Is there semantic conflict in the Stroop task?

Further evidence from a modified two-to-one Stroop paradigm combined with single-letter coloring and cueing

Mariana Burca[1], Virginie Beaucousin[1], Pierre Chausse[2], Ludovic Ferrand[2], Benjamin A. Parris[3] and Maria Augustinova[1]

[1]Université de Rouen Normandie, CRFDP, Rouen, France

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

[3]Department of Psychology, Bournemouth University, Poole, UK

Running head: Further evidence for semantic conflict in the Stroop task

Type of article: Short research article

Word count: 5623 (total including references), 4156 main text excluding references

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: + 33 473 406 438
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.

Data availability:
The data and materials are available at https://osf.io/u72p6/

Abstract

This research addressed current controversies concerning the contribution of semantic conflict to the Stroop interference effect and its reduction by single-letter coloring and cueing-procedure (SLCC). On the first issue it provides, for the first time, unambiguous evidence for a contribution of semantic conflict to the (overall) Stroop interference effect. The reported data remained inconclusive on the second issue, despite being collected in a considerable sample and analyzed with both classical (frequentist) and Bayesian inferential approaches. Given that in all past Stroop studies, *semantic* conflict was possibly 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 [$BLUE_{\text{blue}}$] are used as baseline to derive a magnitude for semantic conflict), its genuine contribution to the Stroop interference effect is the most critical result reported in the present study. Indeed, it leaves no doubt – in complete contrast to dominant single-stage response competition models (e.g., Roelofs, 2003) – that selection occurs at the semantic level in the Stroop task. The immediate implications for the composite (as opposed to unitary) nature of the Stroop interference effect and other still unresolved issues in the Stroop literature are outlined further.

Key words: Stroop interference, Single-letter coloring and cueing, Semantic conflict, Response conflict

Introduction

The typical result in the well-known Stroop task (Stroop, 1935) is that individuals' color-identification times are longer for *color-incongruent* (e.g., "BLUE" displayed in yellow; hereafter $BLUE_{\text{yellow}}$), than for *color-neutral* items (e.g., the word "DEAL" displayed in yellow, hereafter $DEAL_{\text{yellow}}$). This difference – called *the Stroop interference effect* – reveals that, despite being explicitly instructed not to, individuals still attend to the irrelevant word-dimension of compound Stroop items (i.e. process its meaning), which in turn slows their performance. Therefore, numerous studies have tried to identify experimental manipulations that improve attentional selectivity in the Stroop task.

Among such manipulations, the effects of *single-letter coloring and cueing* (SLCC; Besner et al., 1997) have perhaps received the most attention (e.g., Küper & Heil, 2012 for a review). Despite differences in the methodologies employed[1], all of these studies have reported a substantial reduction or even elimination of Stroop interference when only a single letter (as opposed to all letters) of a target display was colored in an incongruent color from the response set and spatially pre-cued (hereafter SLCC vs. ALCC [all-letters colored and cued]). Because SLCC (unlike ALCC) is thought to keep two informational sources (i.e., color and word) separate (i.e., *odd-one-out effect* of SLCC, Besner et al., 1997), a response induced by the irrelevant word-dimension of compound Stroop items (e.g., say "blue"/press blue for $BLUE_{\text{yellow}}$) interferes less with that induced by the relevant color-dimension (e.g., say "yellow"/press yellow for $BLUE_{\text{yellow}}$) in SLCC than in ALCC. Indeed, according to dominant *single-stage response competition* (e.g., Roelofs, 2003) models, a single conflict between responses (hence *response conflict*) is the driving force behind the Stroop interference effect.

[1] In some studies, the coloring manipulation was coupled with variations in spatial pre-cueing such that, for instance, (a) small arrow(s) appeared on the screen to indicate the position(s) subsequently occupied by the color-carrier(s) (i.e., target letter(s)) whose color was to be named. Some studies (e.g., Manwell et al., 2004) also added empty spaces between letters that were filled – in other studies (e.g., Labuschagne & Besner, 2015) – with non-letter characters.

In contrast to this unitary view of Stroop interference, other (i.e., *multi-stage*) models anticipate that incidental processing of the irrelevant word-dimension of Stroop items actually generates an additional (i.e., *semantic or stimulus*) conflict (e.g., Zhang et al., 1999; Zhang & Kornblum, 1998). On the basis of this idea, a handful of studies set out to investigate whether the SLCC-manipulation affects semantic and/or response conflict.

