**Conflict of Interest Statement.** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Acknowledgments.** We are grateful to Mr. Pierre Blache, Inspector of French National Education-Garlaban (Aix-Marseille Academy, France) and Mrs. Marie Bonicel, head of Sainte-Thècle school in Chamalières (France) for their agreement to run and their help to set up the reported experiments – thanks that are extended to all the teachers in participating classes. We would also like to thank Christine Cardot-Baroni, Hugo Julhe, Jordan Chambas, Jeanne-Tiphaine Iltis and Mathilde Cayre for their help in running Experiment 2, as well as Julien Guégan and Mickaël Berthon from the LAPSCO and CATECH (CNRS, Université Clermont Auvergne, France) for their invaluable help at the different stages of the project. Finally, the French part of the byline would like to thank the French football (aka soccer) team for the second star on the national team's jersey.

### **Research Highlights**

- The present study shows that early developmental change in Stroop interference persists even after controlling for existing differences in processing speed.
- This change corresponds to an increase such that 3<sup>rd</sup>- and 5<sup>th</sup>-graders displayed greater Stroop interference compared to 1<sup>st</sup>-graders.
- This increase results from joint contributions of task, semantic and response conflicts to Stroop interference in 3<sup>rd</sup>- and 5<sup>th</sup>-graders (vs. task conflict only in 1<sup>st</sup>-graders).
- The specific developmental trajectories of these different conflicts provide unique evidence for the composite (as opposed to unitary) nature of Stroop interference.

## Abstract

Only one previous developmental study of Stroop task performance (Schiller, 1966) has controlled for differences in processing speed that exist both within and between age-groups. Therefore, the question of whether the early developmental change in the magnitude of Stroop interference actually persists after controlling for processing speed needs further investigation; work that is further motivated by the possibility that any remaining differences would be caused by process(es) other than processing speed. Analysis of data from two experiments revealed that, even after controlling for processing speed using z-transformed reaction times, early developmental change persists such that the magnitude of overall Stroop interference is larger in 3<sup>rd</sup>- and 5<sup>th</sup>- graders as compared to 1<sup>st</sup>-graders. This pattern indicates that the magnitude of overall Stroop interference peaks after two or three years of reading practice (Schadler & Thissen, 1981). Furthermore, this peak is shown to be due to distinct components of Stroop interference (resulting from specific conflicts) progressively falling into place. Experiment 2 revealed that the change in the magnitude of Stroop interference specifically results from joint contributions of task, semantic and response conflicts in 3<sup>rd</sup>- and 5<sup>th</sup>-graders as compared to a sole contribution of task conflict in 1<sup>st</sup>-graders. The specific developmental trajectories of different conflicts presented in the present work provide unique evidence for multiple loci of Stroop interference in the processing stream (respectively task, semantic and response conflict) as opposed to a single (i.e., response) locus predicted by historically-favored response competition accounts.

**Key words:** Stroop interference, Task conflict, Semantic conflict, Response conflict, Developmental trajectories

## 1. INTRODUCTION

The Stroop task (Stroop, 1935) requires individuals to identify, as quickly and accurately as possible, the font color of written words without reading them. Despite this requirement, the typical result is that individuals' identification times are longer and more error-prone for color-incongruent Stroop words (i.e., words displayed in a color that is different from the one they designate such as “BLUE” displayed in green ink; hereafter  $BLUE_{green}$ ), than for color-neutral items (e.g., “DOG”/ “XXX” displayed in green ink,  $DOG/XXX_{green}$ ).

Numerous studies have amply documented that the magnitude of this latter difference – called Stroop interference (e.g.,  $BLUE_{green}-DOG/XXX_{green}$ ) – changes during the course of early development (e.g., Aïte, Cassotti, Linzarini, Osmont, Houdé, & Borst, 2018; Grégoire, Perruchet, & Poulin-Charronnat, 2015; Megherbi, Elbro, Oakhill, Segui, & New, 2018; Wright, 2017). Yet, this early change, as well as the underlying processing, still remain to be elucidated. Therefore, the present study attempted to shed a more direct light on these issues.

### 1.1. Developmental change in Stroop Interference

In their cross-sectional study, Comalli, Wapner, and Werner (1962) were the first to report the change in the magnitude of Stroop interference across the life-span. More specifically, they reported that this magnitude consistently decreases during the course of early development (from 7 to 19 years) and that – after a period of stabilization during early and middle adulthood (from 19 to 44 years) – it increases during late adulthood (from 65 to 80 years). Since this seminal study, developmental studies of Stroop interference continued to flourish, albeit with a specific focus on either early (childhood and adolescence) or late (adulthood and aging) developmental change. Moreover, these lines of research have evolved in separation.

Interestingly, for a period of time, the importance of the late developmental change in Stroop interference had been called into question (e.g., Verhaeghen & De Meersman, 1998; see also Verhaeghen, 2011). Indeed, age-related differences in the magnitude of Stroop interference, where they had been seen at all, were merely attributed to a decrease in processing speed that is strongly associated with normal aging (e.g., Salthouse & Meinz, 1995).

Only more recent studies have demonstrated that age-related differences in Stroop interference actually persist even after controlling for general slowing (e.g., Augustinova, Clarys, Spatola, & Ferrand, 2018; Bugg, DeLosh, Davalos, & Davis, 2007; Jackson & Balota, 2013; Spieler, Balota, & Faust, 1996; Wolf et al., 2014). Given this control, these studies convincingly showed that the generalized slowing account (e.g., Myerson, Hale, Wagstaff, Poon, & Smith, 1990; Salthouse, 1996) is not sufficient to fully explain the age-related differences in the magnitude of Stroop interference in later life. And thus, they reinstated the relevance of a popular account of these differences in terms of the decline in inhibitory control (Hasher & Zacks, 1988). It should be remembered that in the Stroop task, the irrelevant word-dimension of Stroop words (i.e., “blue” for *BLUE*<sub>green</sub>) needs to be inhibited to allow individuals to generate the correct response based on the relevant color dimension of Stroop words (e.g., “green” for *BLUE*<sub>green</sub>). Consequently, a higher level of Stroop interference is often equated with lower inhibitory control (e.g., Miyake et al., 2000).

The inhibition-based account is also dominant in early developmental studies. Circuits in the prefrontal and anterior cingulate cortex continue to mature until the late twenties causing the continuous improvement of the executive functions – including inhibitory control – that are based on these circuits (e.g., Davidson, Amso, Anderson, & Diamond, 2006; Luna, et al., 2001; Luna, Garver, Urban, Lazar, & Sweeney, 2004; Munoz, Broughton, Goldring, & Armstrong, 1998; Prencipe et al., 2011; for reviews see Diamond, 2002; Diamond, 2013). As

a result, in developmental neuropsychological research and practice, inhibitory abilities are thought to present a specific developmental trajectory that is clearly captured by changes in the magnitude of Stroop interference during childhood and adolescence (e.g., Armengol, 2002; Roy et al., 2017; see also e.g., Aïte et al., 2018; Peru, Faccioli, & Tassinari, 2006; see also here below).

However, exactly like during the course of late-development discussed above, the early developmental trajectory of inhibitory abilities and the one of processing speed overlap (i.e., both increase during the course of early development; e.g., Kail, 1991, 2007; Fry & Hale, 1996; Nettelbeck & Burns, 2010) – making it difficult to disentangle the specific contribution of these processes to the magnitude of Stroop interference. Despite this latter fact, only one early developmental study actually controlled for processing speed (Schiller, 1966). Therefore, the aim of Experiment 1 was to examine whether, and the extent to which, the early developmental change in the magnitude of Stroop interference still occurs when age-related processing speed differences are accounted for (see e.g., Lété & Fayol, 2013; Ziegler, Lété, Bertrand, & Grainger, 2014 for this type of endeavor in children’s studies). If this is the case, then a remaining question is which process or processes underlie this change; a question we aimed to address in Experiment 2.

## **1.2. Processes underlying the early developmental change in Stroop interference**

As already mentioned, early developmental change in the magnitude of Stroop interference is often thought to reflect improved inhibitory control during childhood and adolescence. Results of several studies – including the seminal study of Comalli and colleagues (1962) – are consistent with an inhibition-based account (see above). Indeed, they reported the highest Stroop interference in seven-years-olds (i.e., the youngest age-group

investigated), the magnitude of which continued to decrease until the age of nineteen (see also e.g., Armengol, 2002; Peru, Faccioli, & Tassinari, 2006; Roy et al., 2017).

However, it is not the inhibition-based account that is mobilized by Comalli and colleagues (1962) to interpret their results. They argue that there is actually a one-to-one relationship between the magnitude of Stroop interference that children display and their “(...) capacity to maintain a course of action in the face of intrusion by other stimuli” (p.47). Said differently, these authors were the first to conclude that the modification of Stroop interference during childhood is clearly determined by what is now termed as the capacity of goal – or task-set – maintenance (see e.g., Chevalier & Blaye, 2008; see also De Jong, Berendsen, & Cools, 1999; Parris, Bate, Brown, Hodgson, 2012). The distinction between goal maintenance and inhibitory capacity (e.g. De Jong et al., 1999) reflects the debate about whether selective attention operates by enhancing the processing of the relevant dimension or enhancing inhibitory control of the irrelevant dimension (e.g., Egnér & Hirsch, 2006).

Finally, a quick inspection of other early developmental Stroop studies indicates that in some quarters (e.g., developmental psycholinguistics), the early developmental change in Stroop interference is thought to index the automatization of word-recognition and/or reading skills in general (e.g., Ehri, & Wilce, 1983; Stanovich, Cunningham, & West, 1981). In contrast to studies above emphasizing a continuous decrease in the magnitude of Stroop interference across childhood (e.g., Comalli et al., 1962; Roy et al., 2017), such a conceptualization predicts an initial increase in Stroop interference when children are first learning to read. Several past findings – including the one that controlled for age-groups differences in processing speed (Schiller, 1966) – are in line with this conceptualization. A robust Stroop interference is indeed detected in 1<sup>st</sup>-graders (6-7 years old) after only five months of reading instruction (Stanovich et al., 1981) and its magnitude has been reported to peak after two or three years of reading practice (Schiller, 1966; Schadler & Thissen, 1981).

Consistent with this latter finding, Rand, Wapner, Werner, and MacFarland (1963) found the greatest Stroop interference in a group of nine-year-olds, the magnitude of which was significantly greater than the one observed in a group of six-years-old pre-readers.

To sum up, the different lines of research presented above clearly diverge on the nature of the process underlying the early developmental change in the magnitude of Stroop interference. For some this change reflects a developmental trajectory of goal maintenance, for others the development of inhibitory control, and others still, the developing automatization of word-reading. However, these approaches to explaining the early developmental change in Stroop interference do all appear to agree on the fact that Stroop interference is a unitary phenomenon, and such is to be explained by a single process.

