

**Some further clarifications on age-related differences in the Stroop task: new evidence
from the two-to-one Stroop paradigm**

Mariana Burca¹, Pierre Chausse², Ludovic Ferrand², Benjamin A. Parris³ and Maria
Augustinova¹

¹Université de Rouen Normandie, CRFDP, Rouen, France

²Université Clermont Auvergne, CNRS, LAPSCO, Clermont-Ferrand, France

³Department of Psychology, Bournemouth University, Poole, UK

Running head: Age-related differences in the Stroop task

Type of article: Brief Report

Word count: 3665 (Introduction, Results and Discussion)

Contact address:

Mariana Burca or Maria Augustinova
Université de Rouen Normandie
Le Centre de Recherches sur les Fonctionnements et Dysfonctionnements Psychologiques
(CRFDP, EA 7475),
UFR SHS – Bâtiment Freinet, Place Emile Blondel, 76821 Mont Saint Aignan Cedex,
FRANCE
E-mail: mariana.burca2@univ-rouen.fr maria.augustinova@univ-rouen.fr
Phone: + 33 473 406 257
Fax: + 33 473 406 114

Funding Information:

The work reported was supported by ANR Grant ANR-19-CE28-0013 and RIN Doctorant et RIN Tremplin Grant 19E00851 of Normandie Région, France.

Abstract

Previous studies (Augustinova et al., 2018; Li & Bosman, 1996) have shown that the larger Stroop effects reported in older adults is specifically due to age-related differences in the magnitude of response, and not semantic, conflict, both of which are thought to contribute to overall Stroop interference. However, the most recent contribution to the issue of the unitary vs. composite nature of the Stroop effect (Parris et al., 2021) argues that semantic conflict has not been clearly dissociated from response conflict in these or any other past Stroop studies, meaning that the very existence of semantic conflict is at present uncertain. To distinguish clearly between the two types of conflict, the present study employed the two-to-one Stroop paradigm (De Houwer, 2003) with a color-neutral word baseline (Hasshim & Parris, 2014). This addition made it possible to isolate a contribution of semantic conflict that was independent of both response conflict and Stroop facilitation. Therefore, this study provides the first unambiguous empirical demonstration of the composite nature of Stroop interference – as originally claimed by multi-stage models of Stroop interference (Zhang & Kornblum, 1998; Zhang et al., 1999). This permitted the further observation of significantly higher levels of semantic conflict in older adults, whereas the level of response conflict in the present study remained unaffected by healthy aging – a finding that directly contrasts with previous studies employing alternative measures of response and semantic conflict. Two qualitatively different explanations of this apparent divergence across studies are discussed.

Key words: Stroop interference, Aging, Semantic conflict, Response conflict

The *Stroop interference effect* (i.e., longer color-identification times for color-incongruent [e.g., “BLUE” displayed in yellow] than for color-neutral words [e.g., the word “DEAL” displayed in yellow]) is generally larger in healthy older adults than in their younger counterparts (see Comalli et al., 1962, for the first empirical demonstration). Also, and importantly, this age-effect in the Stroop task (Stroop, 1935) persists even after controlling for differences in processing-speed (e.g., Aschenbrenner et al., 2015, 2017; Bugg et al., 2007; Jackson & Balota, 2013; Nicosia & Balota, 2020; Spieler et al., 1996). It is therefore thought to reflect an inhibition deficit (e.g., Hasher & Zacks, 1988) due to which older adults are less efficient at suppressing the word-dimension of color-incongruent Stroop words, leading them to experience greater competition at the response output stage (Spieler et al., 1996).

Indeed, according to dominant single-stage response competition models (e.g., Roelofs, 2003), incidental semantic processing of the irrelevant word-dimension of color-incongruent Stroop items generates a single type of conflict: response conflict. According to this view, the Stroop interference effect is considered a unitary phenomenon generated solely due to competition between two alternative responses indicated by the two dimensions of the Stroop stimulus. In contrast, *multi-stage* models anticipate this incidental processing to generate an additional level of conflict at the level of semantics: *semantic* conflict (e.g., Zhang et al., 1999; Zhang & Kornblum, 1998). They therefore view the Stroop interference effect as a composite phenomenon comprising both response and semantic conflict.

Taking this idea as their starting point, several studies have set out to investigate the level of processing (e.g. response and/or semantic) at which the age-related differences in the Stroop task take their effects and, more specifically, whether semantic conflict is or is not affected by healthy aging. Indeed, the idea proffered by Spieler and colleagues that older adults are less efficient in suppressing the word-dimension of Stroop stimuli leads to the somewhat straightforward prediction that they should (also) experience a greater amount of semantic

conflict. This is not what studies have found.

Li and Bosman (1996) and, later, Augustinova and colleagues (2018) reported greater magnitudes of standard Stroop interference (e.g., $BLUE_{yellow} - DEAL/****_{yellow}$) in healthy older adults, but neither study reported age-related differences in the magnitude of semantic-associative Stroop interference (e.g., $SKY_{yellow} - DEAL/****_{yellow}$)¹. Augustinova and colleagues (2018) subsequently claimed that the locus of the age-effect in the Stroop task is at the level of response conflict rather than the level of semantic conflict or a combination of the two. Contrary to past conceptualizations (e.g., Spieler et al., 1996), these results imply that both older and younger participants are actually equally (in)efficient at suppressing the word-dimension of Stroop stimuli. In line with the most recent contributions to the literature on the above-mentioned inhibition deficit (e.g., Rey-Mermet & Gade, 2018), it further implies that older participants are rather less efficient in inhibiting the irrelevant response that is primed by the (irrelevant) word-dimension. This in turn reinforces the idea that the age-related deficit in inhibition (e.g., Andrés et al., 2008), or, more broadly, the age-related deficit in cognitive control, is not general (e.g., Bugg, 2014).

