Parris, Benjamin A., Sharma, Dinkar, Weekes, Brendan S. Hackett, Momenian, Mohammad, Augustinova, Maria and Ferrand, Ludovic (2019) *Response Modality and the Stroop Task: Are there phonological Stroop effects with manual responses?* Experimental Psychology, 66 (5). pp. 361-367. ISSN 1618-3169.

Response modality and the Stroop task: Are there phonological Stroop effects with manual responses?

Benjamin A. Parris[1*], Dinkar Sharma [2], Brendan Stuart Hackett Weekes[3-7], Mohammad Momenian[3], Maria Augustinova[8], and Ludovic Ferrand[9]

[1] *Department of Psychology, Bournemouth University*

[2] *Department of Psychology, University of Kent*

[3] *Laboratory for Communication Science, University of Hong Kong, Hong Kong*

[4] *Melbourne Graduate School of Education, University of Melbourne*

[5] *School of Psychological Sciences, University of Melbourne*

[6] *State Key Laboratory for Brain and Cognitive Sciences, University of Hong Kong*

[7] *School of Foreign Studies, University of Science and Technology, Beijing (tbc)*

[8] *Normandie Université, UNIROUEN, CRFDP, 76000 Rouen, France*

[9] *Université Clermont Auvergne, CNRS, LAPSCO, 63000 Clermont-Ferrand, France*

*Corresponding author:

Dr. Benjamin A. Parris

Department of Psychology

Faculty of Science and Technology

Bournemouth University

Talbot Campus

Fern Barrow

BH12 5BB

Email: bparris@bournemouth.ac.uk

Word count: 3986

Abstract

A long-standing debate in the Stroop literature concerns whether the way we respond to the colour dimension determines how we process the irrelevant dimension or whether word processing is purely stimulus-driven. Models and findings in the Stroop literature differ in their predictions about how response modes (e.g., responding manually versus vocally) affect how the irrelevant word is processed (i.e. phonologically, semantically) and the interference and facilitation that results, with some predicting qualitatively different Stroop effects. Here we investigated whether response mode modifies phonological facilitation produced by the irrelevant word. In a fully within-subjects design, we sought evidence for the use of a serial print-to-speech prelexical phonological processing route when using manual and vocal responses by testing for facilitating effects of phonological overlap between the irrelevant word and the colour name at the initial and final phoneme positions. The results showed phoneme overlap leads to facilitation with both response modes, a result that is inconsistent with qualitative differences between the two response modes.

Keywords: Stroop; selective attention; response mode; phonological; facilitation

The Stroop task (Stroop, 1935) is a test of the mechanisms of selective attention. It is also used to test predictions about the relative automaticity of written word processing (see Augustinova & Ferrand, 2014; MacLeod, 1991, for reviews). The task yields a Stroop effect which refers to slower naming of the print color of the word when the word refers to another colour (e.g., the word *red* is printed in blue; called an incongruent trial) compared to when the word is the same as the print colour name (e.g., the word *red* is printed in red; called a congruent trial). Furthermore, when trials are compared to neutral non-colour related trials (e.g., *top* printed in blue; called a neutral trial), incongruent trials produce interference whilst congruent trials produce facilitation.

One finding to emerge from Stroop experiments is that the mode of response used to classify the colour of the printed word (e.g., responding manually via a keypress vs. responding vocally by saying the colour name aloud) determines the magnitude and type of facilitation and interference that results (Fennell & Ratcliff, 2019; Glaser & Glaser, 1989; Kinoshita, De Wit & Norris, 2017; McClain, 1983; Redding & Gerjets, 1977; Sharma & McKenna, 1998; Turken & Swick, 1999; Zahedi, Abdel Rahman, Stürmer & Sommer, 2019; see also MacLeod, 1991; Parris, Hasshim, Wadsley, Augustinova, & Ferrand, submitted). For example, Sharma and McKenna (1998) reported evidence they argued showed no semantic Stroop effects with the manual response Stroop task (although see Brown & Besner, 2001). Accounts of differences between these two response modes propose differential access to the systems (i.e., phonological, lexico-semantic, or response level processing) that are assumed to produce interference and facilitation (Glaser & Glaser, 1989; Kinoshita, De Wit & Norris, 2017; Sharma & McKenna, 1998; Sugg & McDonald, 1994; Turken and Swick, 1999; Virzi & Egeth, 1985; see also Fennell & Ratcliff, 2019; Zahedi et al., 2019). Consistent with this, it has recently been argued that vocal and manual responding involve different tasks (naming vs classification, respectively) and as such the type of evidence that is accumulated during

Stroop task performance is different, leading to qualitatively different Stroop effects (Kinoshita et al., 2017). This account predicts that reading and the semantic access that this entails is not an automatic invariant process, as argued by others (Augustinova & Ferrand, 2014; Neely & Kahan, 2001).