Manwell and colleagues (2004) consequently reported that SLCC reduced Stroop interference depicted above (e.g., $BLUE_{\text{yellow}} – DEAL_{\text{yellow}}$) but eliminated *semantic-associative* Stroop interference (e.g., $SKY_{\text{yellow}} – DEAL_{\text{yellow}}$, see also Labuschagne & Besner, 2015). Because these studies viewed the former (hereafter *standard* Stroop interference) as a product of both semantic and response conflicts, and viewed semantic-associative Stroop interference as being free of response conflict (e.g., Neely & Kahan, 2001[2]; but see e.g., Klein, 1964), they consequently concluded that SLCC affects the Stroop interference effect early (e.g., it curtails semantic conflict altogether, Manwell et al., 2004, Account 2; see also Besner et al., 2016 for discussion). But this conclusion contrasts with other studies reporting that SLCC reduces standard but leaves semantic-associative Stroop interference (Augustinova & Ferrand, 2007; Augustinova et al., 2010, 2018), and the associated N400-like ERP-activity[3] (Augustinova et al., 2015) unaffected. These studies subsequently claimed that SLCC affects the Stroop interference effect late (i.e., it reduces or even eliminates response conflict, see Augustinova & Ferrand, 2014 for discussions) but leaves semantic conflict unaffected. In sum, the processing stage at which SLCC-manipulation actually takes its effect (semantic vs. response) remains unclear.

However, this lack of clarity is not entirely due to the contrasting findings across the aforementioned studies. Indeed, as noted above, all of these studies have employed the

[2] Since SKY is not part of the response set, it does 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).
[3] N400-like corresponded to fronto-central negativity occurring from 380 to 480msec after the presentation of a Stroop item taken as evidence of semantic processing in the Stroop task.

semantic-associative Stroop trials (e.g., SKY_{yellow}) to investigate the extent to which SLCC affects semantic conflict. Because the irrelevant word-dimension “sky” is strongly associated with “blue”, attending to its meaning is thought to interfere with the meaning of the relevant color-dimension (yellow for SKY_{yellow}). 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” (Seymour, 1977, p. 263). Yet, this assumption is at odds with the aforementioned unitary models of Stroop interference such as that proposed by Roelofs (2003). This model predicts that both standard and semantic-associative Stroop interference are not qualitatively distinct but both result from different quantities of a single (i.e., response) conflict taking place in the language production unit. Specifically, for semantic-associative items, the incorrect response (e.g., say “blue”/press blue for SKY_{yellow}) is activated through the association of the distractor *SKY* with the response set colors (i.e., blue is a part of the response set) in such a way that it subsequently interferes with the production of the correct response (e.g., “yellow”/press yellow button for SKY_{yellow}). As a result, the influence of SLCC on semantic conflict remains an open issue because the extent to which semantic conflict is the actual driving force behind semantic-associative interference (e.g., $SKY_{\text{yellow}} – DEAL_{\text{yellow}}$) is itself unclear (see Parris et al., 2021 for discussion). To address this more fundamental issue, the present study substituted semantic-associative items with items producing semantic interference in such a way that cannot be accounted for by single-stage response competition models (i.e., cannot be the result of response conflict).

The present study

To this end, the study only used color-words as distractors (i.e., BLUE, RED, GREEN, YELLOW). Importantly, following De Houwer (2003), the same response-key (e.g., actioned with left hand) was used for making responses to blue and red (i.e., target) items, and the other response-key (e.g., actioned with right hand) was used for making responses to green and yellow items. This so-called *two-to-one* response-mapping therefore generated two kinds of color-incongruent trials. For *different-response* trials like $BLUE_{\text{yellow}}$ the correct response ("BLUE") is indicated using the pointing finger on the right hand, and the incorrect response is indicated by using the pointing finger of the left hand. There is no such (response) interference on *same-response* trials like $BLUE_{\text{red}}$, because the responses primed by both the distractor "BLUE" and the target "RED" are indicated using the pointing finger on the left hand[4]. Consequently, any significant interference caused by same-response trials cannot be attributed to response conflict. Indeed, in line with *multi-stage models* of Stroop interference (e.g., Zhang & Kornblum, 1998), it is commonly attributed to semantic conflict (De Houwer, 2003, see e.g., Hershman & Henik, 2020 for the most recent example).

However, despite this consensus, the extent to which same-response trials actually induce semantic conflict is still an open empirical issue. Indeed, with only a few exceptions (see below), all studies employing same-response trials to induce semantic conflict – including De Houwer (2003) – have used color-congruent trials as the baseline. This means that in line with what unitary models of Stroop interference would predict (Roelofs, 2003), the difference between same-response and color-congruent trials might not involve any semantic conflict but be entirely driven by facilitation on color-congruent trials. Consistent with this idea, Hasshim and Parris (2014, 2015) reported significantly longer RTs for same-response trials than for color-congruent trials, but no difference between same-response and color-neutral word trials that are free from facilitation (Brown, 2011; MacLeod, 1991 for discussion)[5]. Somewhat in

[4] Note that these items are likely to involve at least some response facilitation (Hasshim & Parris, 2014; 2015 for a discussion).