However, single-stage response competition models argue that semantic-associative interference ($SKY_{yellow} - DEAL_{yellow}$) measured in these prior studies results entirely from response conflict (e.g., Roelofs, 2003). According to this position, semantic associates indirectly elicit incorrect *response* activity (e.g., say “blue”/press blue for SKY_{yellow}) through their association with the response-set colors (blue in this case) which explains in turn the smaller magnitude of semantic-associative interference ($SKY_{yellow} - DEAL_{yellow}$) compared to its standard ($BLUE_{yellow} - DEAL_{yellow}$) counterpart. Under this account, neither Li and Bosman’s (1996) nor Augustinova and colleagues’ (2018) studies satisfactorily demonstrated that

¹ To control for differences in processing-speed, raw naming latencies were proportionally transformed in these studies into percentages of standard ($[(M_{standard\ color-incongruent\ RT} - M_{color-neutral\ RT}) / M_{color-neutral\ RT}] * 100$) and semantic Stroop interference ($[(M_{color-associated\ incongruent\ RT} - M_{color-neutral\ RT}) / M_{color-neutral\ RT}] * 100$).

semantic conflict is spared by healthy aging (i.e., due specifically to a slowdown that occurs whenever two distinct, yet closely related semantic representations are simultaneously activated in an amodal semantic network, see e.g. Seymour, 1977 for discussion).

To address this issue directly, the present study replaced semantic-associative items with items which induce semantic conflict in a way that cannot be accounted for by single-stage response competition models. Specifically, the study employed the *two-to-one Stroop paradigm* (De Houwer, 2003, hereafter 2:1). In this paradigm, all the distractors are part of the response set (e.g., BLUE, RED, GREEN, YELLOW), while responses for paired target colors are mapped to only one response-key (e.g., ‘f’ for blue and red and ‘j’ for green and yellow). As a result of this response-mapping, trials like *BLUE*_{red} only involve semantic conflict (since both dimensions of the Stroop stimulus provide evidence toward the same response; therefore termed same-response trials), whilst trials like *BLUE*_{yellow} involve both response and semantic conflict (which are therefore termed *different-response* trials), and represent the standard incongruent Stroop trial. Indeed, same-response trials cannot generate (directly or indirectly) any response conflict because of this response mapping manipulation. Consequently, significant *interference* generated by same-response trials is interpreted as representing the independent contribution of semantic conflict to overall Stroop interference (De Houwer, 2003, see e.g., Hershman & Henik, 2020 for the most recent example).

However, with the exception of a few notable studies (see below), all studies employing this measure of semantic conflict – including De Houwer (2003) – have used color-congruent trials as the baseline against which semantic conflict is measured. Problematically, the difference between same-response and color-congruent trials could be entirely driven by facilitation on color-congruent trials and thus not involve any semantic conflict (Hasshim and Parris 2014, 2015) – as unitary models of Stroop interference (Roelofs, 2003) would predict. In line with this interpretation, Hasshim and Parris consistently reported significantly longer RTs

for same-response trials than for color-congruent trials, but no difference between same-response trials and trials that were free of facilitation (i.e., color-neutral word trials, see e.g., Brown, 2012; MacLeod, 1991 for discussion)².

In contrast to Hasshim and Parris, Burca and colleagues' study (in revision) reported a significant difference between same-response and color-neutral trials. This suggests that the difference between same-response and color-congruent trials (i.e., when no color-neutral baseline is included) simply confounds the (*semantic*) *conflict* produced by same-response trials and *facilitation* produced by color-congruent trials (MacLeod, 1991). However, the extent to which this is actually the case remains uncertain, since Burca et al.'s study did not include color-congruent trials. As a result, no study has so far demonstrated that semantic conflict contributes to overall Stroop interference in the 2:1 Stroop paradigm independently of both response conflict and facilitation. Considering this as a necessary prerequisite for any empirical demonstration of the specific age-effect (or lack thereof) on semantic vs. response conflict in the Stroop task, the present study aimed to address this more fundamental issue.

To this end, items that are traditionally included in the 2:1 Stroop paradigm (De Houwer, 2003) were supplemented by color-neutral word trials (Hasshim & Parris, 2014). This addition enabled us to test adequately for the presence of semantic conflict predicted by the multi-stage models of Stroop interference (e.g., Zhang & Kornblum, 1998) that were favored a priori in the current study over the still-dominant single-stage response competition models (e.g., Roelofs, 2003). With this design, the study was able to more unambiguously measure age-related differences in response and semantic conflict. Consequently, if, as reported by past studies (Augustinova et al., 2018; Li & Bosman, 1996), semantic conflict (same-response trials – color-neutral trials) is indeed spared in healthy aging, its magnitude will not differ between

² The absence of semantic conflict was supported further by Bayesian evidence for the null-hypothesis and by the unchanged magnitude of associated pre-response pupillometric measures of effort (i.e., a reliable measure of the potential differences between conditions, Hasshim & Parris, 2015).

young and old adults. In contrast, response conflict (different-response – same-response trials) will be greater in healthy aging adults as compared to their younger counterparts.