This latter view contrasts with several other lines of research suggesting that Stroop interference is a more complex phenomenon, potentially influenced by more than one process. To illustrate, Bub, Masson and Lalonde (2006) argue that the magnitude of Stroop interference reflects both children's inability to consistently maintain the task set of color naming and to suppress irrelevant responses stemming from the word-dimension of Stroop words. However, the specific contribution of goal-maintenance vs. inhibition to overall Stroop interference was not measured directly in this study. Megherbi and colleagues (2018), on the other hand, suggest that the magnitude of Stroop interference displayed by children is a result of both the effect of the mandatory decoding of the distracting words and the capacity of inhibition (see also e.g., Peru et al., 2006; Wright & Wanley, 2003). Again, the specific contribution of word-reading vs. inhibition to the overall Stroop interference was not measured directly in this study.

These latter lines of research agree that the processes underlying the early developmental change in the magnitude of Stroop interference are twofold – one of which being the improvement of inhibitory abilities. But they clearly diverge on the nature of the



other process involved. For Megherbi et al. (2018), the change in the magnitude of Stroop interference during childhood also indexes a developmental trajectory of automatization of word-reading, whereas it is goal-maintenance for Bub and colleagues (2006). However, it is critical to note that, just as in studies subscribing to single-process approaches discussed above, these allegedly *multiple* processes were in fact merely inferred from the overall magnitude of Stroop interference (e.g.,  $BLUE_{\text{green}}-DOG/XXX_{\text{green}}$ ).

Therefore, the extent to which single versus multiple processes actually determine the developmental change in the magnitude of Stroop interference and what the nature of this processing is remain an open issue. It is noteworthy that this issue extends beyond developmental science. Indeed, it reflects a long-lasting debate in the adult Stroop literature about the relevance of single – versus multi-stage accounts of the Stroop interference effect, as discussed in the following section.

### **1.3. Single- vs. Multi-stage accounts of Stroop interference**

The aforementioned popular idea that the magnitude of Stroop interference specifically indexes inhibitory abilities (see section 1.1. and 1.2.) is rooted in *single-stage response competition accounts* that have historically been favored in the Stroop literature (Augustinova, Silvert, Spatola, & Ferrand, 2018; Risko, Schmidt, & Besner, 2006). These kinds of single-stage accounts share the aforementioned idea that word-reading – routinized in skilled readers – provides the basis for a response (i.e., “blue” for  $BLUE_{\text{green}}$ ). Because this incorrect response interferes with the one cued by the relevant color dimension of color-incongruent Stroop words (i.e., “green” for  $BLUE_{\text{green}}$ ), it gives rise to so-called *response conflict* – the magnitude of which is determined by the degree of inhibitory control. In this view, Stroop interference is thought to specifically index the magnitude of this latter conflict

– occurring in the processing stream at the level of response output (e.g., Hommel, 1997; LaBerge & Samuels, 1974; Morton & Chambers, 1973; Shiffrin & Schneider, 1977).

These latter accounts prevailed over another class of single-stage accounts – *early-selection accounts*. They share the idea that a single conflict generating Stroop interference – so-called *stimulus conflict* – actually occurs much earlier in processing (as compared to response conflict depicted above). For instance, Seymour (1977) considers that this (early) conflict occurs at conceptual encoding of color-incongruent words (e.g. *BLUE*<sub>green</sub>) because the meaning of the word dimension (i.e., blue for *BLUE*<sub>green</sub>) and that of the color dimension (i.e., green here) both correspond to colors. Indeed, “(...) delays of processing occur whenever distinct semantic codes are simultaneously activated, and that these delays become acute when the conflicting codes are values on a single dimension or closely related dimensions.” (p. 263; see also e.g. Luo, 1999; Scheibe, Shaver, & Carrier, 1967; Seymour, 1974; 1977; Stirling, 1979).

In sum, “The early-selection account focuses on the similarity between the relevant stimulus and the irrelevant stimulus, whereas the late-selection account focuses on the similarity between the irrelevant stimulus and the response. Both similarity relationships are, of course, present in the Stroop task – in fact, they constitute a confounding that makes distinguishing empirically between the two accounts difficult.” (Zhang & Kornblum, 1998, p. 4). It is thus not surprising that the first multi-stage accounts assumed that color-incongruent words (e.g. *BLUE*<sub>green</sub>) generate both stimulus (SC) and response conflicts (RC; hereafter SC-RC accounts, see Augustinova et al., 2018 for this terminology and review of these accounts). Subsequent studies consistently showed that both conflicts contribute to and are therefore confounded in the overall Stroop interference measure (Augustinova & Ferrand, 2014a; De Houwer, 2003; Manwell, Roberts, & Besner, 2004; Schmidt & Cheesman, 2005; Zhang & Kornblum, 1998; see also e.g., Augustinova et al., 2015; Chen, Lei, Ding, Li, & Chen, 2013;

Killikelly & Szücs, 2013; Szücs & Soltész, 2010; Van Veen & Carter, 2005 for electrophysiological and fMRI evidence).

However, several other multi-stage accounts also assume that Stroop interference results from the simultaneous contribution of two distinct conflicts. In addition to *response conflict* as depicted above, they assume the existence of so-called *task conflict* (TC; hereafter TC-RC accounts, see Augustinova et al., 2018; Parris et al., submitted; for reviews) instead of the semantic conflict assumed by the aforementioned SC-RC accounts. Task conflict is thought to arise for all kinds of readable items (including color-congruent words, e.g., *BLUE*<sub>blue</sub>) and is thus independent of the specific color-incongruency conflict occurring for color-incongruent Stroop words (e.g., *BLUE*<sub>green</sub>). This is because the individual's attention is drawn to an irrelevant task (i.e., word-reading) instead of being fully focused on the relevant task (i.e., color-naming), leading to the two task sets competing (e.g., Goldfarb & Henik, 2006, 2007; Kalanthroff, Goldfarb, & Henik, 2013; Kalanthroff, Goldfarb, Usher, & Henik, 2013; Monsell, Taylor & Murphy, 2001; Parris, 2014 for empirical demonstrations; see also e.g., Bench, Frith, Grasby, Friston, Paulesu, Frackowiak, et al., 1993 for fMRI evidence).

To sum up, the SC-RC and TC-RC multi-stage accounts described above anticipate the contribution of three distinct conflicts to overall Stroop interference. The SC-RC accounts emphasize that semantic and response conflicts need to be de-confounded and – as such – assessed in separation whereas the TC-RC accounts express the very same concerns for task and response conflicts. Given that a considerable behavioral, EEG and fMRI evidence points to the viability of both SC-RC and TC-RC multi-stage accounts of Stroop interference (see above), several lines of research highlight the necessity to adopt an integrative perspective that allows for bridging these two perspectives (Augustinova et al., 2018; Parris, Hasshim, Wadsley, Augustinova & Ferrand, submitted; for reviews).

This integrative perspective posits that all three conflicts (i.e., task, semantic and response conflicts) contribute to standard Stroop interference and are thus confounded in its overall magnitude. By providing direct and simultaneous evidence for each of the three conflicts, Augustinova and colleagues (2018) showed that is indeed the case (see section 3 for further details). Therefore, this work not only strongly reaffirmed that the standard (i.e., overall) Stroop interference constitutes a composite and not a unitary phenomenon, but also clearly showed the relevance of an integrative perspective bridging SC-RC and TC-RC multi-stage accounts.

So far, however, this integrative perspective is supported by only a single empirical study. Additional converging evidence is therefore required to strengthen its case. If it can be shown that task, semantic and response conflicts actually present with specific developmental trajectories, this would provide unique evidence in favor of an integrative multi-stage perspective of Stroop interference (Augustinova et al., 2018; Parris et al., submitted). Therefore, Experiment 2 was specifically aimed at providing this kind of evidence. As noted above however, before addressing this more complex question, we first set out to determine whether the oft-reported developmental change in Stroop interference (color-incongruent – color-neutral trials;  $BLUE_{green}-XXXX_{green}$ ) survives controlling for response speed.

## 2. EXPERIMENT 1

The studies of the late developmental change in the magnitude of Stroop interference discussed above have significantly increased Stroop researchers' interest in the need to control for processing speed. However, as already mentioned, at least to our knowledge, only one early developmental study had actually involved this type of control (Schiller, 1966). Because all other past studies compared untransformed raw RTs, the early developmental change in Stroop interference itself, as well as the implication of its underlying processes

(e.g., inhibition) might have been exaggerated. Additionally, the previously reported discrepancies across studies concerning the actual shape of this early developmental change (see section 1.2.) might also result – at least in part – from important differences in processing speed that exist across different age-groups, in addition to existing inter-individual differences within these groups of children (e.g., Faust, Balota, Spieler, & Ferrero, 1999 for this reasoning about Stroop interference that groups of younger vs. older adults display). Therefore, Experiment 1 was designed to examine the extent to which the early developmental change in the magnitude of Stroop interference still occurs when age-related processing speed differences are accounted for.

To this end, participating children (N=218) presenting different levels of automatization of word-recognition and/or reading skills (i.e., 1<sup>st</sup>-graders (6-7 years old), 3<sup>rd</sup>-graders (8-9 years old) and 5<sup>th</sup>-graders (10-11 years old)) were administered a standard Stroop task. Collected mean RTs to color-incongruent words (*BLUE*<sub>green</sub>) and color-neutral items (*XXXX*<sub>green</sub>) were subsequently analyzed before and after being converted into z-scores. This latter transformation was based on Faust and colleagues (1999, p. 788, equation 5; see also Jackson & Balota, 2013).

$$z = \frac{[M_{\text{participant } X}(\text{on given condition}) - M_{\text{gen}}(\text{on all conditions for participant } X)]}{SD_{\text{gen}}(\text{on all conditions for participant } X)}$$

Considering the findings of the only other study that controlled for processing speed (Schiller, 1966; see also late developmental studies cited above), we, a priori, expected the early developmental change to be reflected in the magnitude of z-transformed Stroop interference. However, given the discrepancy of past studies concerning the trajectory of early developmental change in the magnitude of Stroop interference (a consistent decrease as a function of age versus a peak after a few years of reading practice), we did not further predict how the magnitude would change across those age-groups showing robust Stroop interference.

## 2.1. Method

### 2.1.1 Participants and inclusion criteria

234 right-handed native-speakers (presenting normal or corrected-to-normal vision) were recruited from two public elementary schools in Aix-en-Provence (France). The standardized reading score in *Alouette* test (Lefavrais, 2005) was used as the inclusion criterion. Thus, sixteen children –whose score was 18 months below the expected level (i.e., formal criterion used by dyslexia assessment centers in France)– were excluded from data analyses. Information about gender, chronological and reading age of the remaining 218 children is provided in Table 1.

### 2.1.2. Design and Stimuli

The data was collected using a 2 (Stimulus-type: standard color-incongruent words vs. color-neutral letter strings)  $\times$  3 (Age-group: 1<sup>st</sup>-graders vs. 3<sup>rd</sup>-graders and 5<sup>th</sup>-graders) design, with Stimulus-type as a within-participants factor. There were 24 trials for each Stimulus-type factor condition, whose presentation order was randomly determined for each participant within a single block of 48 experimental trials.