[5] 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).

contrast, Hershman and Henik (2020) reported evidence for a difference between same-response and *non-word* color-neutral trials but in pupillometric measures only[6], whereas Burca found preliminary – as yet unpublished – evidence suggesting the presence of semantic conflict in RTs even when color-neutral *words* are used as baseline. In sum, because of the theoretical implications for composite (e.g., Zhang & Kornblum, 1998) as opposed to unitary (e.g., Roelofs, 2003) Stroop interference, the unambiguous demonstration of same-response trials inducing (or not) *semantic conflict* still remains an empirical void to fill.

Therefore, the present study used color-neutral words to measure the magnitude of interference induced respectively by *same*- (e.g., $BLUE_{\text{red}}$) and *different-response* ($BLUE_{\text{yellow}}$) trials. Specifically, color-neutral trials were matched with color-incongruent ones in such a way that they appeared in the same colors an equal number of times (e.g., *BLUE* only appeared in red, green, yellow, and so did its counterpart *DEAL*). This allowed for comparisons of *same*- (e.g., $BLUE_{\text{red}}$) and *different-response* ($BLUE_{\text{yellow}}$) trials with their corresponding (in terms of response contingency) color-neutral baselines (e.g., $DEAL_{\text{red}}$, and $DEAL_{\text{yellow}}$ respectively). In sum, as in past studies of SLCC-effect outlined above, the magnitudes of both *semantic* (e.g., $BLUE_{\text{red}} – DEAL_{\text{red}}$) and *standard Stroop interference* ($BLUE_{\text{yellow}} – DEAL_{\text{yellow}}$) were derived. However, this was done without using the same color-neutral baseline twice (see Shichel & Tzelgov, 2018 for discussion).

If, in agreement with single-stage response competition models (e.g., Roelofs, 2003), there is no contribution of semantic conflict to the overall Stroop interference effect, same-response trials will not induce any Stroop interference (i.e., their RTs will be comparable to those observed for color-neutral trials – as previously reported by Hasshim and Parris (2014, 2015)). Different-response trials will be responded to slower than same-response trials (De

[6] In the case of RTs, this study perfectly replicated Hasshim and Parris (2014, 2015) including conclusions based on Bayes factors. Based on this dissociation, these authors concluded that RTs are not the best-suited measures to addressed a more fine-grained components of the Stroop interference effect (see also Hershman & Henik, 2019).

Houwer, 2003), and this in the absence of a significant difference between the same-response and their neutral counterparts. In this case – confirming that the response conflict is the unique driving force of the Stroop interference effect –, only the magnitude of standard Stroop interference (the significant difference between different-response and color-neutral trials) will be reduced in SLCC compared to ALCC.

In contrast, if the multi-stage models of Stroop interference (e.g., Zhang & Kornblum, 1998) are correct and semantic conflict genuinely contributes to overall Stroop interference, then same-response trials will produce significant semantic Stroop interference such that the difference between same-response and color-neutral trials will be significant (Burca, unpublished; see also Hershman and Henik's pupillometry data). The semantic conflict could then be affected by SLCC in several ways. In line with past studies reviewed above, SLCC might *eliminate* semantic interference altogether (Manwell et al.'s Account 2; Labuschagne & Besner, 2015), or leave its magnitude unaffected (Augustinova et al., 2010, 2015, 2018). Finally, SLCC might simultaneously *reduce* both semantic and standard interference (Manwell et al.'s Account 1)[7] – a finding that would suggest that SLCC affects both semantic and response conflicts. The aim of the experiment reported below was to directly test these different hypotheses regqrding the contribution of semantic conflict to Stroop interference and to the effect of SLCC on semantic conflict.

Method

Participants

Ninety-two native French-speakers (64 females and 28 males; M_{age}=20.95; SD=1.612) with normal or corrected-to-normal color-vision volunteered to take part in this experiment, which was approved by the local ethics committee (2018-02-A).

[7] This causes the elimination of semantic-associative Stroop interference (SKY_{yellow} – $DEAL_{yellow}$) as it is usually proportionally smaller than its standard counterpart ($BLUE_{yellow}$ – $DEAL_{yellow}$).

Design

The study used a 4 (Stimulus-Type: different response vs. same response vs. neutral different response vs. neutral same response) × 2 (Coloring: ALCC vs. SLCC) within-participant design for data collection.

Stimuli

The (French) stimuli were presented in lowercase Courier font, size 18, on a black background and subtended an average visual angle of 0.9° high × 3.0° wide. They consisted of four color-words: *rouge* [red], *jaune* [yellow], *bleu* [blue], and *vert* [green]; and four non-color counterparts: *plomb* [lead], *liste* [list], *page* [page], *cave* [basement], that were paired in length and frequency via *Lexique 3.38* (New et al., 2004).