In contrast to the differential access position, Roelofs (2003) argued for *similar* effects with manual and vocal response modes based on his WEAVER++ model of Stroop task performance positing mandatory verbal mediation of manual responses; under this account participants subvocally produce the colour name when responding manually and thus any effect found with vocal responses should be observed with manual responses. In the present study we investigated whether phonological processing of the irrelevant word differs between manual and vocal responses.

Previous investigations have provided evidence for phonological processing with a vocal response. To seek evidence for a serial grapheme-to-phoneme processing route in word reading, Coltheart, Woollams, Kinoshita and Perry (1999) developed a novel Stroop paradigm by creating stimuli that either shared the initial phoneme or end phoneme of a colour name. For example, if the colour to be named was red, the to-be-ignored word would be *rat* (sharing initial phoneme) or *pod* (sharing the end phoneme) or a word that shares no phoneme at all e.g. *fit.* Words that share a phoneme with a colour name have been shown to produce a naming latency advantage when naming the colour not the printed word (Regan, 1978). Coltheart and colleagues reasoned that if, as assumed by the Dual Route Cascaded (DRC) model (see Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), there is a serial order component in processing of print-to-speech, there will be an advantage for colour names that share an *initial* phoneme with the to-be-ignored written word compared to items that share an *end* phoneme because the computation of phonology from print is left to. Consistent with their predictions, Coltheart and colleagues reported that phonemic overlap was significant at both positions. However, facilitation was greater for items with overlap in the initial phoneme which they argued is incompatible with models that do not assume a serial grapheme-to-phoneme processing route in reading (see also Marmurek et al., 2006, for a partial replication of this effect).

Evidence for automatic phonological processing in the Stroop task was also sought by Besner and Stolz (1998) who employed a manual response. Besner and Stolz (1998) employed nonwords that sound like colour words (e.g. *bloo*), known as pseudohomophones, as the irrelevant stimuli, and found substantial Stroop effects when compared to a neutral baseline (see also Tzelgov, Henik, Sneg, & Baruch, 1996). They argued that phonological computation cannot be controlled, even with a manual response. However, a limitation of their design given present purposes is the fact that whilst phonological processing must be activated during responding to produce a pseudohomophone effect, there are alternative sources of interference and facilitation (orthographic/semantic) that cannot be ruled out, which could have led to any subsequent Stroop effects. As such, on the basis of this study alone it is not possible to determine whether phonological processing of the irrelevant word contributed to the Stroop effects with the manual response.

Recent work has presented evidence suggesting that the manual response Stroop task does not produce phonologically-based Stroop effects (Kinoshita et al., 2017). Kinoshita, De Wit and Norris (2017) asked what aspects of the reading process is triggered by the irrelevant word stimulus to produce interference in the colour-word Stroop task. They compared performance on five types of colour-neutral letter strings to incongruent words. They included real words (e.g., HAT), pronounceable pseudowords (e.g., HIX), consonant strings (e.g., HDK), nonalphabetic symbol strings (e.g., &@£), and a row of Xs. They reported that there was a wordlikeness or pronounceability gradient where real words and pseudowords showed an equal amount of interference (with interference increasing with string length) and more than that produced by the consonant strings which in turn produced more interference than the symbol strings and the row of Xs (which did not differ from each other). Under their account, the phonological encoding (the segment-to-frame association processes in articulation planning) of the colour name was slowed by the computation of phonology that occurs independent of lexical status and as long as the word is pronounceable. Importantly for present purposes, the authors reported this effect when participants responded vocally, but did not observe it when participants responded manually. Kinoshita et al. (2017) suggested that it is the process of generating a speech code from the irrelevant letter string that produced the pronounceability in the vocal task. They argued that because the task in the manual Stroop task is to categorize and not name the relevant colour, generation of the speech code of the irrelevant word does not cause interference. This task-dependent effect is consistent with the Bayesian reader model (Norris, 2006; Norris & Kinoshita, 2008) that stipulates that it is the task goal and not some invariant automatic reading processing that determines the outcome of processing. And the task goal in vocal and manual responding is different (naming vs. classification, respectively). Such a finding is inconsistent with Roelofs' (2003) lexical mediation account of manual responses.