The stimuli (presented in lowercase 24-point Courier New font on a black background) consisted of four color words (*rouge* [red], *jaune* [yellow], *bleu* [blue], and *vert* [green]) and strings of *xxx* of the same length as the color-incongruent trials. Color-incongruent words always appeared in colors that were incongruent with the meaning of their word-dimension.

### 2.1.3. Apparatus and Procedure

All aspects of this experiment were approved by the Statutory Ethics Committee of Aix-Marseille University. Parents had consented to the participation of their children by signing and returning an informed consent form. Children then participated individually during a single session lasting about 15 min and taking place in a familiar quiet room of their own school during the months of April and May.

Children were seated approximately 60 cm from the computer screen. A 17-inch HP laptop computer was used for stimulus presentation and data collection was done with DMDX software (Version 2.9.01; Forster & Forster, 2003). Their task was to identify the color of the letter-strings presented on the screen (while ignoring their meanings), as quickly and accurately as possible, by pressing with their dominant hand one of four color-keys on a button box specifically designed for this experiment. To this end, children were instructed to focus on the fixation cross (i.e., not to move their eyes from it). This white cross (“+”) appeared in the center of the (black) screen for 500 ms and it was then replaced by a letter-string that continued to be displayed until the child responded (or until 3500 ms had elapsed). Once the participant had responded, the screen was cleared and a new trial began after a 1000-ms delay. The children were familiarized with these requirements during a set of 12 practice trials consisting of strings of various letters (e.g., “bbbb” presented in green color) that was then followed by a block of experimental trials.

## 2.2. Results and Discussion

Response times (RTs) greater than 3 *SDs* above or below each participant’s mean latency for each condition (i.e., less than 2% of the total data) were excluded from the analyses. The data were subsequently analyzed in a 2 (Stimulus-type: standard color-

incongruent words vs. color-neutral letter strings)  $\times$  3 (Age-group: 1<sup>st</sup>-graders vs. 3<sup>rd</sup>-graders vs. 5<sup>th</sup>-graders) within-participants ANOVA.

The first analysis on raw RTs revealed main effects of Stimulus-Type [ $F(1, 215) = 15.22, p < .001, \eta_p^2 = .066$ ] and of Age-group [ $F(2, 215) = 67.28, p < .001, \eta_p^2 = .385$ ] but no Stimulus-Type  $\times$  Age-group interaction [ $F(2, 215) = .66, p = .52, \eta_p^2 = .006$ ] (see Table S1 in Supporting Information for further depiction of raw RTs results). To further examine, the extent to which the early developmental change in Stroop interference still occurs when age-related processing speed differences are accounted for, the same analysis was performed on  $z$ RTs (see above). It revealed a main effect of Stimulus-Type [ $F(1, 215) = 24.66, p < .001, \eta_p^2 = .103$ ] such that overall children's identification times were longer for color-incongruent Stroop words (i.e., *BLUE*<sub>green</sub>), than for color-neutral items (e.g., "XXX"<sub>green</sub>, see Table 2). This analysis no longer revealed an effect of Age-group [ $F(2, 215) = 1.68, p = .189, \eta_p^2 = .015$ ] and again, the Stimulus-Type  $\times$  Age-group interaction remained non-significant [ $F(2, 215) = 0.512, p = .60, \eta_p^2 = .005$ ] (see Table 2 for descriptive statistics).

The analysis on error rates mirrored the previous (RT) results as it revealed a main effect of Stimulus-Type [ $F(1, 215) = 8.17, p = .005, \eta_p^2 = .037$ ] such that overall children's identification times were more error-prone for color-incongruent Stroop words than for color-neutral items (see Table 2). This analysis further revealed that both the effect of Age-group [ $F(2, 215) = 0.25, p = .78, \eta_p^2 = .002$ ] and the Stimulus-Type  $\times$  Age-group interaction [ $F(2, 215) = 0.45, p = .64, \eta_p^2 = .004$ ] remained non-significant.

Taken together, these results provide no substantial evidence for early developmental change in Stroop interference across early childhood. This intermediary conclusion points to the possibility that the amplitude of the early development change in the magnitude of Stroop interference might have been exaggerated in past studies. It is however important to note at



Accepted Article

this point that the present experiment used manual responses (see also e.g., Aïté et al., 2018; Peru et al., 2006) whereas the majority of past early developmental studies – including those within neuropsychological clinical perspective (e.g., Armengol, 2002; Roy et al., 2017) – have actually used vocal responses. Given that this latter response modality is known to produce larger magnitudes of Stroop interference (see e.g., Brown, Joneleit, Robinson, & Brown, 2002; Sharma & McKenna, 1998), a possibility remains that the use of manual responses reduced the likelihood of detecting more important differences across age groups. To address this possibility, vocal responses were consequently used in Experiment 2. In addition to addressing the response mode issue, Experiment 2 was specifically designed to shed a more direct light on processes underlying the potential early developmental change in Stroop interference.

### 3. EXPERIMENT 2

Recall that the SC-RC accounts anticipate that semantic and response conflicts both contribute to standard Stroop interference, whereas the TC-RC accounts anticipate the contribution of task and response conflicts. In line with an integrative perspective bridging these two classes of account (Augustinova et al., 2018; Parris et al., submitted, see section 1.3.), we predicted that all three conflicts (i.e., task, semantic and response conflicts) contribute to standard Stroop interference that children display.

To test this hypothesis, and as in Augustinova et al. (2018), we supplemented the standard color-incongruent words (e.g., *BLUE*<sub>green</sub>) and color-neutral words (e.g., *DOG*<sub>green</sub>) that are commonly used in the standard Stroop task (see section 2) with two additional types of Stroop items. This *extended form of the semantic Stroop paradigm* (e.g., Augustinova & Ferrand, 2014 for the original version) thus also comprises color associated incongruent words (e.g., *SKY*<sub>green</sub>) and color-neutral letter-strings (e.g., *XXX*<sub>green</sub>). The inclusion of those

two additional types of Stroop items allows for the de-confounding of the three types of conflict, as explained below (see Figure 1).

First, the inclusion of the color-neutral letter-strings allows the separation of the effect of task conflict from the other two conflicts. Because the irrelevant dimension of most stimuli that this paradigm contains is readable (i.e., composed of letters), it is assumed that they all generate task conflict (see section 1.3.). Also, and importantly, they do so to the same extent, except for the non-readable color-neutral letter-strings (e.g.,  $XXX_{\text{green}}$ ). In line with the bimodal, interactive activation model with (amodal) semantics (McClelland & Rumelhart, 1981; Grainger & Ferrand, 1996; Ferrand & New, 2003), the processing of the written dimension of these color-neutral letter-strings (i.e., xxx) stops at the orthographic pre-lexical level. The processing of the written dimension for all other stimuli composed of words (e.g., dog, sky and blue) stops on the other hand with access to meaning (i.e., after a full chain of visual, orthographic, lexical and semantic processing has come to completion).

Consequently, and in line with the subtractive logic of this paradigm, the significant difference in mean response latencies between Stroop color-neutral words and letter-strings (e.g.,  $DOG_{\text{green}} - XXX_{\text{green}}$ ) that was observed in Augustinova et al.'s study was taken to solely reflect differences in activation of the irrelevant reading task set and hence of the differential amount of the task conflict that this entails (see Figure 1). Indeed, because the meaning of color-neutral words (e.g., dog for  $DOG_{\text{green}}$ ) is not related to a color (unlike sky or blue), the aforementioned contribution of task conflict to overall Stroop interference is not intermixed with that of the semantic and response conflicts that are generated by color-incongruency.

Turning now to the separation of semantic and response conflicts, numerous studies have shown that the semantic conflict is caused by color-incongruency (see Seymour's reasoning depicted in section 1.3. and e.g., Augustinova & Ferrand, 2014b for a review).

Also, and importantly, in line with Seymour (1977), semantic conflict is generated to the same extent by associated (e.g.,  $SKY_{\text{green}}$ ) as compared to standard (e.g.,  $BLUE_{\text{green}}$ ) Stroop words (e.g., Augustinova et al., 2015 for N400-like evidence). Consequently, the significant difference in mean response latencies between color-associated and color-neutral trials (e.g.,  $SKY_{\text{green}} - DOG_{\text{green}}$ , see Figure 1) that was also observed in the study of Augustinova and colleagues (see Figure 1) was taken as evidence of the semantic conflict that color-associated (e.g.,  $SKY_{\text{green}}$ ) unlike color-neutral ( $DOG_{\text{green}}$ ) Stroop words generate. Indeed, given that associated color-incongruent words do not activate (pre-)motor responses linked to the associated color (e.g., press a blue button on seeing SKY; see Schmidt & Cheesman, 2005 for a direct demonstration), the aforementioned contribution of semantic conflict to overall Stroop interference is not confounded with that of response conflict – generated by standard color-incongruent words only (e.g.,  $BLUE_{\text{green}}$ , but see Hasshim & Parris, 2014 for discussion).

Once the irrelevant word dimension of standard incongruent trials has been semantically processed, it indeed primes the aforementioned (pre-)response tendency that for these words (e.g., blue for  $BLUE_{\text{green}}$ ) is in the response set. It therefore interferes with the (pre-)response tendency primed by the meaning of the relevant color-dimension (green here). Consequently, the significant difference in mean response latencies between standard and associated color-incongruent trials (e.g.,  $BLUE_{\text{green}} - SKY_{\text{green}}$ , see Figure 1) observed in the study of Augustinova and colleagues (see Figure 1) solely results from this (pre-)motor (i.e., response) conflict occurring at the level of response processing and/or output. Indeed, both task and semantic conflict are equal in those two types of color-incongruent items ( $BLUE_{\text{green}}$  and  $SKY_{\text{green}}$ , see above).

Therefore, the positive difference in mean response latencies between color-neutral words and letter-strings (e.g.,  $DOG_{\text{green}} - XXX_{\text{green}}$ ) was used to capture the specific contribution of *task* conflict to overall Stroop interference displayed by the children participating in this experiment (see Figure 1). Additionally, the positive difference in mean response latencies between color-associated and color-neutral trials (e.g.,  $SKY_{\text{green}} - DOG_{\text{green}}$ , see Figure 1) was used to isolate the specific contribution of *semantic* conflict to overall Stroop interference. And finally, the positive difference in mean response latencies between standard color-incongruent and color-associated trials (e.g.,  $BLUE_{\text{green}} - SKY_{\text{green}}$ , see Figure 1) was used to capture the specific contribution of *response* conflict to overall Stroop interference (e.g.,  $BLUE_{\text{green}} - SKY_{\text{green}}$ ).