Apparatus and Procedure

The participants were seated 70cm in front of a 17" computer screen. Eprime 2.0 software was used for data presentation and recording. Because the experimental trials only used words as distractors, the participants were instructed to identify the color of the letter indicated by the arrow as quickly and accurately as possible by pushing the appropriate color-button and to ignore everything else in the display. To this end, and similarly to Augustinova et al. (2010, Experiment 1), the participants were instructed to concentrate on the fixation cross ("+") that appeared for 2000msec in the center of the screen at the beginning of each trial. A small white arrow (height of 1.2° of visual angle, displayed 0.6° below) then appeared for 150msec. As it served as a spatial pre-cue, it was located at the position that was subsequently occupied by the target letter. This position varied randomly from trial-to-trial, being located at the initial, optimal viewing position (OVP), middle, or final letter of the distractor-word (Parris et al., 2007). In order to avoid an additional color-color interference in the SLCC-condition[8],

[8] An additional color-color interference occurs when remaining letters are colored in other (incongruent) colors from the response set. Since this is known to increase RTs on color-neutral trials in the SLCC-condition (e.g. Küper & Heil, 2012; Monahan, 2001) it might inflate the SLCC-induced reduction of both standard and semantic Stroop interference (e.g. Manwell et al., 2004).

the spatially pre-cued letter was the only one that appeared in an incongruent color from the response set and the rest of the letters appeared in white (i.e., a color that was not part of the response set; see e.g., Augustinova & Ferrand, 2007; Besner et al., 1997; Brown et al., 2002 for the use of gray or black). In the ALCC-condition, all letters (including the one that was spatially pre-cued) appeared in incongruent colors from the response set. The entire display remained on screen until the participant responded or until 3500msec had elapsed.

The participants responded manually using a modified SRBox® consisting of two handles, each of which had a single response button at the top (placement of the handles in the right or left hand, respectively, was counterbalanced across participants). The participants pushed these response buttons with their thumbs. This allowed them to hold each handle comfortably in their palm with the remaining four fingers (see Supplementary Materials, left panel) while resting their arms on an armrest.

One response button was flanked by blue and red color-stickers and the other by yellow and green color-stickers (see Supplementary Materials, middle and right panels). Because of this color assignment, together with the fact that color-stimuli only appeared in incongruent colors, a single presentation of the full set of color-stimuli in all possible colors resulted in eight same-response items (i.e., 4 in ALCC and 4 in SLCC; e.g. “blue” presented entirely in red, “red” presented entirely in blue, “green” presented entirely in yellow and “yellow” presented entirely in green in ALCC) and sixteen different-response items. To control for contingency, the same presentation was used for the items' color-neutral counterparts (e.g., *PAGE* only appeared in red, yellow and green, exactly like its color-incongruent counterpart *BLUE*). All stimulus types were therefore seen an equal number of times, meaning that 5 repetitions (making it possible to fully control for the letter-position effects, see above) of the full set of 48 different stimuli resulted in a total of 240 experimental stimuli. In each coloring-condition, we therefore collected RTs on 20 same-response (SR) incongruent trials, 20 color-neutral trials paired with

SR-incongruent trials, 40 different-response (DR) incongruent trials, and 40 color-neutral trials paired with DR-incongruent trials all included in a single block. Before completing this experimental block, the participants first completed 128 practice trials (MacLeod, 2005) consisting of asterisks in order to learn the color-button correspondence. They proceeded to the experimental block once their accuracy rate was above 95% (2 participants had to redo the training, one of them was later excluded from further analyses).

Results and Discussion

Four participants were excluded from the analysis due to faulty data recordings. The data of the remaining eighty-eight participants were first analyzed in an omnibus 4 (Stimulus-Type: DR-incongruent trials vs. SR-incongruent trials vs. DR-neutral trials vs. SR-neutral trials) × 2 (Coloring: ALCC vs. SLCC) ANOVA. This analysis on RTs was conducted on mean correct latencies using both 2.5 (e.g., Labuschagne & Besner, 2015) and 3*SD* cut-offs (Augustinova and colleagues' studies). Given their convergence and to permit comparisons with our past studies, only analyses using the 3*SD* cut-off (leading to the exclusion of 1.12% of the total data) are reported below. In omnibus ANOVA, this latter analysis revealed main effects of Stimulus-Type [$F(3,261)=38.09$; $p<.001$, $\eta_p^2=0.304$; $BF_{10}=4.160e+18$[9], indicating extreme evidence of an effect of Stimulus-Type] and of Coloring [$F(1,87)=11.57$; $p=.001$, $\eta_p^2=0.117$; $BF_{10}=29.223$, indicating strong evidence of an effect of Coloring]. It also revealed a significant Stimulus-Type × Coloring interaction, [$F(3,261)=3.236$; $p=.023$, $\eta_p^2=0.036$; $BF_{10}=2.229e+20$]. This extreme Bayesian evidence, is in favor the model where the both main effects and their interaction are significant. Nevertheless, an effect of the interaction alone was not supported by Bayesian evidence ($BF_{10}=0.606/BF_{01}=1.648$[10], see Table S1 in Supplementary Materials for a

[9]BF_{10} corresponds to the Bayesian probability of the occurrence of a hypothesis (H1) and the likelihood of another null hypothesis (H0). It was calculated with JASP (JASP Team, 2017) and interpreted according to Lee and Wagenmakers (2013 adjusted from Jeffreys, 1961). All priors were equal.