The present study was designed to test the prediction that manual and vocal Stroop tasks lead to qualitatively different phonological Stroop effects. We employed Coltheart et al.'s (1999) paradigm to investigate whether phonological facilitation arises with a manual response when the initial or final phoneme is shared by the irrelevant word and colour name. In a fully within-subjects design, participants were asked to complete the Coltheart et al. paradigm using both manual and vocal responses (in counterbalanced order). A finding of phonological facilitation with a vocal response but not with a manual response would be supportive of a qualitative difference between response modes perhaps resulting from the different task goals and resulting differential access associated with responding vocally (naming) and manually (classification). In contrast, a finding of phonological facilitation with manual and vocal responses would be consistent with similar mechanisms underpinning both response modes (Roelofs, 2003).

Method

Participants. 44 participants (30 females) were recruited from Universities of Bournemouth and Sussex (UK). Based on the effect size reported by Coltheart et al., 44 participants were considered sufficient to achieve a power estimate of 0.95. Ages ranged from 18 to 35 years (mean = 23.5, SD=3.2). Participants were right-handed, undergraduate psychology students.

Design. The experiment used a 2 (Condition: Phoneme overlap, No phoneme overlap) by 2 (Overlap Position: initial, end) by 2 (Response type: manual, vocal) within-subjects design.

Materials. The 184 stimuli used in the present study were taken directly from the Coltheart et al. study. The full list of items, with a full description, is reported in Coltheart et al. (1999). There were 92 stimuli in the Phoneme overlap condition, half of which possessed phonemic overlap with the initial phoneme of one of the three colours names (red, blue, green; e.g. rack, bait, gasp) the other half overlap with the final phoneme (e.g. cud, skin, few). The stimuli were selected on the bases of having meeting the following criteria: 1) Being between 3 and 5 letters in length; 2) Having the same number of phonemes as the colour name; 3) Having an absence of phonemic overlap with any other phoneme in the colour name; 4) Lack any obvious colour association. The remaining 92 stimuli were match to the critical stimuli on number of letters and phonemes and contained no phonemic overlap with the colour names. Each control stimulus was presented only in the same colour as its matched critical stimulus.

All stimuli were presented in lowercase via Superlab Pro Version 2.0 software running on a PC desktop computer. The ink colours chosen were red (RGB: 255, 0, 0), green (RGB: 0, 255, 0), and blue (RGB: 0, 0, 255). Three-letter stimuli were approximately 15mm long x 11mm high, four-letter stimuli were approximately 22mm long x 10mm high, and five-letter stimuli were approximately 29mm long x 10mm high. During a practice session, the stimuli consisted of a string of five repeated X symbols presented on a gateway solo P2-550 laptop pc using a TFT screen with a refresh rate of 60 Hz. For the manual response condition the response keys were labelled green, blue and red using colour patches placed over the 'z', 'v', and 'm' keys, respectively. Half of the participants used two fingers of their right hand to respond and one finger from their left hand, whilst the other half used two fingers from their left hand and one finger from their right hand. For the vocal response condition, a voice key connected to the laptop using a microphone with a voice activated trigger was used to record the responses and responses were also taped to enable the experimenter to check for errors.

Procedure. Each participant sat approximately 60 cm away from a screen. They were asked to identify the colour of the stimulus print, which would be red, green, or blue, quickly and as accurately. Each stimulus was preceded by a fixation cross appearing at the centre of the screen for 497 msec. The stimulus then appeared and was replaced by a blank screen as soon as a participant initiated their response. The blank screen was subsequently replaced after 1000msec by a fixation cross. Participants completed 184 trials of both manual and vocal responses giving a total of 368 trials. The same stimuli were presented in each condition i.e. for both response modes. Half of the participants performed the manual task first and half the vocal task first.