Method

Participants and Design

Fifty-one older (i.e., over 65 years old) and fifty younger (i.e., below 35 years old) native French-speakers reporting normal or corrected-to-normal vision and presenting no impairment in color discrimination initially volunteered to participate in the study approved by the local ethics committee. One older participant presented a medical history that included a head injury and one other was undergoing a medical treatment for depression. Six months prior to inclusion in the study, none of the other participants suffered from other psychiatric and/or neurological disorders. None of them declared taking any drug and/or following any medical treatment that is known to impact the nervous system during the 48h prior to inclusion. To ensure that the remaining participants fitted the inclusion criteria, they completed a psychometric evaluation battery. To this end, the older adults completed the Mini Mental State Examination (Folstein, 1975). The scores of two participants were lower than the cutoff score of 25 points. The older adults also completed the Frontal Assessment Battery (Dubois et al., 2000). None of them presented with a cutoff score of 16 (or 15 depending on the participant's sociocultural level). A depression scale was then administered to both the older and younger adults. No older adults reached the cutoff score of 7 on the short version (15 items) of the Geriatric Depression Scale (Sheikh & Yesavage, 1986). In addition, none of the younger adults reached the cutoff score of 8 on Beck's Depression Inventory (Beck, 1988). In both groups, working memory was assessed with the forward and backward digit span (WAIS, Wechsler, 2008). All participants had scores within the norm, recalling seven plus or minus two items.

Finally, to further assess differences in processing speed, the French equivalent (Bugajska et al., 2007) of the letter-comparison test (Salthouse, 1990) was administered in both age-groups. After the exclusion of 5 participants in total (one was unable to perform the manual 2:1 Stroop task due to reduced hand mobility), the Stroop data of forty-six healthy older (36 females and 10 males; $M_{\text{age}} = 74.04$) and fifty younger adults (41 females and 9 males; $M_{\text{age}}=21.48$) were analyzed in a 4 (Stimulus-Type: different-response vs. same-response vs. neutral vs. congruent) \times 2 (Age-Group: older vs. younger) ANOVA, with the former factor as within-participants factor.

Apparatus, Stimuli and Procedure

After the psychometric evaluation presented above, the participants completed a computerized version of the Stroop Task run using Eprime 2.0 software (Schneider et al., 2002). The participants were seated 70cm in front of a 13" portable computer and instructed to identify the color of the stimulus presented on the screen, as quickly and accurately as possible, by pressing the appropriate color-button and to ignore everything else in the display. To this end, they were instructed to concentrate on the fixation cross (“+”) that appeared for 2000msec in the center of the screen at the beginning of each trial. The stimulus remained on the screen until the participant responded or until 3500msec had elapsed.

All stimuli were presented in lowercase Courier font, size 18, on a black background and subtended an average visual angle of 0.9° high \times 3.0° wide. The participants responded manually using a modified SRBox® consisting of two handles, each of which had a single response button at the top flanked by two color-stickers (blue and red on one handle, yellow and green on the other, see Supplementary Materials). The participants pushed these response buttons with their thumbs. This allowed them to hold each handle comfortably in their palms with the remaining four fingers. The placement of the handles in the right or left hand, respectively, was counterbalanced across participants.

To familiarize themselves with the color-button correspondence before completing the experimental block, the participants first completed 96 practice trials consisting of asterisks. Due to the low accuracy rate, 8 older participants had to repeat this practice block (3 of them were later excluded from further analyses) before proceeding to the experimental trials. As in Hasshim and Parris (2014, Exp. 2A), these consisted of 96 different-response, 48 same-response, 48 color-neutral and 48 color-congruent trials. The trials were randomly intermixed in a single block. To this end, four (French) color-words: *rouge* [red], *jaune* [yellow], *bleu* [blue], and *vert* [green] presented in both congruent and incongruent colors and four non-color words: *plomb* [lead], *liste* [list], *page* [page], *cave* [basement] presented in all the colors were used. They were paired on length and frequency via *Lexique 3.38* (New et al., 2004).

Results and Discussion

Five older participants were excluded from further analyses: one due to faulty recording, and the four others due to the fact that more than 33% of their data was removed from the analysis after the 3SD correction and the exclusion of the wrong answers (see Table 1S in the Supplementary Materials for demographic and psychometric data of the remaining participants). *RTs* greater than 3SDs above or below each participant's mean latency for each condition were excluded from the analysis (i.e., less than 2% of the total data, corresponding to 0.9% of younger adults' data and 1.5% of older adults' data). Consequently, *RTs* and errors of the remaining 91 participants (41 older and 50 younger) were first analyzed in an omnibus 4 (Stimulus-Type: different-response vs. same-response vs. neutral vs. congruent) \times 2 (Age-Group: older vs. younger) standard and Bayesian ANOVA. The values for this latter ANOVA were calculated with JASP (JASP Team, 2017) and interpreted according to Lee and Wagenmakers (2013 adjusted from Jeffreys, 1961). All priors were equal. Recall that further reported BF_{10} is the Bayes factor giving the evidence for H1 over the null hypothesis (H0),

whereas BF_{01} is evidence for H_0 over H_1 .