[10]BF value reported for interaction only is the one associated with the model including the interaction and the two

comparison of models in a Bayesian repeated-measures ANOVA].

These same analyses of error percentages (see Table 1) revealed only significant main effects of Stimulus-Type [$F(3, 261)=6.91$; $p< .001$, $\eta_p^2=0.74$; $BF_{10}=182.463$, indicating extreme evidence of an effect of Stimulus-Type]. Both the main effect of Coloring [$F(1,87)=2.83$; $p=.096$, $\eta_p^2= 0.32$; $BF_{10}=0.210/BF_{01}=4.810$], and Stimulus-Type × Coloring interaction [$F(3, 261)=.12$; $p=.950$, $\eta_p^2= 0.01$] remained non-significant. Indeed, BFs provided very strong evidence against the effect of interaction alone ($BF_{10}=0.013/BF_{01}=68.52$, see Table S2 in the Supplementary Materials for a comparison of the models).

Is there semantic conflict in the Stroop task?

To answer this key question, we first decomposed the aforementioned main effect of the Stimulus-Type on mean correct latencies. It revealed the presence of both semantic and standard Stroop interference. The SR-incongruent trials were indeed responded to slower than their SR-neutral counterparts ($p<.001$; $BF_{10}=44197.61$) and so were the DR-incongruent trials compared to their DR- neutral counterparts ($p<.001$; $BF_{10}=1.342e+12$)[11]. To explore this issue of semantic conflict further, we then decomposed the aforementioned Stimulus-Type × Coloring interaction (i.e., interaction supported at least by frequentist statistics) by testing the simple main-effect of Stimulus-Type at each level of Coloring. This simple main-effect was significant in both ALCC [$F(3,85)=22.762$; $p<.001$, $\eta_p^2=0.445$; $BF_{10}=6.986e+11$] and SLCC [$F=(3,85)=12.898$; $p<.001$, $\eta_p^2=0.313$; $BF_{10}=368930.025$[12]]. In addition to significant standard Stroop interference (DR-incongruent – DR-neutral trials) in both ALCC- and SLCC-condition

main effects divided by BF value observed for the model including the two main effects only.

[11] The difference between the RT's of DR and SR incongruent trials (10msec), which is representative of the response conflict, did not reach significance ($p=.135$). However, $BF_{10}=2.962/BF_{01}=0.338$ values of paired samples T-Tests provided at least anecdotal evidence in favor of such an effect.

[12] A Bayesian repeated measures ANOVA compares a series of different models against a null model. It therefore cannot be decomposed by testing simple main effects per se. Therefore, BFs provided test main effect of Stimulus-Type in ALCC and SLCC respectively – which explains why a Bayesian approach that is generally more conservative, actually yields larger effects.

(all p_s<.01, see Table 1 for descriptive statistics), further pairwise comparisons, again unequivocally revealed the presence of significant semantic Stroop interference (SR-incongruent – SR-neutral trials) in both ALCC- and SLCC-conditions (see Table 1 for magnitudes). This latter conclusion is reinforced by BF_{10}=130.036, providing extreme evidence of semantic interference in the ALCC-condition and BF_{10}=19.354, providing strong evidence of semantic interference in the SLCC-condition[13]. The presence of significant semantic interference overall, and in both coloring conditions, is thus consistent with the idea that a semantic conflict indeed genuinely contributes to the overall Stroop interference effect even when color-neutral *words* are used as baseline (see also Hershman & Henik, 2020' pupillometric measures, but see Hashim & Parris, 2014, 2015; Hershman & Henik, 2020' RTs).

<Insert Table 1 about here>

How does SLCC affect semantic conflict?

The aforementioned significant semantic interference in SLCC-conditions runs counter to Manwell et al.'s Account 2 (see also Labuschagne & Besner, 2015) suggesting that SLCC *eliminates* semantic interference altogether. However, in line with Manwell and colleagues' Account 1 (and counter to past results observed by Augustinova, Ferrand and colleagues), SLCC (as opposed to ALCC) shortened RTs in both DR-incongruent trials [$F(1,87)=13.396$;