At least some specific contribution of *task* conflict to overall Stroop interference was expected in all age-groups. Even in pre-readers, visual expertise for letters is known to be present (Maurer, Brem, Bucher, & Brandeis, 2005), suggesting that for all our age-groups (all of whom had at least begun to receive reading instruction) one would expect written stimuli to trigger task irrelevant processing. In contrast, the contribution of *semantic* and *response* conflicts was expected to differ across age-groups. It should be remembered that both of these latter conflicts are generated by color-incongruency – involved in color-incongruent words ( $BLUE_{\text{green}}$  and  $SKY_{\text{green}}$ ) – occurring once the written-dimension of these words has been not only visually, orthographically, lexically but most importantly semantically processed. Because this latter processing chain – typical of fully developed word-recognition – only occurs in more advanced readers, the specific contribution of the *semantic* and *response* conflicts to overall Stroop interference was expected to be significant in only in 3<sup>rd</sup>-graders (8-9 years old) and 5<sup>th</sup>-graders (10-11 years old), as opposed to 1<sup>st</sup>-graders (6-7 years old) that are only starting to receive reading instruction in France.

### 3.1. Method

#### 3.1.1. Participants and inclusion criteria

173 native-speakers (presenting normal or corrected-to-normal vision) were recruited from a single private elementary school in the suburbs of Clermont-Ferrand (France). The standardized reading score in *Timé* test (Ecalte, 2003; 2006) was used as the inclusion criterion. Thus, twelve children –whose score was 18 months below the expected level (i.e., as in Experiment 1)– were excluded from data analyses. Twenty-four other children were also excluded from further analyses due to a malfunctioning microphone. Information on gender, chronological and reading age of the remaining 137 children is provided in Table 3.

#### 3.1.2. Design and Stimuli

The data were collected using a 4 (Stimulus-type: standard color-incongruent words vs. associated color-incongruent words vs. color-neutral words vs. color-neutral letter strings) × 3 (Age-group: 1<sup>st</sup>-graders vs. 3<sup>rd</sup>-graders and 5<sup>th</sup>-graders) design, with Stimulus-type as a within-participants factor. There were 36 trials for each Stimulus-type factor condition, whose presentation order was randomly determined for each participant within a single block of 144 experimental trials. Thus, as in Experiment 1, 50% of the experimental trials involved color-incongruency (i.e., standard color-incongruent words and associated color-incongruent words), whereas 50% of the experimental trials were color-neutral (color-neutral words and color-neutral letter strings).

The stimuli (presented in lowercase 18-point Courier font on a black background) consisted of four color words: *rouge* [red], *jaune* [yellow], *bleu* [blue], and *vert* [green]; four color-associated words: *tomate* [tomato], *maïs* [corn], *ciel* [sky], and *salade* [salad]; four color-neutral words: *balcon* [balcony], *robe* [dress], *pont* [bridge] and *chien* [dog]; and four strings of xxx of the same length as the color-incongruent trials. In each condition, all the

stimuli were similar in length (4.5, 5, 4.75 and 4.75 letters on average for the color-incongruent words, the color-associated words, the color-neutral words and the strings of Xs respectively) and were of very high frequency (154, 114 and 142 occurrences per million for the color-incongruent words, the color-associated words and the color-neutral words, respectively) according to *MANULEX* (Lété, Sprenger-Charolles, & Colé, 2004)<sup>1</sup>. Color-incongruent and color-associated words always appeared in colors that were incongruent with the meaning of their word-dimension.

### 3.1.3. Apparatus and Procedure

All aspects of this experiment were approved by the South-Eastern Statutory Ethics Committee (*Comité de Protection des Personnes*, CPP, Sud-Est 6, France). Parents had consented to the participation of their children by signing and returning an informed consent form. Children then participated individually during a single session lasting about 20 min and taking place in a familiar quiet room of their school during January and February.

The children were seated approximately 50 cm from the computer screen. A 15-inch HP laptop computer was used for stimulus presentation and data collection was done with E-Prime 2.1 (Psychology Software Tools, Pittsburgh). Their task was to name the color of the letter-strings presented on the screen (while ignoring their meanings). To this end, the children were instructed to focus on the fixation cross (i.e., not to move their eyes from it). The white cross (“+”) appeared in the center of the (black) screen for 1000 ms which was then replaced by a letter-string that continued to be displayed until the participant responded (or until 3500 ms had elapsed). The participants’ vocal responses were recorded via a Philips SBC ME570 microphone mounted on a headset specially designed for children (KAMA

---

<sup>1</sup> *MANULEX* is a grade-level lexical database based on French elementary school textbooks. To ensure that all the children were familiar with the words, only entries that overlapped in the 1<sup>st</sup> through the 5<sup>th</sup>-grade lexicon were selected. Note that, by definition, no frequencies for strings of Xs are reported in this database.

stereo headset) and stored on a Sony recorder model ICD-PX333. The children were familiarized with these requirements during a set of 16 practice trials consisting of strings of asterisks (presented in the four colors), followed by 12 additional experimental-like practice trials and then by a block of experimental trials.

### 3.2. Results and Discussion

Response times (RTs) greater than 3 *SDs* above or below each participant's mean latency for each condition (i.e., less than 2% of the total data) were excluded from the analyses. In order to control for differences in processing speed, these were then transformed into *zRT* (see section 2 for further details). Mean raw and *zRT* along with error rates were subsequently analyzed in a 4 (Stimulus-type: standard color-incongruent words vs. associated color-incongruent words vs. color-neutral words vs. color-neutral letter strings)  $\times$  3 (Age-group: 1<sup>st</sup>-graders vs. 3<sup>rd</sup>-graders and 5<sup>th</sup>-graders) within-participants ANOVA (see Table 4 for descriptive statistics).

On mean raw RT, this analysis revealed main effects of Stimulus-Type [ $F(3, 402) = 100.53, p < .001, \eta_p^2 = .43$ ] and of Age-group [ $F(2, 134) = 15.29, p < .001, \eta_p^2 = .19$ ]; as well as a Stimulus-Type  $\times$  Age-group interaction [ $F(6, 402) = 3.76, p = .001, \eta_p^2 = .05$ ] (see Table S2 in Supporting Information for further depiction of results on raw RTs). On mean *zRTs*, this analysis revealed a main effect of Stimulus-Type [ $F(3, 402) = 142.42, p < .001, \eta_p^2 = .515$ ] but as in Experiment 1, the effect of Age-group was no longer significant [ $F(2, 134) = .000, p = 1.000, \eta_p^2 = .001$ ]. However, and in line with the idea that the early developmental change in the magnitude of Stroop interference still occurs when age-related processing speed differences are accounted for, the Stimulus-Type  $\times$  Age-group interaction was also significant [ $F(6, 402) = 10.05, p < .001, \eta_p^2 = .130$ ]. Finally, the analysis of error rates revealed main effects of Stimulus-Type [ $F(3, 402) = 99.48, p < .001, \eta_p^2 = .426$ ] and of Age-group [ $F(2, 134) = 7.24,$

$p=.001, \eta_p^2=.098$ ] as well as Stimulus-Type  $\times$  Age-group interaction [ $F(6, 402)=8.12, p<.001, \eta_p^2=.108$ ].

### 3.2.1. Analyses of the overall Stroop interference (observed with vocal responses)

Further contrast analyses (with Bonferroni corrections to counteract problems associated with multiple comparisons) revealed that, the mean  $z$ RTs for standard color-incongruent words (e.g.,  $BLUE_{\text{green}}$ ) was significantly greater than the one for color-neutral letter-strings (e.g.,  $XXX_{\text{green}}$ ) in 1<sup>st</sup>-graders (6-7 years old) [ $F(1,52)=22.87, p<.001, \eta_p^2=.305$ ], in 3<sup>rd</sup>-graders (8-9 years old) [ $F(1,45)=169.69, p<.000, \eta_p^2=.790$ ] and in 5<sup>th</sup>-graders (10-11 years old) [ $F(1,37)=141.27, p<.001, \eta_p^2=.792$ ] (see Table 4 for descriptive statistics). The same contrast analyses conducted on mean error-rates reported above mirrored those on  $z$ RTs. Indeed, in 1<sup>st</sup>-graders, mean error-rates for standard color-incongruent trials were significantly greater than those observed for color-neutral color-strings, [ $F(1,52)=21.35, p<.001, \eta_p^2=.291$ ] (see Table 4 for descriptive statistics). This latter contrast was also significant in 3<sup>rd</sup>-graders [ $F(1,45)=42.38, p<.001, \eta_p^2=.485$ ] and in 5<sup>th</sup>-graders (i.e., skilled-readers) [ $F(1,37)=57.23, p<.001, \eta_p^2=.607$ ]. Taken together, these latter results suggest that Stroop interference was present in both  $z$ -RTs and errors (see also Table S2 in Supporting Information for depiction of raw RTs).

Because the overall Stroop interference ( $BLUE_{\text{green}}-XXX_{\text{green}}$ , see Table 4) was significant in all age-groups, computed mean magnitudes of this interference were further analyzed in a single-factor ANOVA as a function of this factor. The effect of Age-group was significant [ $F(2,134)=15.47, p<.001, \eta_p^2=.188$ ] such that (see Figure 2), the magnitude of overall Stroop interference was significantly smaller in 1<sup>st</sup>-graders (6-7 years old) as compared to 3<sup>rd</sup>-graders (8-9 years old) [ $F(1, 97)=22.06, p < .001, \eta_p^2=.185$ ]. It was also significantly smaller in 1<sup>st</sup>-graders (6-7 years old) as compared to 5<sup>th</sup>-graders (10-11 years



old) [ $F(1,89)=18.92, p<.001, \eta_p^2=.175$ ] but there was no difference between magnitudes of the overall Stroop interference observed in 3<sup>rd</sup>-graders (8-9 years old) as compared to 5<sup>th</sup>-graders (10-11 years old),  $F(1,82)=.007, p = .94; \eta_p^2<.001$ ].

### 3.2.2. Analyses of different components of the overall Stroop interference

Further contrast analyses (with Bonferroni corrections to counteract problems associated with multiple comparisons) conducted with the aforementioned 4-Stimulus-type  $\times$  3-Age-group ANOVA (see above) revealed that – as expected – the overall Stroop interference displayed by 1<sup>st</sup>-graders (6-7 years old) solely resulted from the significant contribution of the task conflict (see Figure 3). Indeed, the mean  $z$ RTs for color-neutral words (e.g.,  $DOG_{\text{green}}$ ) was significantly greater than the one for color-neutral letter-strings (e.g.,  $XXX_{\text{green}}$ ) in this age-group, [ $F(1,52)=24.84, p<.001, \eta_p^2=.323$ ], whereas all the other contrasts used as proxies for respectively semantic [ $F(1,52)=2.30, p=.136, \eta_p^2=.042$ ] and response conflict [ $F(1,52)=.26, p=.610, \eta_p^2=.005$ ] remained non-significant (see Table 4 for descriptive statistics).