Data analysis. Following Coltheart et al. (1999) the data were analysed using a repeated measures ANOVA and for which the dependent variable was reaction time. However, in response to reviews of a previous version of this manuscript we also carried out

Linear Mixed Effects (LME) Modelling of the data. The benefit of LME over ANOVA is that is models random effects, reducing the Type 1 error rate. A full description of the steps taken during the analysis are available in the Supplementary Materials and on request**.** We used the lme4 package (https://cran.r-project.org/web/packages/lme4/) with R software (R Development Core Team, 2008) to analyze RT data. We followed the practice recommended by Bates, Kliegl, Vasishth, and Baayen (2015). In contrast to the ANOVA analysis**, o**ur dependent variable was *log* RT following the mainstream tradition in psycholinguistics (Baayen, 2008; Baayen, Davidson, & Bates, 2008. Finally, we calculated *p* values for this model (see Table 2) with the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017).

Results

Following Coltheart et al. (1999) analysis was conducted on correct reaction times (RTs) only. 6.3% and 2.6% of the vocal and manual data, respectively, were classed as errors. Following Coltheart et al. (1999) reaction times 2 standard deviations from the mean were defined as outliers and removed before analysis. This outlier removal procedure was done independently for the manual and vocal responses. For the manual responses, 3.9% of the data were classed as outliers. For the vocal response, 7.4% of the data were classed as outliers. See Table 1 for reaction times and standard errors in each condition.

Analysis of latencies. The data were entered into a 2 (Condition: Phoneme overlap, No phoneme overlap) x 2 (Response type: manual, vocal) x 2 (Overlap position: initial-letter, end-letter) repeated measures ANOVA. There was a main effect of Condition $F(1, 43) = 23.225$, $p < .001$, partial $\eta^2 = .351$, indicating that the words with phoneme overlap were responded to faster than the words with no phoneme overlap, and a main effect of Response type $F(1, 43) = 106.390$, $p < .001$, partial $\eta^2 = .712$, indicating that the vocal responses were longer than the manual responses, and a main effect of Overlap position $F(1, 43) = 10.275$,

p = .003, partial η^2 = .193.

There was a Condition x Response type interaction, F(1, 43) = 8.023, p = .007, partial η^2 = .157, indicating that the vocal response led to greater overall phonological facilitation (13ms; irrespective of position of overlap) than the manual response (3ms). The two-way Condition x Overlap position interaction was significant, F(1, 43) = 41.290, p<.001, partial η^2 = .490, indicating that there was Stroop facilitation at the initial position (15ms), but not at the end position (-4ms) when collapsed across response modality. The Overlap position x Response type interaction was not significant, F(1, 43) = 2.020, p > .1, partial η^2 = .045. There was a trend toward a three-way interaction where F(1, 43) = 3.795, p = .058, partial η^2 = .081.

Since the main aim of the present work was to establish whether phonological processing contributes to Stroop effects with manual responses we decomposed the three-way interaction into two 2 (Condition: Phoneme overlap, No phoneme overlap) x 2 (Overlap position: initial-letter, end-letter) repeated measures ANOVAs. For the manual response, there was no main effect of Condition F(1, 43) = 1.049, p = .311, partial η^2 = .024, and no main effect of Overlap position F(1, 43) = 1.264, p = .267, partial η^2 = .029, but the interaction was significant (F(1, 43) = 7.733, p = .008, η^2 = .152. The interaction was the result of significant phonological facilitation at the initial letter position (9ms; t(43) = 2.653, p = .011), but not at the end position (-4ms; t(43) = -.989, p = .328). The significant facilitation at the initial letter position supports the notion that phonological processing of the irrelevant word occurs when participants respond manually in the Stroop task.

For the vocal response, there was a main effect of Condition F(1, 43) = 30.967, p < .001, partial η^2 = .419, and a main effect of Overlap position F(1, 43) = 10.780, p = .002, partial η^2 = .2. The interaction was also significant (F(1, 43) = 32.299, p < .001, η^2 = .429. As with the manual response, the interaction was the result of significant phonological facilitation at the initial letter position (27ms; t(43) = 7.831, $p < .001$), but not at the end position (1ms; t(43) = .201, $p = .841$).

Linear Mixed Effects Modelling

As in other analyses using LME we first started with a maximal model (Barr, Levy, Scheepers, & Tily, 2013) entering all the fixed variables including Word Type, Overlap Position and Response Type and the control variable (colour), together with random intercepts for the items and the subjects, plus by-subject and by-item random slopes for all the fixed effects. The interaction between the fixed variables was also included in the models following Barr et al. (2013).