[13] In line with a specific contribution of response conflict to standard Stroop interference – latencies for DR-incongruent trials tended to be longer than those for SR-incongruent trials (see Table 1). But although the observed 12msec difference was indeed marginally significant (p=.072) in ALCC, BF10=.569/BF_{01}=1.758 tend to suggest otherwise. Despite this, response conflict in ALCC is still evident in percentages of errors [$t(87)=3.80$, $p<.001$; BF_{10}=77.45]. In SLCC, the 7msec difference between mean response latencies for DR-incongruent and SR-incongruent trials was not significant (p=.173, BF10=.291/BF_{01}=3.436), but response conflict was however present in percentages of errors [$t(87)=2.79$, $p<.05$; BF_{10}=4.34]. The additional sequential analyses of response conflict in ALCC vs. SLCC condition suggested that the probability of finding this type conflict if more observations were added, is unlikely (see Figure 2 in Supplementary Materials). Therefore, for each level of Stimulus and of Coloring, we binned slow vs. fast trials. The additional analyses (see Supplementary Materials) suggest that the lack of response conflict reported above is driven by the lack of response conflict in the fast trials (see Table 3S in Supplementary Materials). Again, the additional sequential analyses of response conflict in fast vs. slow trials suggested that the probability of finding this type conflict if more observations were added, is unlikely (see Figure 3 in Supplementary Materials).

p<.001, η_p^2=0.133; BF_{10}=50.129, which indicates very strong evidence of Coloring] and SR-incongruent trials [F(1,87)=5.912; p =.017, η_p^2=0.064; BF_{10}=1.981/BF_{01}=0.529]. It should be noted however, that the latter Bayesian evidence for SLCC shortening SR-incongruent trials was anecdotal. Finally, this simple main effect of Coloring on both DR-neutral [F(1,87) = 0.117; p =.733, η_p^2=0.001; BF_{10} =0.125/BF_{01}=8.025] and SR-neutral trials [F(1,87) = 1.519; p=.221, η_p^2=0.017; BF_{10} = 0.244/BF_{01}=4.091] revealed that color-neutral trials remained unaffected by Coloring manipulation.

To further examine the SLCC-effect in the two-to-one Stroop paradigm, we then conducted a 2 (Interference-Type: Standard vs. Semantic) × 2 (Coloring: ALCC vs. SLCC) ANOVA to analyze the extent to which magnitudes of both standard and semantic Stroop interference were affected by SLCC (see Table 1). This analysis revealed main effects of Interference-Type [F(1,87)=7.285; p=.008, η_p^2=0.077; BF_{10}=2.207], and Coloring [F(1,87)=6.066; p=.016, η_p^2=0.065; BF_{10} = 5.243] but only a non-significant Interference-Type × Coloring interaction, [F(1,87)=1.925; p=.169, η_p^2=0.0322]. Indeed, in line with the aforementioned Stimulus × Coloring interaction, the specific BF value for the interaction (see Table S4 in Supplementary Materials for a comparison of the models) was only BF_{10}=0.401 and again BF_{01}=2.491 actually provided anecdotal evidence for a null effect of the Interference-Type × Coloring interaction. These results are therefore clearly inconsistent with Augustinova and colleagues' past results showing that SLCC leaves the magnitude of semantic Stroop interference unaffected (Augustinova et al., 2010, 2015, 2018). Indeed, the present results suggest that SLCC reduces both semantic and standard Stroop interference in tandem.

General Discussion and Conclusions

With regards to the first aim of addressing controversies concerning the role of semantic conflict in contributing to Stroop interference, the results reported above provided extreme

(ALCC) and strong (SLCC) Bayesian evidence for differences between same-response trials and color-neutral word baselines (e.g., $DEAL_{red}$). This means that the present study reports – for the first time – a genuine contribution of semantic conflict to overall Stroop interference. Indeed, in all past Stroop studies this contribution was possibly confounded either with response conflict (e.g., when semantic-associative items [SKY_{blue}] were used to induce semantic conflict) or facilitation (when color-congruent items [$BLUE_{blue}$] were used as the baseline from which to derive the magnitude of semantic conflict generated by same-response items). Also, and importantly, magnitudes of semantic interference reported in the present paper are not inflated either by the use of color-neutral *non-word* baselines (see Brown, 2011 for discussion of this latter issue) or by the inclusion of color-congruent trials that are also known to amplify interference (e.g., Roelofs, 2014). This unambiguous contribution of a genuine semantic conflict to the (overall) Stroop interference effect – at least with manual responses that the two-to-one paradigm (De Houwer, 2003) necessarily involves – constitutes the most significant result reported in the present study.

Whilst it is unclear why it was observed in the present study and not in other studies comparing the interfering effect of same-response trials against a color-neutral word base-line (Hasshim & Parris, 2014, 2015), the unambiguous presence of a genuine semantic conflict has at least two immediate implications for several still unresolved issues in the Stroop literature. First, it provides clear evidence that so-called *informational conflict* (e.g., MacLeod & MacDonald, 2000) in the Stroop task includes semantic conflict. Indeed, because of the aforementioned confounds, the evidence in favor of this contribution from semantic conflict available was still inconclusive prior to the present paper (see Parris et al., 2021, for a thorough discussion of this issue). Second, the unambiguous presence of genuine semantic conflict clearly implies that selection occurs at the level of semantics – a finding that runs counter to the dominant single-stage response competition models of Stroop interference. Therefore, it is likely that these latter models need to be modified to make room for this type of conflict.