For errors (see Table 1), these analyses revealed a main effect of Stimulus-Type [$F(3,267) = 19.03$; $p < .001$, $\eta_p^2 = 0.176$; $BF_{10} = 4.450e+7$], but not of Age-Group [$F(1,89) = .018$; $p = .894$, $\eta_p^2 < .000$; $BF_{10} = 0.227/BF_{01} = 4.396$]. The Stimulus Type \times Age-Group interaction was also significant [$F(3,267) = 3.11$; $p = .041$, Greenhouse-Geisser corrected, $\eta_p^2 = 0.034$; $BF_{10} = 1.130/BF_{01} = 0.884^3$]. However, the BF evidence in favor of an interaction was only anecdotal⁴.

Given that the analysis of *RTs* showed a considerable but expected (see Table 2S in the Supplementary Materials) general slowing in older adults (i.e., the significant Stimulus-Type \times Age-Group interaction [$F(3, 267) = 14.78$; $p < .001$; $\eta_p^2 = 0.142$; $BF_{10} = 1.378e+6$]), which was qualified by a significant simple main effect of Age-Group for each type of Stimulus (all p 's $< .001$, see Table 2S in the Supplementary Materials), these *RTs* were z-scored (e.g., Jackson & Balota, 2013). The same omnibus ANOVA then revealed a main effect of Stimulus-Type [$F(3,267) = 128.59$; $p < .001$, $\eta_p^2 = 0.591$; $BF_{10} = 1.459e+59$] that was also included in the significant Stimulus-Type \times Age-Group interaction [$F(3,267) = 10.36$; $p < .001$, $\eta_p^2 = 0.104$, $BF_{10} = 706286.31$], thus indicating that age-related differences persist even after controlling for generalized slowing (see Table 1).

Is there any Semantic Conflict in the two-to-one Stroop paradigm?

³ Since the different models that a Bayesian ANOVA compares against a null model never include interaction alone, the BF values reported for all the interactions correspond to values obtained by dividing the BF value of the model containing the two main effects and their interaction by the BF value of the model with the two main effects only.

⁴ The simple-main effect of Stimulus-Type was significant in both older [$F(3,87) = 9.09$; $p < .001$, $\eta_p^2 = 0.239$; $BF_{10} = 66195.29$] and younger [$F(3,87) = 3.26$; $p = .025$, $\eta_p^2 = 0.101$; $BF_{10} = 35.97$] participants. Unsurprisingly, and in line with the main effect of Stimulus-Type, further pairwise comparisons revealed that the most errors in both age-groups were committed for different-response incongruent items (see Table 1 for descriptive statistics and simple main-effects of Age-group). However, in the younger adults, %ER for these latter items differed only marginally from those observed for color-congruent ones – yielding only a marginally significant overall Stroop effect ($M_{\text{different-response}} - M_{\text{congruent}}$) on errors ($p = .062$; $BF_{10} = 1.751/BF_{01} = 0.571$).

To answer this key question, we first analyzed the aforementioned main effect of Stimulus-Type. This analysis revealed that, as in De Houwer's original study, the total Stroop effect ($M_{\text{different-response}} - M_{\text{congruent}}$, $p < .001$; $BF_{10} = 1.814e+24$) resulted from a significant contribution of both response conflict ($M_{\text{different-response}} - M_{\text{same-response}}$; $p < .001$; $BF_{10} = 4.134e+11$) and the difference between same-response and congruent trials ($p < .001$; $BF_{10} = 2.880e+10$) – taken in previous studies as evidence for semantic conflict. However, the crucial addition of color-neutral trials enabled us to show that, overall, this latter difference did indeed confound the contribution of semantic conflict ($M_{\text{same-response}} - M_{\text{neutral}}$; $p < .001$; $BF_{10} = 27038.729$) and that of Stroop facilitation ($M_{\text{neutral}} - M_{\text{congruent}}$, $p = .016$), which was moderate ($BF_{10} = 7.835$). This finding is consistent with MacLeod's reasoning (1991) that in the absence of color-neutral trials, the total Stroop effect ($M_{\text{different-response}} - M_{\text{congruent}}$) is likely to confound two qualitatively distinct phenomena: the Stroop interference ($M_{\text{different-response}} - M_{\text{neutral}}$) and facilitation ($M_{\text{neutral}} - M_{\text{congruent}}$) effects.

The decomposition of the Stimulus-Type \times Age-Group interaction further revealed that the simple main effect of Stimulus-Type was significant in both older [$F(3,87) = 76.86$; $p < .001$, $\eta_p^2 = 0.726$; $BF_{10} = 1.876e+33$] and younger [$F(3,87) = 35.65$; $p < .001$, $\eta_p^2 = 0.551$; $BF_{10} = 3.019e+26$] participants. Further pairwise comparisons conducted in both age-groups revealed that the significant total Stroop effect had the same structure, although excluding Stroop facilitation (see Table 1 for descriptive statistics and magnitudes), which was no longer significant in younger adults ($p = .114$, $BF_{10} = 0.621/BF_{01} = 1.610$). The Stroop interference effect – which was significant in both age-groups (young group: $p < .001$; $BF_{10} = 5.746e+10$; older group: $p < .001$; $BF_{10} = 3.195e+10$) – again resulted from the significant contribution of semantic ($M_{\text{same-response}} - M_{\text{neutral}}$) and response ($M_{\text{different-response}} - M_{\text{same-response}}$) conflicts (see Table 1)⁵.