Next, we calculated a Singular Value Decomposition (SVD) on the covariance matrix of the maximal model using the PCA (Bates et al., 2015). SVD analysis reveals which slopes in the random effects structure do not contribute to the model. This removes unnecessary complications such as over-specification from the model, elimination of redundant analyses, and makes the model easier to converge overall. SVD analysis is accompanied by a Likelihood Ratio Test (LRT) during evaluation. After the SVD specifies which slopes have a variance close to zero, they are then removed from the model one at a time. Using a series of LRTs, the new models are compared with previous models to determine whether the removal of a slope(s) makes any significant changes to the model fit.

SVD analysis and then LRTs suggested a random effects structure which consisted of both random intercepts for subjects and items, plus by-subject and a by-item random slopes for Response Type. This means that the effect of Response Type was variable across subjects and items. However, because of a singularity problem and a ceiling correlation between the item intercept and its slope, we needed to remove the by-item slope from the model. This did not have a significant impact on the model estimates because the removed variance was contributing to less than 5% of the whole random effects variance. The resulting structure is then assumed to explain our data using the best fit from among most other alternatives.

After specifying the best random effects structure for our data, we then started looking into the main effects and their interactions. Determining which main effects and interactions are significant is a controversial issue in LME research. We followed the practice suggested by Baayen et al., (2008). LRT showed that the main effects of colour ($x^2(2)=146.85$, $p <0.001$), Word Type ($x^2(1)=12.75$, $p <0.001$), Overlap Position ($x^2(2)=8.94$, $p <0.01$), and Response Type ($x^2(1)=41.79$, $p <0.001$) were significant. When it comes to interactions, Word Type* Overlap Position ($x^2(1)=24.89$, $p <0.001$), Response Type* Overlap Position ($x^2(1)=9.45$, $p <0.01$), and Word Type*Response Type ($x^2(1)=7.56$, $p <0.01$) were significant. The three-way interaction was not significant ($x^2(1)=1.57$, $p =0.20$). We used Akaike Information Criterion (AIC) as the criterion to select the best model which explained our data best. The following model was the best model since it had the lowest AIC:

RTlog ~ Colour + Word Type + Overlap Position + Response Type + Word Type * Overlap Position + Overlap Position*Response Type + Word Type*Response Type + (1 +Response Type|Subject) + (1|Item).

Summary of the analyses

The results from the ANOVA and LME modelling were consistent for all key effects including the non-significant three-way interaction. Whilst the three-way interaction trended towards significance with the ANOVA, this was not the case for the LME modelling which has better control over Type 1 error rate.

Discussion

The aim of the present work was to investigate whether the manual response colour-word Stroop task leads to phonological Stroop effects. Using a fully-within subjects design, we replicated the serial position Stroop facilitation effect reported by Coltheart et al. (1999)

using a vocal response with both manual and vocal responses,. The results support the notion that manual responding in the Stroop task does produce phonological Stroop facilitation when the irrelevant word shares phonemes with the colour name despite the fact that participants are not required to produce the name of the relevant colour. The results are consistent with models that predict verbal mediation of manual responses and thus similar components comprising Stroop interference in both response modes (Roelofs, 2003). Our data suggest that both manual and vocal responding lead to phonological processing of the irrelevant word and that this processing facilitates the process of phonologically encoding the colour name.

Our data are inconsistent with the notion that manual and vocal responding involve different task goals that determine the nature of the evidence to be accumulated (Kinoshita et al., 2017) and that manual and vocal responding have differential access to the systems that produce Stroop effects (Fennell & Ratcliff, 2019; Glaser & Glaser, 1989; Kinoshita, De Wit & Norris, 2017; McClain, 1983; Redding & Gerjets, 1977; Sharma & McKenna, 1998; Zahedi, Abdel Rahman, Stürmer & Sommer, 2019). However, one clear and obvious difference between the manual and vocal response modes is the need to actually utter the colour name with a vocal response. This requires the full repertoire of speech production including syllabification and articulation. The engagement of speech production processes might modify the way the irrelevant word is processed (Burt, 2002; VanVoorhis & Dark, 1995). Thus, whilst our data and those from others (e.g., Brown & Besner, 2001) support the notion that both manual and vocal responding might involve access to phonology and semantics, leading to phonologically and semantically based Stroop effects, a key difference might be the increased access to these systems produced by feedback from the phonological encoding of the relevant colour name. Whilst we do not report a significant difference in the amount of phonological processing between the two response modes our data are somewhat consistent with a potential quantitative difference.