Despite the data being collected in a considerable sample (Brysbaert & Stevens, 2018) and analyzed with both classical (frequentist) and Bayesian inferential approaches, the results reported above are clearly less conclusive with regards to the second aim which was to address controversies concerning the reduction of semantic conflict by SLCC-manipulation. Indeed, the lack of Interference-Type × Coloring interaction (along with a Stimulus-Type × Coloring only being present in standard frequentist ANOVA) seems most consistent with Manwell et al.'s Account 1 suggesting that SLCC reduces both semantic and standard interference in tandem[14]. In line with Kinoshita et al. (2018; see also e.g. Besner et al., 2016), this means that SLCC reduces semantic conflict in the Stroop task. However, the results with regards to the SLCC-effect on SR-incongruent trials come with caveats. Indeed, the Bayesian analyses reported above provided anecdotal evidence for the simple main-effect of Coloring on these trials whereas this effect was strong for DR-incongruent trials. This points to the fact the while the interpretation in terms of reduced semantic conflict (i.e., the common denominator of both semantic and standard Stroop interference) is possible, it is likely to be incomplete. In line with past studies (e.g., Augustinova et al., 2010, 2015, 2018), it therefore remains possible that SLCC also reduces response conflict as the aforementioned strong simple main-effect of Coloring on DR-incongruent trials suggest. Although the response conflict was not significant overall[15], this possibility cannot be ruled out entirely, as it was significant on slow trials (see

[14] This conclusion is reinforced by the exploratory 2 (Interference-Types: Stroop interference vs. Semantic interference) x 2 (Coloring: ALCC vs. SLCC) ANOVA conducted on slow trials only (i.e., trials revealing response conflict). Indeed, it yielded a marginally significant main effect of Interference-Type [$F(1,87)=3.50$; $p < .065$, $\eta_p^2= 0.039$], a significant main effect of Coloring [$F(1,87) =5.39$; $p=.023$, $\eta_p^2= 0.058$], but a non-significant Interference-Type × Coloring interaction [$F(1,87)=.294$; $p< .589$, $\eta_p^2=0.003$] suggesting again that Coloring manipulation affects both types of interference in tandem.

[15] This unexpected absence of significant response conflict (that our group has actually replicated since) is at odds with past studies using the two-to-one Stroop paradigm – including De Houwer's initial study (2003). This could be explained by differences in methodology applied in the present study (e.g., absence of color-congruent items that are known to amplify the interfering effects of color-incongruent items (e.g., Roelofs, 2014), the equal percentage of color-incongruent vs color-neutral trials (MacLeod, 2005) with no contingency issues involved (Hasshim & Parris, 2014; Schmidt, 2013). More studies are needed to assess these different possibilities directly.

Supplementary Materials). This reduction is again probably too small to fully explain the present pattern of results. The remaining possibility is that SLCC actually reduces 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; Hershman & Henik, 2019). The lack of SLCC-effects on color-neutral stimuli in the present study is seemingly at odds with this latter idea (but see e.g., Manwell et al., 2004). Indeed, any reduction of task conflict should in principle facilitate processing of all compound stimuli (i.e., including the color-neutral items) and even more so in the present study given that only one color from the response set was used to implement the SLCC-manipulation. It is, however, noteworthy that two colors are still present in our SLCC-condition (a color from the response set and white) – making it plausible that an additional color-color interference still occurs in SLCC as compared to ALCC. Its known lengthening effect on color-neutral items (e.g. Küper & Heil, 2012; Monahan, 2001) might be cancelled in the RTs by the concomitant shift in attentional focus toward the relevant color-dimension in SLCC, which in turn reduces the task conflict. In sum, future studies – which should include a direct measure of task conflict – need to address these possibilities directly, perhaps with a more fine-grained measures than RTs (e.g., Hershman & Henik, 2019). The decrease in the magnitude of overall Stroop interference that is specifically due to the reduction of task conflict implies – contrary to the reduction of semantic conflict – a simultaneous increase in Stroop facilitation (Parris, 2014), which was also not measured in the present study. It is of course possible that SLCC actually affects all the components of the overall Stroop interference (response, semantic and task conflict).

In conclusion, whilst the influence of SLCC on semantic conflict (but also on other types of conflicts) still remains an open issue, the presence of semantic conflict in the Stroop task – at least when administered in the form of the two-to-one Stroop paradigm –, is no longer one. Therefore, the present study provides impetus for future empirical work on SLCC (see above). Furthermore, the present work also strongly encourages the development of new integrative models of the Stroop inference effect, as only one existing model effectively accounts for either semantic (Zhang & Kornblum, 1998; Zhang et al., 1999) or task conflict (Kalanthroff et al., 2018), respectively, but none currently accounts for the probable coexistence of task, semantic and response conflicts (see Parris et al., 2021 for further discussion).