As can be seen in Figure 3, the overall Stroop interference displayed by 3<sup>rd</sup>- and 5<sup>th</sup>-graders (respectively 8-9 and 10-11-years old) resulted from a joint contribution of all three conflicts. Indeed, mean  $z$ RTs for color-neutral words (e.g.,  $DOG_{\text{green}}$ ) being significantly greater than the one for color-neutral letter-strings (e.g.,  $XXX_{\text{green}}$ ) in both 3<sup>rd</sup>- [ $F(1, 45)=64.58, p<.001, \eta_p^2=.589$ ] and 5<sup>th</sup>-graders [ $F(1, 37)= 42.75, p<.001, \eta_p^2=.536$ ] is in line with the significant contribution of task conflict. Mean  $z$ RTs for associated color-incongruent words (e.g.,  $SKY_{\text{green}}$ ) being significantly greater than the one for color-neutral words (e.g.,  $DEAL_{\text{green}}$ ) in both 3<sup>rd</sup>- [ $F(1, 45)=6.00, p=.018, \eta_p^2=.118$ ] and 5<sup>th</sup>-graders [ $F(1, 37)= 6.23, p=.017, \eta_p^2=.144$ ] is in line with the significant contribution of the semantic conflict. Finally, mean  $z$ RTs for standard color-incongruent words (e.g.,  $BLUE_{\text{green}}$ ) being significantly greater than the one for associated color-incongruent words (e.g.,  $SKY_{\text{green}}$ ) in both 3<sup>rd</sup>- [ $F(1,$

45)=33.37,  $p<.001$ ,  $\eta_p^2=.426$ ] and 5<sup>th</sup>-graders [ $F(1, 37)= 38.78$ ,  $p<.001$ ,  $\eta_p^2=.512$ ] is in line with the significant contribution of the response conflict.

#### 4. CROSS-EXPERIMENT ANALYSES

It should be remembered at this point that the Stimulus-Type  $\times$  Age-group interaction on zRTs – that is likely to epitomize the early developmental change in the magnitude of Stroop interference – remained non-significant in Experiment 1, whereas it was significant in Experiment 2. Given these discrepancies (that are not uncommon, see e.g., Tse & Neely, 2007), the very first aim of the cross-experiment analyses was to examine whether the Stimulus-Type  $\times$  Age-group interaction on zRTs actually replicates across multiple independent experiments. To this end, we used a procedure known as Winer's z test (Winer, 1971, pp. 49–50; see e.g., Augustinova, Flaudias, & Ferrand, 2010; Tse & Neely, 2007, for applications). When the F values of Experiment 1 and 2 were combined using this approach, the Stimulus-Type  $\times$  Age-group interaction was indeed significant ( $z = 3.87$ ,  $p <.01$ ).

In light of this significant interaction – suggesting that the early developmental change in the magnitude of Stroop interference is indeed likely to replicate across studies –, the second aim of the cross-experiment analyses was to examine further its trajectory. To this end, mean zRTs for standard color-incongruent words (e.g., BLUE<sub>green</sub>) and color-neutral letter-strings (e.g., XXX<sub>green</sub>) observed respectively in Experiment 1 (using manual responses) and in Experiment 2 (using vocal responses) were combined in a single analysis.

This 2 (Stimulus-type: standard color-incongruent words vs. color-neutral letter strings)  $\times$  3 (Age-group: 1st-graders vs. 3rd-graders vs. 5th-graders) ANOVA revealed a main effect of Stimulus-Type [ $F(1,352)=154.384$ ,  $p<.001$ ,  $\eta_p^2=.305$ ], a main effect of Age-group [ $F(2,352)=6.673$ ,  $p=.001$ ,  $\eta_p^2=.037$ ] and a Stimulus-Type  $\times$  Age-group interaction [ $F(2, 352)=6.160$ ,  $p=.002$ ,  $\eta_p^2=.034$ ]. Further contrast analyses revealed that overall Stroop

interference was significant in 1<sup>st</sup>-graders [ $F(1,128)=24.075, p<.001, \eta_p^2=.158$ ], 3<sup>rd</sup>-graders [ $F(1,126)=75.234, p<.001, \eta_p^2=.374$ ] and 5<sup>th</sup>-graders [ $F(1,98)=58.319, p<.001, \eta_p^2=.373$ ].

However, in line with the change in overall Stroop interference (e.g.,  $BLUE_{green}-XXX_{green}$ ) in the course of early development, its magnitude was smaller in 1<sup>st</sup>-graders (6-7 years old) than in 3<sup>rd</sup>-graders (8-9 years old) [ $M_s=.202$  vs.  $.395; F(1,254)=9.95, p=.002, \eta_p^2=.038$ ] and 5<sup>th</sup>-graders (10-11 years old) [ $M_s=.202$  vs.  $.392; F(1,226)=8.56, p=.004, \eta_p^2=.036$ ]. The magnitude observed in 3<sup>rd</sup>-graders did not differ from the one observed in 5<sup>th</sup>-graders [ $M_s=.395$  vs.  $.392; F(1,224)=0.002, p=.963, \eta_p^2<.001$ ]. Nevertheless, the magnitude of Stroop interference increased linearly as a function of grade as indicated by a significant linear contrast [ $F(1,352)=8.22, p=.004$ ] and an only marginally significant quadratic contrast [ $F(1,352)=3.18, p=.075$ ].

## 5. GENERAL DISCUSSION

The results presented above provide no statistical evidence for early developmental change in the magnitude of Stroop interference with the manual response Stroop task (Experiment 1). However, Experiment 2 (using the vocal response Stroop task) and data combined from both experiments indicate that this change still occurs such that the magnitude of overall Stroop interference is larger in 3<sup>rd</sup>-graders (8-9 years old) and 5<sup>th</sup>-graders (10-11 years old) as compared to 1<sup>st</sup>-graders (6-7 years old). Inspection of these combined results and of Figure 2 (see above) indicates that – in line with past findings (e.g., Brown et al., 2002; Sharma & McKenna, 1998), the use of vocal (instead of manual) responses is likely to accentuate the overall pattern of early developmental change that was undetected with manual responses (see Figure 2).

Also, and importantly, the trajectory of this developmental change convincingly shows that – independently of differences in processing speed that undeniably exist both within and between age-groups – the magnitude of Stroop interference peaks after two or three years of reading practice (Schiller, 1966; Schadler & Thissen, 1981; Rand et al., 1963). Perhaps only after the peak reported here will its magnitude start decreasing as reported by numerous studies reviewed above (see section 1.2.). Given the age-groups included in our study, we are unable to demonstrate the actual onset of this decrease that some have claimed is likely to stop with maturation of executive functions (Aïte et al., 2018; Diamond, 2013; Prencipe et al., 2011). These remaining issues are of considerable importance and thus remain to be addressed by future empirical studies mobilizing age-groups beyond the 5<sup>th</sup>-grade (10-11 years old, i.e., the oldest age-group in the present study).

Despite this limitation, another strength of the present study is that it addressed, for the first time, the isolable processes underlying Stroop interference and their different developmental trajectories. Indeed, all past studies investigating the early development of Stroop interference simply inferred those processes from the change in its overall magnitude. Therefore, the measures used in adult Stroop studies (e.g., Augustinova et al., 2018; see also Augustinova & Ferrand, 2014a, b) to assess these underlying processes were applied in Experiment 2 while the potentially confounding effect of processing speed was controlled for. The ensuing results point to perhaps the most novel aspect of the present study.

It lies in the fact that the aforementioned peak in the magnitude of Stroop interference reliably resulted from the joint contributions of task, semantic and response conflicts. Indeed, significantly smaller magnitudes of Stroop interference observed in 1<sup>st</sup>-graders (6-7 years old) resulted from a sole contribution of task conflict. This change reflects the fact that all distinct components of Stroop interference (resulting from specific conflicts) are progressively falling into place, yet not at the same pace. Initially only task conflict contributes to Stroop

interference in 1<sup>st</sup>-graders, but then after two or three years of reading practice (i.e., in 3<sup>rd</sup>- and 5<sup>th</sup>-graders), the magnitude of Stroop interference peaks substantially (Schiller, 1966; Schadler & Thissen, 1981; Rand et al., 1963) – at least when the vocal response-modality is used – and this peak is specifically due to the joint contributions of task, semantic and response conflicts.

It is important to note that the present findings are exempt from the common limitations arising from the employment of color-congruent trials (e.g.,  $BLUE_{blue}$ ). Even though this type of trial is often used in the Stroop literature to supposedly measure Stroop *interference*, the positive difference in color-naming times between color-incongruent and color-congruent items (e.g.,  $BLUE_{blue} - BLUE_{green}$ ) actually corresponds to “(...) the sum of *facilitation* and *interference*, each in *unknown* amounts” (MacLeod, 1991, p.168, *italics added*). Furthermore, processing underlying facilitation on congruent trials still remains unclear and has been interpreted variously as resulting from response convergence (Cohen et al., 1990), inadvertent reading (Kane & Engle, 2003; MacLeod & McDonald, 2000), or competing influences of interference and facilitation in different portions of the RT distribution (Heathcote et al., 1991). Finally, the inclusion of congruent trials can also lead to response contingency effects (Melara & Algom, 2003; Schmidt & Besner, 2008) and it remains unclear whether facilitation would remain if response contingency was controlled for. The aforementioned concerns have been successfully avoided by the deployment of an appropriate color-neutral base-line in the present study (see Augustinova et al., 2016; T. L. Brown, 2011; Parris et al., submitted, for discussions). Therefore, the results of the present study have several straightforward implications that are methodological, theoretical and applied in nature.

In line with what is now common practice in adult developmental Stroop studies (e.g., Augustinova et al., 2018; Bugg et al., 2007; Spieler et al., 1996), these results (1) invite early developmental researchers' using the Stroop task to increase their interest in the need to control for processing speed (e.g., Lété & Fayol, 2013; Ziegler et al., 2014 for this type of endeavor in psycholinguistic children's studies). These results also (2) invite researchers to preferentially use vocal responses especially when the composite nature of Stroop interference is under scrutiny (see also Augustinova, Parris, & Ferrand, 2019; Augustinova & Ferrand, 2014b for discussions of this issue in adult studies). Another implication of considerable methodological importance is that (3) in order to avoid any conclusions about the (early) developmental trajectory of *inhibition* (i.e., an executive function that deals with interference) remaining tentative, congruent trials (i.e., trial producing facilitation) should be avoided.