⁵ It should be noted, however, that Bayesian evidence for semantic conflict in younger adults remained anecdotal ($BF_{10} = 1.893/BF_{01} = 0.528$) despite the fact that moderate evidence in support of such a conflict was found in a recent study by our research group (Burca et al., in revision)

Taken together, these results are therefore consistent with the idea that both semantic conflict and response conflict contribute to Stroop interference. This pre-requisite being satisfied (see Introduction), we can now go on to investigate the extent to which these independent components of Stroop interference are influenced by healthy aging.

<Insert Table 1 about here>

How does healthy aging influence semantic vs. response conflict in the Stroop task?

To address this issue, the magnitudes of semantic and response conflicts (see Table 1) were analyzed in a 2 (Conflict-Type) \times 2 (Age-Group: older vs. younger) ANOVA. This revealed a non-significant main effect of Conflict-Type [$F(1,89)=1.13$; $p=.292$, $\eta_p^2=0.012$; $BF_{10}=0.481/BF_{01}=2.078$] as well as a significant [$F(1,89)=11.94$; $p=.001$, $\eta_p^2=.118$], although anecdotal ($BF_{10}=1.529/BF_{01}=0.654$), main effect of Age-Group. It also revealed a marginally significant [$F(1,89)=3.38$; $p=.069$, $\eta_p^2=0.037$], although anecdotal ($BF_{10}=2.330/BF_{01}=0.429$), Conflict-Type \times Age-Group interaction. Even though evidence for this interaction was only anecdotal, we decomposed it further by testing the simple main effect of Age-Group at each level of Conflict-Type. Contrary to our expectations, this effect was significant for semantic conflict [$F(1,89)=9.288$; $p=.003$, $\eta_p^2=0.094$; $BF_{10}=11.683/BF_{01}=0.086$], with older adults presenting a much greater magnitude of semantic conflict than young adults. Additionally, and also contrary to our expectations, the simple main effect of Age-Group remained non-significant for response conflict [$F(1,89)=0.010$; $p=.922$, $\eta_p^2=0.000$; with evidence for the null effect of aging, $BF_{10}=0.222/BF_{01}=4.512$ (see Table 1)]⁶. Thus, the present study clearly extends the dissociative nature of the age-effect to the two-to-one Stroop paradigm. However,

⁶ Additionally, the simple main effect of Conflict-Type was significant in younger adults [$F(1,89)=4.67$; $p=.034$, $\eta_p^2=0.050$; $BF_{10}=9.318$] but not in older adults [$F(1,89)=0.275$; $p=.602$, $\eta_p^2=0.003$; $BF_{10}=0.278/BF_{01}=3.597$]. The additional $BF_{+0}=3.338$ in younger adults indicates that they displayed more response than semantic conflict, whereas $BF_{01}=5.298$ in older adults indicates comparable magnitudes of both conflicts (see Table 1).

completely unlike past studies using the semantic Stroop paradigm (Augustinova et al., 2018; Li & Bosman, 1996), it points to a greater magnitude of semantic conflict in older adults.

General Discussion and Conclusion

Given that in all past Stroop studies, *semantic* conflict was potentially confounded with either *response conflict* (e.g., when semantic-associative items [*SKY*_{blue}] are used to induce semantic conflict) or with *facilitation* (when color-congruent items [*BLUE*_{blue}] are used as a baseline to derive a magnitude for semantic conflict), its contribution to the Stroop interference effect has so far been uncertain. Using the 2:1 Stroop paradigm (De Houwer, 2003) with a color-neutral baseline, the present study clearly demonstrated that the contribution of semantic conflict is independent of both response conflict and Stroop facilitation. Therefore, the present study provides an unambiguous empirical basis for the composite nature of Stroop interference – as originally claimed by De Houwer (2003) based on the multi-stage models of Stroop interference (Zhang & Kornblum, 1998; Zhang et al., 1999)⁷.

Given that no such basis was available in past studies of age-related differences in the Stroop task (Augustinova et al., 2018; Li & Bosman, 1996), the present study also investigated the extent to which healthy aging influences these independent constituents of Stroop interference. The reported results suggest a dissociative pattern opposite to that reported in past studies: whilst response conflict was not affected by healthy aging, greater semantic conflict was found in older adults.

It remains possible that this reverse pattern is due to the fact that the present study mobilized different processes from those at work in past studies. Indeed, both Augustinova et al. (2018) and Li and Bosman (1996) employed a vocal response, which is known to induce

⁷Note that the unambiguous presence of Stroop facilitation additionally implies that magnitudes of semantic conflict observed without color-neutral baseline are clearly inflated. This also concerns magnitudes of a general conflict – central in cognitive control studies (e.g. Egner et al., 2010) – since this type of conflict is inferred from the so-called Stroop congruency effect (*BLUE*_{yellow}–*BLUE*_{blue}) using the same color-congruent baseline.

greater phonological processing of the irrelevant word than a manual one (Kinoshita et al., 2017; Parris et al. 2019). Therefore, the pattern that these studies report could be due to less efficient control of this phonological processing in older adults. Such an effect would not have been observed in the present study due to the use of manual responses. Despite this, the issue surrounding the use of semantic-associative Stroop trials remains.