References

Augustinova, M., & Ferrand, L. (2007). Influence de la présentation bicolore des mots sur l'effet Stroop. *L'Année Psychologique*, *107*(02), 163-179. https://doi.org/10.4074/S0003503307002011

Augustinova, M., & Ferrand, L. (2014). 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*(6), 827-833. https://doi.org/10.3758/PBR.17.6.827

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., 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*(2), 543-549. https://doi.org/10.3758/s13423-014-0697-z

Besner, D., Risko, E. F., Stolz, J. A., White, D., Reynolds, M., O'Malley, S., & Robidoux, S. (2016). Varieties of Attention: Their Roles in Visual Word Identification. *Current Directions in Psychological Science*, *25*(3), 162-168. https://doi.org/10.1177/0963721416639351

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, *4*(2), 221-225. https://doi.org/10.3758/BF03209396

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.

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. *The American Journal of Psychology*, *115*(4), 515. https://doi.org/10.2307/1423526

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.10

Burca, M. (unpublished). *Is there any Semantic Conflict in the two-to-one Stroop paradigm? (working title).* Doctoral dissertation in progress. University of Rouen.

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

Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1170-1176. https://doi.org/10.1037/0096-1523.33.5.1170

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 preresponse pupillary measures. *Attention, Perception, & Psychophysics*, *77*(8), 2601-2610. https://doi.org/10.3758/s13414-015-0971-9

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

Hershman, R., & Henik, A. (2020). Pupillometric contributions to deciphering Stroop conflicts. *Memory & Cognition*, *48*(2), 325-333.

Kalanthroff, E., Davelaar, E., Henik, A., Goldfarb, L., & Usher, M. (2018). Task conflict and proactive control: A computational theory of the Stroop task. *Psychological Review, 125*(1), 59-82.

Kinoshita, S., Mills, L., & Norris, D. (2018). The semantic Stroop effect is controlled by endogenous attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44(11)*, 1730-1742.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. T*he American journal of psychology*, *77*(4), 576-588.

Küper, K., & Heil, M. (2012). Attentional Focus Manipulations Affect Naming Latencies of Neutral But Not of Incongruent Stroop Trials. *Swiss Journal of Psychology*, *71*(2), 93-100. https://doi.org/10.1024/1421-0185/a000075

Labuschagne, E. M., & Besner, D. (2015). Automaticity revisited: When print doesn’t activate semantics. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00117

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin, 109(2),* 163.

MacLeod, C. M. (2005). The Stroop Task in Cognitive Research. In A. Wenzel & D. C. Rubin (Eds.), *Cognitive methods and their application to clinical research. (pp. 17-40)*. American Psychological Association. https://doi.org/10.1037/10870-002

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*(10), 383-391.

Manwell, L. A., Roberts, M. A., & Besner, D. (2004). Single letter coloring and spatial cuing eliminates a semantic contribution to the Stroop effect. *Psychonomic Bulletin & Review*, *11*(3), 458-462. https://doi.org/10.3758/BF03196595

Monahan, J. S. (2001). Coloring Single Stroop Elements: Reducing Automaticity or Slowing Color Processing? *The Journal of General Psychology*, *128*(1), 98-112. https://doi.org/10.1080/00221300109598901

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 (Eds.), *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

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.*

Parris, B. A., Sharma, D., & Weekes, B. (2007). An Optimal Viewing Position Effect in the Stroop Task When Only One Letter Is the Color Carrier. *Experimental Psychology, 54*(4), 273-280. https://doi.org/10.1027/1618-3169.54.4.273

Parris, B. A., Augustinova, M., & Ferrand, L. (2020). *The Locus of the Stroop Effect*. Frontiers Media SA, Lausanne. https://doi.org/10.3389/978-2-88963-445-3

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 colour-word Stroop effects and the implications for the loci of attentional selection. *Psychological Research.* https://doi.org/10.1007/s00426-021-01554-x

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

Roelofs, A. (2014). Tracking eye movements to localize Stroop interference in naming: Word planning versus articulatory buffering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1332-1347.

Schmidt, J. R. (2013). Questioning conflict adaptation: Proportion congruent and Gratton effects reconsidered. *Psychonomic Bulletin & Review*, *20(4*), 615-630.

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.

Shichel, I., & Tzelgov, J. (2018). Modulation of conflicts in the Stroop effect. *Acta Psychologica*, *189*, 93-102. https://doi.org/10.1016/j.actpsy.2017.10.007

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., & 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.

Zhang, H. H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of stimulus–stimulus and stimulus–response compatibility. *Cognitive Psychology, 38*(3), 386- 432.