At the theoretical level, the findings reported in this paper clearly point (4) to a composite (as opposed to unitary) nature of Stroop interference, and therefore the fact that (5) historically favored single-stage response accounts – in which the customary measures of Stroop interference ( $BLUE_{green} - DOG_{green} / BLUE_{blue}$ ) are rooted – are likely incorrect. Indeed, in addition to differences in magnitude, (6) the specific contributions of task, semantic and response conflicts to overall interference were evidenced by their distinct developmental trajectories. These contributions and their specificity thus reinforce the idea that (7) each component of Stroop interference reflects a unique underlying process that is likely to correspond to goal-maintenance (task conflict), automatization of word-reading (semantic conflict) and response-inhibition (response conflict) respectively. Taken together, these results consequently (8) provide unique developmental evidence in favor of an integrative multiple-stage perspective placing the locus of Stroop interference at multiple levels in the processing stream (respectively task, semantic and response conflict; Augustinova et al.,

2018; Parris et al., submitted). Because of its integrative nature, this latter perspective allows the reconciling of different perspectives on processes thought to underlie (9) the magnitude of Stroop interference (see section 1.3.) and (10) the early developmental change in this magnitude (see section 1.2.)

The results from the present study also suggest that (11) potentially different neural substrates underlie each conflict type. Consistent with this notion, van Veen and Carter (2005) observed no overlap of activation between semantic and response conflict. They showed that semantic conflict activated dorso-lateral prefrontal cortex, posterior parietal cortex and the anterior cingulate cortex (ACC), whereas response conflict activated more inferior lateral prefrontal cortex, left premotor areas and regions of the ACC more anterior and ventral to that activated by semantic conflict. The authors concluded that their data were evidence for separate but analogous mechanisms for dealing with different kinds of representational conflict (see also Chen et al., 2013; Milham et al. 2001). Consistently, task conflict might also have a unique neural substrate. MacLeod and MacDonald (2000) noted that the ACC appeared to be more activated by incongruent and congruent stimuli when compared to repeated letter neutral stimuli (e.g., xxx; see also, Bench et al., 1993) indicating the detection of task conflict. That said, no study has yet directly investigated this possibility using the required contrast of color-neutral words to color-neutral letter-string, so the precise location of activation within the ACC associated with task conflict is not known. The present study therefore (12) motivates a comparison of the neural substrates of all three conflict types in the same neuroimaging study.

Finally, the reported findings are equally important from an applied point of view. Indeed, (12) the fact that the specific contribution of all three types of conflict can be clearly seen within the semantic Stroop paradigm administered with vocal responses (Experiment 2) might make it possible to construct a more sensitive evaluation tool that is simple enough

(i.e., a card version) to be administered in both lab and field (i.e., clinical) settings. Indeed, the evaluation tools that are currently used in developmental neuropsychological practice (e.g., Armengol, 2002, Roy et al., 2017) all confound the contributions of task, semantic and response conflicts. The extra sensitivity of the Stroop test (stemming from the ability to detect and rate each of these conflicts separately) would provide clinical practitioners with invaluable information since the different forms of conflict are detected and resolved by different neural regions (see above).

It is indeed likely that the different conflicts involved in Stroop interference are selectively impacted by various clinical conditions. To illustrate, the consistently poorer performance of children with attention deficit hyperactivity disorder (ADHD) on the Stroop test is likely to result from increased task and response (as opposed to semantic) conflicts since these conditions impact attentional selectivity and executive functioning (Homack & Riccio, 2004). Extra specificity of the Stroop test would give clinicians new insights including regarding potential improvement of the condition (or the lack of) over time. Given the important individual differences in responsiveness of children with ADHD to both behavior therapy (e.g., Heath, Curtis, Fan & McPherson, 2015) and drugs (e.g., Burton et al., 2015), these issues are of utmost importance and should therefore be addressed directly in future studies.

## **6. CONCLUSION**

While we urge caution in extrapolating from data from a single task to make claims about development more generally, the present study makes one important contribution. It shows that researchers interested in the development of executive functions, especially inhibitory control, should focus more on how conflict is created in the tasks they use, and what specific types of conflict these tasks generate. Different types of conflict may emerge,



and be successfully resolved, at different ages. This point echoes that recently made by Simpson et al. (2012), who found that 4-year-old children in a non-linguistic Stroop-like task (the Day-Night task) were insensitive to semantic conflict, but were sensitive to response conflict. These findings are broadly consistent with the lack of semantic conflict observed in younger children in Experiment 2 of the present study. A more detailed theoretical treatment of these ideas has been made recently in Simpson and Carroll (in press).

Given the ubiquitous use of the Stroop task to study a wide variety of phenomena in psychology and cognitive science broadly defined, the present study potentially makes another important point. Currently, the Stroop task is considered as “a prototypical inhibition task (...) in which one needs to inhibit or override the tendency to produce a more dominant or automatic response (i.e., name the color word)” and the magnitude of Stroop interference as reflecting “one’s ability to deliberately inhibit dominant, automatic, or prepotent responses when necessary” (Miyake et al., 2000, p.57). Therefore, perhaps the most immediate conclusions to be drawn from the present study is that Stroop interference is complex and that its underlying processes might remain unseen and/or be misinterpreted when observed using the standard Stroop paradigm. The use of a more fine-grained implementation of a Stroop task (see also e.g. De Houwer, 2003; Hasshim & Parris, 2014; 2015; 2018) should therefore be considered.

#### **Data Availability Statement**

Data (for each participant and each condition) for both Experiment 1 and 2 are publicly available via the Open Science Framework and can be accessed at <https://osf.io/69fkj/>

Doi: 10.17605/OSF.IO/69FKJ (Contact: Ludovic Ferrand)

## References

- Aïte, A., Cassotti, M., Linzarini, A., Osmont, A., Houdé, O., & Borst, G. (2018). Adolescents' inhibitory control: Keep it cool or lose control. *Developmental Science, 21*, doi: 10.1111/desc.12491
- Armengol, C. G. (2002). Stroop test in Spanish: Children's norms. *Clinical Neuropsychologist, 16*, 67–80.
- Augustinova, M., & Ferrand, L. (2014a). Social priming of dyslexia and reduction of the Stroop effect: What component of the Stroop effect is actually reduced? *Cognition, 130*, 442–454.
- Augustinova, M., & Ferrand, L. (2014b). Automaticity of word reading: Evidence from the semantic Stroop paradigm. *Current Directions in Psychological Science, 23*, 343–348.
- Augustinova, M., Flaudias, V., & Ferrand, L. (2010). Single-letter coloring and spatial cuing do not eliminate or reduce a semantic contribution to the Stroop effect. *Psychonomic Bulletin & Review, 17*, 827–833.
- Augustinova, M., Parris, B., & Ferrand, L. (2019). The loci of Stroop interference and facilitation effects with manual and vocal responses. *Frontiers in Psychology, 10*:1786. doi: 10.3389/fpsyg.2019.01786
- Augustinova, M., Silvert, L., Ferrand, L., Llorca, P. M., & Flaudias, V. (2015). Behavioral and electrophysiological investigation of semantic and response conflict in the Stroop task. *Psychonomic Bulletin & Review, 22*, 543–549.
- Augustinova, M., Clarys, D., Spatola, N., & Ferrand, L. (2018). Some further clarifications on age-related differences in Stroop interference. *Psychonomic Bulletin & Review, 25*, 767–774.
- Augustinova, M., Silvert, S., Spatola, N., & Ferrand, L. (2018). Further investigation of distinct components of Stroop interference and of their reduction by short response

stimulus intervals. *Acta Psychologica*, 189, 54–62.

- Augustinova, M., Almeida, E., Clarys, D., Ferrand, L., Izaute, M., Jalenques, I., Juneau, C., Normand, A., Silvert, L. (2016). Que mesure l'interférence Stroop ? Quand et comment ? Arguments méthodologiques et théoriques en faveur d'un changement de pratiques dans sa mesure. [What the Stroop interference actually measures? When and how? Methodological and theoretical arguments in favor of change in its measurement.], *L'Année Psychologique/Topics in Cognitive Psychology*, 116, 45-66.
- Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Paulesu, E., Frackowiak, R. S. J., et al. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia*, 31, 907–22.
- Brown, T. L. (2011). The Relationship Between Stroop Interference and Facilitation Effects: Statistical Artifacts, Baselines, and a Reassessment. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 85-99.
- Brown, T. L., Joneleit, K., Robinson, C. S., & Brown, C. R. (2002). Automaticity in reading and the Stroop task: Testing the limits of involuntary word processing. *American Journal of Psychology*, 115, 515–543.
- Bub, D. N., Masson, M. E. J., & Lalonde, C. E. (2006). Cognitive control in children: Stroop interference and suppression of word reading. *Psychological Science*, 17, 351–357.
- Bugg, J. M., DeLosh, E. L., Davalos, D. B., & Davis, H. P. (2007). Age differences in Stroop interference: Contributions of general slowing and task-specific deficits. *Aging, Neuropsychology, and Cognition*, 14, 155–167.
- Burton, B., Grant, M., Feigenbaum, A., Singh, R., Hendren, R., Siriwardena, K., ... & Prasad, S. (2015). A randomized, placebo-controlled, double-blind study of sapropterin to treat ADHD symptoms and executive function impairment in children and adults with sapropterin-responsive phenylketonuria. *Molecular Genetics and Metabolism*, 114, 415–

- Chen, Z., Lei, X., Ding, C., Li, H., & Chen, A. (2013). The neural mechanisms of semantic and response conflicts: An fMRI study of practice-related effects in the Stroop task. *NeuroImage*, *66*, 577–584.
- Chevalier, N., & Blaye, A. (2008). Cognitive flexibility in preschoolers: The role of representation activation and maintenance. *Developmental Science*, *11*, 339–353.
- 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*, 332–361.
- Comalli, P. E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *The Journal of Genetic Psychology*, *100*, 47–53.
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037–2078.
- De Houwer, J. (2003). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition*, *31*, 353–359.
- De Jong, R., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: dissociable causes of interference effects in conflict situations. *Acta Psychologica*, *101*, 379–394.
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. *Principles of Frontal Lobe Function*, 466–503.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

Ecalle, J. (2003). *Timé 2: Test d'Identification de Mots Écrits pour enfants de 6 à 8 ans*.

[Timé 2: A visual word identification test for children from 6-to-8 years old]. Paris, France: ECPA.

Ecalle, J. (2006). *Timé 3: Test d'Identification de Mots Écrits pour enfants de 7 à 15 ans*.

[Timé 3: A visual word identification test for children from 7-to-15 years old]. Paris, France: Mot-à-Mot.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, 8, 1784–1790.

Ehri, L. C., & Wilce, L. S. (1979). Does word training increase or decrease interference in a Stroop task? *Journal of Experimental Child Psychology*, 27, 352–364.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, 125, 777–799.

Ferrand, L., & New, B. (2003). Associative and semantic priming in the mental lexicon. In P. Bonin (Ed.), *The mental lexicon: Some words to talk about words* (pp. 25-43). New York: Nova Science Publishers.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35, 116–124.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, 7, 237–241.

Goldfarb, L., & Henik, A. (2006). New data analysis of the Stroop matching task calls for a theory reevaluation. *Psychological Science*, 17, 96–100.

Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 1170–1176.