Indeed, we have shown above that according to single-stage models of the Stroop task, the semantic associative Stroop trials used in these previous studies induce only indirect response conflict (e.g., Roelofs, 2003). This would then mean that the only conclusion that can be drawn from the studies by Augustinova et al. (2018) and Li and Bosman's (1996) is that overall response conflict is greater in older adults but that some still remains (as measured indirectly by semantic associate trials) unaffected by healthy aging. However, since the present study unequivocally documented the existence of semantic conflict for the first time, it is possible semantic-associative trials actually induce a level of semantic conflict, because we now know there is semantic conflict to be detected, but in unknown quantities.

If we assume that the present and past studies mobilized the same processes (i.e., induced comparable levels of semantic conflict; Augustinova & Ferrand, 2014), the absence of an age-effect on semantic associative interference could be potentially linked to the method used to control for age-related general slowing. Indeed, proportional transformation – applied first by Li and Bosman (1996) and later by Augustinova et al. (2018) – might actually (and counterintuitively) create an advantage for older adults in the presence of slower RTs (Hedge et al., 2018). This spurious advantage is no longer present when general slowing is controlled by means of a more suitable transformation (i.e., z-scores; Faust et al., 1999; Hedge et al., 2018) applied in the present study. To address this possibility directly, the data from Augustinova et al. (2018) were z-scored and re-analyzed in the same way as the 2:1 data reported above (see Supplementary Materials for a full description and results of these analyses, pp.4-9). In line

with Hedge et al.'s reasoning about proportional transformation, not only did the originally significant Conflict-Type \times Aging interaction become non-significant, but the additional Bayesian analyses actually provided moderate evidence against this interaction. This suggests that the magnitudes of both semantic and response conflict in Augustinova et al.'s z-scored data tended to be greater in older adults than in their younger counterparts (see Table 3S).

While the results regarding semantic conflict are in line with those reported above, discrepancies remain regarding the effect of healthy aging on response conflict. Although these differences could be accounted for by the response mode difference highlighted above, we also conducted cross-study analyses on the merged data sets (see Supplementary Materials for a full description and results of the analyses, pp.9-11). Again, Bayesian analyses provided moderate evidence against a Conflict-Type \times Aging interaction, suggesting that across two studies, healthy aging affected both the semantic and response conflicts. It should, however, be noted that a Bayesian Independent Samples t-Test conducted for exploratory purposes actually revealed anecdotal evidence against the age-effect on response conflict (see Table 4S), a finding that appears consistent with the results obtained using the 2:1 paradigm reported above. Alternatively, it also remains plausible that response conflict is unaffected in the 2:1 Stroop paradigm, not because of its specific nature but simply because its magnitude (i.e., smaller in the manual task than in the vocal tasks used in past studies) is too small to be affected.

Although not our favored a priori hypothesis, the fact that the present study could have mobilized different processes compared to past studies emphasizes the importance of choosing the correct critical and control trials for measuring the variable under test. Of course, no measure is perfect and we must therefore consider a limitation of the 2:1 paradigm that could provide an alternative explanation for the apparently greater semantic conflict in older adults. Because both dimensions of same-response trials provide evidence towards the same response, they cannot (unlike semantic associates) generate response conflict. However, they can still

produce response facilitation. This opens up the possibility that the larger difference between same-response and color-neutral trials observed in older adults in the present study could actually be driven by greater response *facilitation* in younger adults, and not greater semantic *conflict* in older adults. Importantly however this account would directly predict greater Stroop facilitation (which involves both response and semantic facilitation) in younger adults; the present study actually reports the opposite – rendering this latter account unlikely.

To sum up, the present study has provided the clearest evidence yet of a contribution of semantic conflict to overall Stroop interference (see Parris et al., 2021, for a thorough discussion of this issue). Moreover, this has enabled us to investigate the effect of healthy aging on the independent constituents of the composite Stroop interference effect. In contrast to previous studies, the present study showed that semantic conflict is affected by healthy aging. This finding prompted a re-analysis of the data from a previous study (Augustinova et al. 2018) using a more suitable method of controlling for the effect of general slowing in healthy aging (the same method as that employed in the present study). This re-analysis revealed that, as indicated by the present study, there is evidence of modified semantic conflict in healthy aging. Whilst the two studies diverge on the issue of the effect of aging on response conflict, the difference might be explained by the fact that a vocal response mode was used in both Augustinova et al. (2018) and Li & Bosman (1996), giving rise to the possibility that the control of phonological processing is reduced in healthy aging. Although both studies converged on the issue of semantic conflict, we would still recommend that future studies use the 2:1 paradigm rather than the semantic-associates method given that only the results from the present study show an unambiguous effect of aging on semantic conflict. However, to address the still-open issue of the characteristics shared (or otherwise) between same-response and semantically associated trials, future studies could combine the two (Schmidt & Cheesman, 2005) and measure the interference they generate against a color-neutral word baseline with

more response-sensitive measures (e.g., EMG, mouse-tracking). Given that these latter measures are also more sensitive to the actual time course of interference, they are particularly suitable for further addressing the age-related differences in the Stroop task. Indeed, the issue of the extent to which a greater magnitude of a given conflict is due specifically to its greater *activation* (i.e., lower attentional selectivity, also implying an age-related deficit in proactive control) or to its less efficient *resolution* (i.e., less efficient inhibitory control, also implying an age-related deficit in reactive control) as yet remains unresolved (see e.g. Coderre et al., 2011 for this type of distinction). In the light of past research demonstrating an age-related deficit in proactive (e.g., Braver et al., 2001) as opposed to reactive (e.g., Bugg, 2014) cognitive control, the first possibility seems more plausible than the second. This reasoning is reinforced by the fact that healthy aging might actually amplify task conflict (i.e., a more general conflict that – for all readable Stroop items including color-neutral ones – derives from the simultaneous preparation of two task sets: word-reading vs. color-naming, e.g., Goldfarb & Henik, 2007; Kalanthroff et al., 2018). Although, the significant age-effect on z-scored color-neutral stimuli observed in the present study is consistent with this idea, future studies – which should include more appropriate measures of task conflict – will need to address these possibilities directly.

The significant magnitudes of both semantic and response conflict observed in both younger and older adults clearly suggest that the historically favored single-stage response accounts of the Stroop interference effect are likely to be obsolete (e.g., Augustinova et al., 2018; De Houwer, 2003; Risko et al., 2006). Also, and importantly, so too are the customary implementations of Stroop interference/effect ($BLUE_{green} - DEAL_{green} / BLUE_{blue}$) that are rooted in these unitary models and from which the involvement of response and semantic processes and their modulation are merely inferred. Thus, in conclusion, the present study strongly encourages both the development of new integrative models of the Stroop interference effect (i.e., models that make room for relatively new types of conflict, e.g., Parris et al., 2021 for

discussion) and further empirical work addressing the processes underlying age-related differences in the Stroop task based on such integrative models.

References

- Andrés, P., Guerrini, C., Phillips, L. H., & Perfect, T. J. (2008). Differential Effects of Aging on Executive and Automatic Inhibition. *Developmental Neuropsychology*, *33*(2), 101-123. <https://doi.org/10.1080/87565640701884212>
- Aschenbrenner, A. J., & Balota, D. A. (2015). Interactive effects of working memory and trial history on Stroop interference in cognitively healthy aging. *Psychology and Aging*, *30*(1), 1-8. <https://doi.org/10.1037/pag0000012>
- Aschenbrenner, A. J., Balota, D. A., Weigand, A. J., Scaltritti, M., & Besner, D. (2017). The first letter position effect in visual word recognition : The role of spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(4), 700-718. <https://doi.org/10.1037/xhp0000342>
- Augustinova, M., Clarys, D., Spatola, N., & Ferrand, L. (2018). Some further clarifications on age-related differences in Stroop interference. *Psychonomic Bulletin & Review*, *25*(2), 767-774. <https://doi.org/10.3758/s13423-017-1427-0>
- Augustinova, M., & Ferrand, L. (2014b). Automaticity of Word Reading: Evidence from the Semantic Stroop Paradigm. *Current Directions in Psychological Science*, *23*, 343-348. <https://doi.org/10.1177/0963721414540169>
- Beck A.T. (1988). *Beck Hopelessness Scale*. The Psychological Corporation.
- Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., Janowsky, J. S., Taylor, S. F., Yesavage, J. A., Mumenthaler, M. S., Jagust, W. J., & Reed, B. R. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General*, *130*(4), 746-763. <https://doi.org/10.1037/0096-3445.130.4.746>
- Brown, T. L. (2011). The relationship between stroop interference and facilitation effects : Statistical artifacts, baselines, and a reassessment. *Journal of Experimental Psychology:*

- Human Perception and Performance*, 37(1), 85-99. <https://doi.org/10.1037/a0019252>
- Bugaiska, A., Clarys, D., Jarry, C., Tacconat, L., Tapia, G., Vanneste, S., & Isingrini, M. (2007). The effect of aging in recollective experience: The processing speed and executive functioning hypothesis. *Consciousness and Cognition*, 16(4), 797-808.
doi:10.1016/j.concog.2006.11.007
- Bugg, J. M. (2014). Evidence for the sparing of reactive cognitive control with age. *Psychology and Aging*, 29(1), 115–127. <https://doi.org/10.1037/a0035270>
- 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(2), 155-167.
<https://doi.org/10.1080/138255891007065>
- Burca M., Beaucousin V., Chausse P., Ferrand L., Parris B. A, Augustinova M. (in revision). *Is there semantic conflict in the Stroop task? Further evidence from the two-to-one Stroop paradigm combined with single letter coloring and cueing.*
- Coderre, E. L., Conklin, K., & van Heuven, W. J. B.(2011). Electrophysiological measures of conflict detection and resolution in the Stroop task. *Brain Research*, 1413, 51-59.
- 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(1), 47-53. <https://doi.org/10.1080/00221325.1962.10533572>
- De Houwer, J. (2003). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition*, 31(3), 353-359.
<https://doi.org/10.3758/BF03194393>
- Dubois, B., Slachevsky, A., Litvan, I., Pillon, B. (2000). The FAB: A Frontal Assessment Battery at bedside. *Neurology*, 55(11), 1621-1626. DOI: 10.1212/wnl.55.11.1621
- Egner, T., Ely, S. & Grinband, J. (2010). Going, going, gone: characterizing the time-course