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual

- word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*, 35, 623–647.
- Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2015). How does Stroop interference change with practice? A reappraisal from the musical Stroop paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 417–425.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193-225). San Diego, CA: Academic Press.
- Hasshim, N., & Parris, B. A. (2014). Two-to-one color-response mapping and the presence of semantic conflict in the Stroop task. *Frontiers in Psychology*, 5:1157. doi: 10.3389/fpsyg.2014.01157
- Hasshim, N., & Parris, B. A. (2015). Assessing stimulus–stimulus (semantic) conflict in the Stroop task using saccadic two-to-one color response mapping and prereponse pupillary measures. *Attention, Perception, & Psychophysics*, 77, 2601–2610.
- Hasshim, N., & Parris, B. A. (2018). Trial type mixing substantially reduces the response set effect in the Stroop task. *Acta Psychologica*, 189, 43–53.
- Heath, C. L., Curtis, D. F., Fan, W., & McPherson, R. (2015). The association between parenting stress, parenting self-efficacy, and the clinical significance of child ADHD symptom change following behavior therapy. *Child Psychiatry & Human Development*, 46, 118–129.
- Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: an example using the Stroop task. *Psychological Bulletin*, 109, 340–347.
- Hommel, B. (1997). Interactions between stimulus-stimulus congruence and stimulus-response compatibility. *Psychological Research*, 59, 248–260.
- Jackson, J. D., & Balota, D. A. (2013). Age-related changes in attentional selection: Quality

- of task set or degradation of task set across time? *Psychology and Aging*, 28, 744-753.
- Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin*, 109, 490–501.
- Kail, R. V. (2007). Longitudinal evidence that increases in processing speed and working memory enhance children's reasoning. *Psychological Science*, 18, 312–313.
- Kalanthroff, E., Goldfarb, L., & Henik, A. (2013). Evidence for interaction between the stop-signal and the Stroop task conflict. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 579–592.
- 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, 1356–1367.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47–70.
- Killikelly, C., & Szűcs, D. (2013). Asymmetry in stimulus and response conflict processing across the adult lifespan: ERP and EMG evidence. *Cortex*, 49, 2888–2903.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.
- Lefavrais, J. (2005). *Alouette–R: Test d'analyse de la lecture et de la dyslexie* [Alouette–R: Test for the analysis of reading and dyslexia]. Paris, France: Editions du Centre de Psychologie Appliquée.
- Lété, B., & Fayol, M. (2013). Substituted-letter and transposed-letter effects in a masked priming paradigm with French developing readers and dyslexics. *Journal of Experimental Child Psychology*, 114, 47–62.
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). Manulex: A grade-level lexical database

from French elementary-school readers. *Behavior Research Methods, Instruments, and Computers*, *36*, 156–166.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*, 1357–1372.

Luna, B., Thulborn, K. R., Munoz, D. P., Merriam, E. P., Garver, K. E., Minshew, N. J., ... & Sweeney, J. A. (2001). Maturation of widely distributed brain function subserves cognitive development. *NeuroImage*, *13*, 786–793.

Luo, C.R. (1999). Semantic competition as the basis of Stroop interference: Evidence from color-word matching tasks. *Psychological Science*, *10*, 35–40.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*, 383–391.

Manwell, L. A., Roberts, M. A., & Besner, D. (2004). Single letter coloring and spatial cueing eliminates a semantic contribution to the Stroop effect. *Psychonomic Bulletin & Review*, *11*, 458–462.

Maurer, U., Brem, S., Bucher, K., & Brandeis, D. (2005). Emerging neurophysiological specialization for letter strings. *Journal of Cognitive Neuroscience*, *17*, 1532–1552.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic finding. *Psychological Review*, *88*, 375–407.

Megherbi, H., Elbro, C., Oakhill, J., Segui, J., & New, B. (2018). The emergence of automaticity in reading: effects of orthographic depth and word decoding ability on an adjusted Stroop measure. *Journal of Experimental Child Psychology*, *166*, 652–663.

Melara, R. D., & Algom, D. (2003). Driven by information: a tectonic theory of Stroop effects. *Psychological Review*, *110*, 422–471.



- Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., & Kramer, A. F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research, 12*, 467–473.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.
- Monsell, S., Taylor, T. J., & Murphy, K. (2001). Naming the color of a word: Is it responses or task sets that compete? *Memory & Cognition, 29*, 137–151.
- Morton, J., & Chambers, S. M. (1973). Selective attention to words and colours. *Quarterly Journal of Experimental Psychology, 25*, 387–397.
- Munoz, D. P., Broughton, J. R., Goldring, J. E., & Armstrong, I. T. (1998). Age-related performance of human subjects on saccadic eye movement tasks. *Experimental Brain Research, 121*, 391–400.
- Myerson, J., Hale, S., Wagstaff, D., Poon, L. W., & Smith, G. A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review, 97*, 475–487.
- Nettelbeck, T., & Burns, N. R. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Personality and Individual Differences, 48*, 379–384.
- Parris, B. A. (2014). Task conflict in the Stroop task: When Stroop interference decreases as Stroop facilitation increases in a low task conflict context. *Frontiers in Psychology, 5*:1182. doi: 10.3389/fpsyg.2014.01182
- Parris, B. A., Bate, S., Brown, S. D., & Hodgson, T. L. (2012). Facilitating goal-oriented behaviour in the Stroop task: When executive control is influenced by automatic processing. *PloS one, 7*(10), e46994.
- Parris, B. A., Hasshim, N., Wadlsey, M., Augustinova, M., & Ferrand, L. (submitted). *The*

*loci of the Stroop effect? A critical review of evidence for varieties of conflict (and facilitation) in the colour-word Stroop task.*

- Peru, A., Faccioli, C., & Tassinari, G. (2006). Stroop effects from 3 to 10 years: The critical role of reading acquisition. *Archives Italiennes de Biologie, 144*, 45–62.
- Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011). Development of hot and cool executive function during the transition to adolescence. *Journal of Experimental Child Psychology, 108*, 621–637.
- Rand, G., Wapner, S., Werner, H., & McFarland, J. H. (1963). Age differences in performance on the Stroop color-word test. *Journal of Personality, 31*, 534–558.
- Risko, E. F., Schmidt, J. R., & Besner, D. (2006). Filling a gap in the semantic gradient: Color associates and response set in the Stroop task. *Psychonomic Bulletin & Review, 13*, 310–315.
- Roy, A., Kefi, M.-Z., Bellaj, T., Fournet, N., Le Gall D., & Roulin J.-L. (2016). The Stroop test: A developmental study in a French children sample aged 7 to 12 years. *Psychologie Française, 63*, 129–143.
- Salthouse, T. A. (1996). A processing-speed theory of adult age differences in cognition. *Psychological Review, 103*, 403–428.
- Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *Journal of Gerontology: Psychological Sciences and Social Sciences, 50B*, P297–P306.
- Schadler, M., & Thissen, D. M. (1981). The development of automatic word recognition and reading skill. *Memory & Cognition, 9*, 132–141.
- Scheibe, K. E., Shaver, P. R., & Carrier, S. C. (1967). Color association values and response interference on variants of the Stroop test. *Acta Psychologica, 26*, 286–295.
- Schiller, P. H. (1966). Developmental study of color-word interference. *Journal of Experimental Psychology, 72*, 105–108.

- Schmidt, J. R., & Cheesman, J. (2005). Dissociating stimulus-stimulus and response-response effects in the Stroop task. *Canadian Journal of Experimental Psychology*, *59*, 132–138.
- Schmidt, J. R., & Besner, D. (2008). The Stroop effect: why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 514–523.
- Seymour, P. H. (1974). Stroop interference with response, comparison, and encoding stages in sentence-picture comparison task. *Memory & Cognition*, *2*, 19–26.
- Seymour, P. H. (1977). Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology*, *29*, 245–265.
- Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition*, *26*, 1033–1040.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.
- Simpson, A., Riggs, K. J., Beck, S. R., Gorniak, S. L., Wu, Y., Abbott, D., & Diamond, A. (2012). Refining the understanding of inhibitory control: How response prepotency is created and overcome. *Developmental Science*, *15*, 62–73.
- Simpson, A., & Carroll, D. J. (in press). Understanding early inhibitory development: distinguishing two ways that children use inhibitory control. *Child Development*.
- Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 461–479.
- Stanovich, K. E., Cunningham, A. E., & West, R. F. (1981). A longitudinal study of the development of automatic recognition skills in first graders. *Journal of Reading Behavior*, *13*, 57–74.

- Stirling, N. (1979). Stroop interference: An input and an output phenomenon. *Quarterly Journal of Experimental Psychology*, *31*, 121–132.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.
- Szücs, D., & Soltész, F. (2010). Stimulus and response conflict in the color-word Stroop task: A combined electro-myography and event-related potential study. *Brain Research*, *1325*, 63–76.
- Tse, C-H., & Neely, J. H. (2007). Semantic and repetition priming effects for Deese/Roediger-McDermott (DRM) critical items and associates produced by DRM and unrelated study lists. *Memory & Cognition*, *35*, 1047-1066.
- van Veen, V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: A functional MRI study. *NeuroImage*, *27*, 497–504.
- Verhaeghen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science*, *20*, 174–180.
- Verhaeghen, P., & De Meersman, L. D. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology and Aging*, *13*, 120–126.
- Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.
- Wolf, D., Zschuschke, L., Scheurich, A., Schmitz, F., Lieb, K., Tüscher, O., & Fellgiebel, A. (2014). Age-related increases in Stroop interference: Delineation of general slowing based on behavioral and white matter analyses. *Human Brain Mapping*, *35*, 2448–2458.
- Wright, B. C. (2017). What Stroop tasks can tell us about selective attention from childhood to adulthood. *British Journal of Psychology*, *108*, 583–607.
- Wright, B. C., & Wanley, A. (2003). Adults' versus children's performance on the Stroop task: Interference and facilitation. *British Journal of Psychology*, *94*, 475–485.

Zhang, H., & Kornblum, S. (1998). The effects of stimulus–response mapping and irrelevant stimulus–response and stimulus–stimulus overlap in four-choice Stroop tasks with single-carrier stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 3–19.

Ziegler, J. C., Bertrand, D., Lété, B., & Grainger, J. (2014). Orthographic and phonological contributions to reading development: tracking developmental trajectories using masked priming. *Developmental Psychology*, 50, 1026–1036.