- of congruency sequence effects. *Frontiers in Psychology*, 1:154.
<https://doi.org/10.3389/fpsyg.2010.00154>
- 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. <http://dx.doi.org/10.1037/0033-2909.125.6.777>
- Folstein, M., F. (1975). Mini Mental State: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research*, 12(3), 189-198. DOI: 10.1016/0022-3956(75)90026-6
- 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.
<https://doi.org/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 prerresponse pupillary measures. *Attention, Perception, & Psychophysics*, 77(8), 2601-2610.
<https://doi.org/10.3758/s13414-015-0971-9>
- Hedge, C., Powell, G., & Sumner, P. (2018). The mapping between transformed reaction time costs and models of processing in aging and cognition. *Psychology and Aging*, 33(7), 1093-1104. <http://dx.doi.org/10.1037/pag0000298>
- Hershman, R., & Henik, A. (2020). Pupillometric contributions to deciphering Stroop conflicts. *Memory & Cognition*, 48(2), 325-333. <https://doi.org/10.3758/s13421-019-00971-z>
- 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(3), 744-753.
<https://doi.org/10.1037/a0033159>
- Li, K. Z. H., & Bosman, E. A. (1996). Age differences in stroop-like interference as a function of semantic relatedness. *Aging, Neuropsychology, and Cognition*, 3(4), 272-284.
<https://doi.org/10.1080/13825589608256630>
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect : An integrative review. *Psychological Bulletin*, 109(2), 163-203.
- Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Éds.), *The nature of remembering : Essays in honor of Robert G. Crowder*. (p. 69-93). American Psychological Association. <https://doi.org/10.1037/10394-005>
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2 : A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, 36(3), 516-524.
<https://doi.org/10.3758/BF03195598>
- Nicosia, J., & Balota, D. (2020). The consequences of processing goal-irrelevant information during the Stroop task. *Psychology and Aging*, 35(5), 663-675.
<https://doi.org/10.1037/pag0000371>
- Parris, B. A., Hasshim, N., Wadsley, M., Augustinova, M., & Ferrand, L. (2021). The loci of Stroop effects: A critical review of methods and evidence for levels of processing contributing to color-word Stroop effects and the implications for the loci of attentional selection. *Psychological Research*. <https://doi.org/10.1007/s00426-021-01554-x>
- Rey-Mermet, A., & Gade, M. (2018) Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, 25, 1695–1716.
<https://doi.org/10.3758/s13423-017-1384-7>
- Risko, E. F., Schmidt, J. R., & Besner, D. (2006). Filling a gap in the semantic gradient : Color

- associates and response set effects in the Stroop task. *Psychonomic Bulletin & Review*, *13*(2), 310-315. <https://doi.org/10.3758/BF03193849>
- Roelofs, A. (2003). Goal-referenced selection of verbal action : Modeling attentional control in the Stroop task. *Psychological Review*, *110*(1), 88-125. <https://doi.org/10.1037/0033-295X.110.1.88>
- Salthouse, T. A. (1990). Working memory as a processing resource in cognitive aging. *Developmental Review*, *10*(1), 101–124. [https://doi.org/10.1016/0273-2297\(90\)90006-P](https://doi.org/10.1016/0273-2297(90)90006-P)
- Schmidt, J. R. (2013). The Parallel Episodic Processing (PEP) model : Dissociating contingency and conflict adaptation in the item-specific proportion congruent paradigm. *Acta Psychologica*, *142*(1), 119-126. <https://doi.org/10.1016/j.actpsy.2012.11.004>
- Schmidt, J. R., & Cheesman, J. (2005). Dissociating stimulus-stimulus and response-response effects in the Stroop task. *Canadian Journal of Experimental Psychology*, *59*(2), 132-138.
- Schneider W., Eschman A., Zuccolotto A. (2002). *E-Prime user's guide*. Pittsburgh, USA: Psychology Software Tools Inc.
- Seymour, P. H. K. (1977). Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology*, *29*(2), 245–265.
- Sheikh, J., I., & Yesavage, J., A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontology: A Guide to Assessment and Intervention* (pp. 165-173). New York, NY: The Haworth Press.
- 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*(2), 461-479. <https://doi.org/10.1037/0096-1523.22.2.461>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental*

Psychology, 18(6), 643-662. <https://doi.org/10.1037/h0054651>

Zhang, H., Zhang, J., & Kornblum, S. (1999). A Parallel Distributed Processing Model of Stimulus–Stimulus and Stimulus–Response Compatibility. *Cognitive Psychology*, 38(3), 386-432. <https://doi.org/10.1006/cogp.1998.0703>

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(1), 3-19.

Wechsler, D., Psychological Corporation., & PsychCorp (Firm). (2008). *WAIS-IV technical and interpretive manual*. San Antonio, Texas: Pearson.