[dataset] Ferrand, Ducrot, Chausse, Maïonchi-Pino, O’Connor, Parris, Perret, Riggs, &

Augustinova (2019). Stroop interference is a composite phenomenon. <https://osf.io/69fkj/>;

Doi: 10.17605/OSF.IO/69FKJ

## Table legends

**Table 1.** Demographic Variables and Mean Reading Scores as a Function of Age-group in Experiment 1

**Table 2.** Color-Identification Performance in  $z$ RTs observed as a Function of Stimulus- or Interference-Type and Age-group in Experiment 1 using Manual Responses

**Table 3.** Demographic Variables and Mean Reading Scores as a Function of Age-group in Experiment 2

**Table 4.** Color-Identification Performance in  $z$ RTs observed as a Function of Stimulus-, Interference- or Conflict-Type and Age-group in Experiment 2 using Vocal Responses

**Table S1.** Color-Identification Performance in raw RTs observed as a Function of Stimulus and Age-group in Experiment 1 using Manual Responses

**Table S2.** Color-Identification Performance in raw RTs observed as a Function of Stimulus and Age-group in Experiment 2 using Vocal Responses

## Figure legends

**Figure 1.** The subtractive logic of the extended semantic Stroop interference applied to the data from a vocal naming task (see e.g., Augustinova, Silvert, Spatola, & Ferrand, 2018; Experiment 2). Significant differences in color-naming RTs are thought to result from the selective presence of different types of conflict across these four types of items.

**Figure 2.** Magnitudes of the overall Stroop interference observed as a function of Age-group respectively in Experiment 1 (using manual responses) and 2 (using vocal responses).

**Figure 3.** Magnitudes of different conflicts involved in the overall Stroop interference observed as a function of Age-group in Experiment 2 (using vocal responses).

Table 1. *Demographic Variables and Mean Reading Scores as a Function of Age-Group in Experiment 1*

<b>Age-Group</b>	<b>N</b>	<b>Gender</b> (boys/girls)	<b>Chronological Age</b> (in months)	<b>Reading Age</b> (in months)
1 (6-7 years old)	76	39/37	83 (5)	82 (6)
3 (8-9 years old)	81	41/40	106 (7)	103 (9)
5 (10-11 years old)	61	32/29	130 (10)	125 (10)

*Note.* Standard Deviations are reported in parentheses.

Table 2

*Color-Identification Performance in zRTs observed as a Function of Stimulus- or Interference-Type and Age-Group in Experiment 1 using Manual Responses*

	1 <sup>st</sup> -graders (6-7 years old)			3 <sup>rd</sup> -graders (8-9 years old)			5 <sup>th</sup> -graders (10-11 years old)		
	M in zRT (SE)	CI	%ER	M in zRT (SE)	CI	%ER	M in zRT (SE)	CI	%ER
<b>Standard Color-Incongruent Words</b> ( <i>BLUE</i> <sub>green</sub> )	.052 (.023)	[.007, .098]	5.9	.091 (.025)	[.043, .14]	5.6	.074 (.026)	[.023, .126]	4.5
<b>Color-Neutral Letter-String</b> ( <i>XXXX</i> <sub>green</sub> )	-.048 (.022)	[-.090, -.005]	3.6	-.072 (.021)	[-.114, -.030]	3.5	-.067 (.024)	[-.114, -.020]	3.4
	M in zRT (SE)	CI		M in zRT (SE)	CI		M in zRT (SE)	CI	
<b>Overall Stroop Interference</b> ( <i>BLUE</i> <sub>green</sub> - <i>XXXX</i> <sub>green</sub> )	.101* (.045)	[.013, .188]		.164** (.045)	[.075, .253]		.142** (.050)	[.043, .240]	

Note. \*significant at  $p < .05$ ; \*\*significant at  $p < .01$



Table 3. *Demographic Variables and Mean Reading Scores as a Function of Age-Group in Experiment 2*

<b>Age-Group</b>	<b>N</b>	<b>Gender</b> (boys/girls)	<b>Chronological Age</b> (in months)	<b>Reading Age</b> (in months)
1 (6-7 years old)	53	25/28	77 (4)	80 (4)
3 (8-9 years old)	46	25/21	100 (4)	112 (20)
5 (10-11 years old)	38	11/27	124 (4)	134 (25)

*Note.* Standard deviations are reported in parentheses.

Table 4

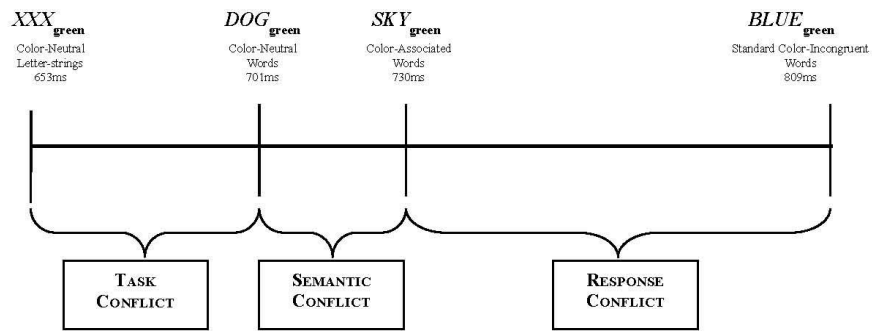
*Color-Identification Performance in zRTs observed as a Function of Stimulus-, Interference- or Conflict-Type and Age-Group in Experiment 2 using Vocal Responses*

	1 <sup>st</sup> -graders (6-7 years old)			3 <sup>rd</sup> -graders (8-9 years old)			5 <sup>th</sup> -graders (10-11 years old)		
	M in zRT (SE)	CI	%ER	M in zRT (SE)	CI	%ER	M in zRT (SE)	CI	%ER
<b>Standard Color-Incongruent Words</b> ( <i>BLUE</i> <sub>green</sub> )	.114 (.039)	[.037, .190]	3.74	.410 (.042)	[.328, .492]	9.13	.418 (.046)	[.328, .508]	9.20
<b>Associated Color-Incongruent Words</b> ( <i>SKY</i> <sub>green</sub> )	.088 (.025)	[.038, .138]	0.86	.038 (.027)	[-.016, .091]	0.92	.029 (.030)	[-.030, .088]	1.32
<b>Color-Neutral Words</b> ( <i>DOG</i> <sub>green</sub> )	.032 (.024)	[-.015, .080]	0.79	-.055 (.026)	[-.106, -.004]	1.56	-.069 (.028)	[-.126, -.013]	1.03
<b>Color-Neutral Letter-String</b> ( <i>XXX</i> <sub>green</sub> )	-.234 (.033)	[-.300, -.168]	0.67	-.393 (.036)	[-.464, -.322]	0.45	-.378 (.039)	[-.455, -.300]	0.30
	M in zRT (SE)	CI		M in zRT (SE)	CI		M in zRT (SE)	CI	
<b>Overall Stroop Interference</b> ( <i>BLUE</i> <sub>green</sub> - <i>XXX</i> <sub>green</sub> )	<b>.348**</b> (.064)	[.222, .474]		<b>.803**</b> (.068)	[.668, .938]		<b>.796**</b> (.075)	[.647, .944]	
resulting from the contribution of:									
<b>Task conflict</b> ( <i>DOG</i> <sub>green</sub> - <i>XXX</i> <sub>green</sub> )	.266** (.045)	[.176, .356]		.338** (.049)	[.241, .434]		.308** (.054)	[.202, .414]	

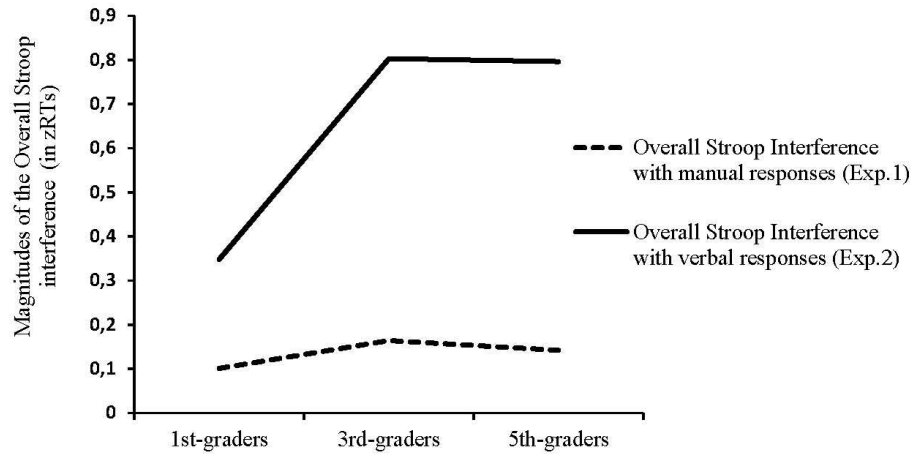
<b>Semantic conflict</b> ( $SKY_{green} - DOG_{green}$ )	.056 <sup>ns</sup> (.035)	[-.014, .126]	.093 <sup>*</sup> (.038)	[.018, .168]	.098 <sup>*</sup> (.042)	[.016, .181]
<b>Response conflict</b> ( $BLUE_{green} - SKY_{green}$ )	.026 <sup>ns</sup> (.054)	[-.082, .133]	.372 <sup>**</sup> (.058)	[.257, .488]	.389 <sup>**</sup> (.064)	[.262, .516]

---

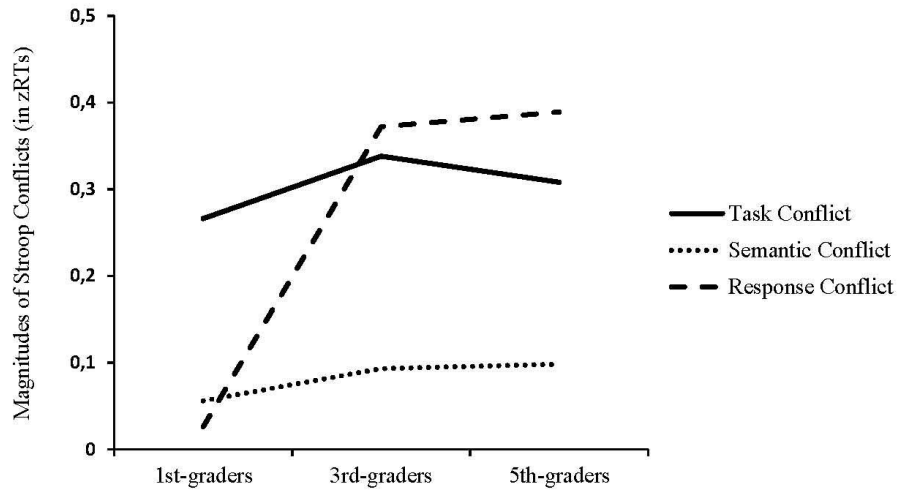
Note. <sup>ns</sup> non-significant <sup>\*</sup> significant at  $p < .05$ ; <sup>\*\*</sup> significant at  $p < .01$



**Figure 1.** The subtractive logic of the extended semantic Stroop interference applied to the data from a vocal naming task of Augustinova, Silvert, Spatola, and Ferrand (2018; Experiment 2). Significant differences in color-naming RTs are thought to result from the selective presence of different types of conflict across these four types of items.



**Figure 2.** Magnitudes of the overall Stroop interference observed as a function of age-group respectively in Experiment 1 (using manual responses) and 2 (using vocal responses).



**Figure 3.** Magnitudes of different conflicts involved in the overall Stroop interference observed as a function of age-group in Experiment 2 (using vocal responses).