In contrast to Coltheart et al.'s results, facilitation at the end letter position in the present study was small (vocal) and even negative (manual; see Figure 1). Closer examination of Marmurek et al. (2006)'s data also reveals they observed a similar lack of facilitation from final position overlap (with a vocal response). We used a Bayes Factor (Dienes, 2011) to determine if each response type showed no evidence of an effect of phoneme overlap at the end position. Here, $B_{H(0, x)}$ refers to a Bayes factor in which the predictions of H1 were modelled as a half-normal distribution with an SD of *x* (see Dienes, 2014, 2016). To calculate the Bayes Factors we used a half normal of 12ms which was the raw effect size reported by Coltheart et al. (1999). For the manual response a Bayes Factor of $B_{H\ (0,12)} = .15$ was returned confirming no facilitation. For the vocal response a Bayes Factor of $B_{H\ (0,12)} = .31$ was returned again confirming no facilitation. Our data are therefore somewhat inconsistent with the findings from Coltheart et al. (1999) who reported a final position overlap effect. Our data do not enable us to determine why this difference has been observed but this could be an avenue for future research.

In conclusion, the present results are not consistent with a qualitative difference between manual and vocal responding in the Stroop task with regards to phonological processing. When there is an overlap between the phonemes of the irrelevant word and the colour name at the initial letter position, phonological facilitation is observed with both response modes. We interpret our data as being consistent with Roelofs' (2003) model of the Stroop task which argues for verbal mediation of the manual Stroop task.

Open Practices Statement

The minimal dataset underlying the findings in the manuscript (i.e. the values behind statistics, graphs, figures, and tables) and a description of the Linear Mixed Effects analysis steps are freely available to other researchers on the Open Science Framework (https://osf.io/2xzqd/?view_only=c9fae46897fa49e9823c91767e78e8d2). There are no ethical or legal restrictions. Requests for data may be sent to the corresponding author. The authors declare no conflict of interest. This experiment was not preregistered. Any original materials used to conduct the research (including analysis code) will be made available to other researchers for purposes of replicating the procedure or reproducing the results.

References

Augustinova, M., & Ferrand, L. (2014). Automaticity of Word Reading: Evidence from the Semantic Stroop Paradigm. *Current Directions in Psychological Science, 23,* 343 348.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics.* Cambridge: Cambridge University Press.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390-412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keeping it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. *arXiv e-prints*. Retrieved from https://ui.adsabs.harvard.edu/\#abs/2015arXiv150604967B

Besner, D., & Stolz, J. A. (1998). Unintentional reading: Can phonological computation be controlled? *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 52*(1), 35-43.

Brown, M., & Besner, D. (2001). On a variant of Stroop's paradigm: Which cognitions press your buttons? *Memory & Cognition, 29*(6), 903-904.

Burt, J. S. (2002). Why do non-colour words interfere with colour naming? *Journal of Experimental Psychology-Human Perception and Performance, 28*(5), 1019-1038.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.

Coltheart, M., Woollams, A., Kinoshita, S., & Perry, C. (1999). A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. *Psychonomic Bulletin & Review, 6*(3), 456-463.

Fennell, A. & Ratcliff, R. (2019). Does response modality influence conflict? Modelling vocal and manual response Stroop interference. *Journal of Experimental Psychology: Learning, Memory and Cognition.* Advanced online publication. http://dx.doi.org/10.1037/xlm0000689

Glaser, W. R., & Glaser, M. O. (1989). Context Effects in Stroop-like Word and Picture Processing. *Journal of Experimental Psychology-General, 118*(1), 13-42.

Keele, S. W. (1972). Attention demands of memory retrieval. *Journal of Experimental Psychology, 93*, 245-248.

Kinoshita, S. (2003). The nature of masked onset priming effects in naming: A review. In S. Kinoshita & S. J. Lupker (Eds.), *Macquarie monographs in cognitive science. Masked priming: The state of the art* (pp. 223-238). New York, NY, US: Psychology Press.

Kinoshita, S. De Wit, B., & Norris, D. (2017). The magic of words reconsidered: Investigating the automaticity of reading colour-neutral words in the Stroop task. *Journal of Experimental Psychology-Learning, Memory and Cognition, 43*(3), 369-384.

Kuznetsova, Brockhoff, & Christensen (2017) lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26.

MacLeod, C. M. (1991). Half a Century of Research on the Stroop Effect: An Integrative Review. *Psychonomic Bulletin, 109*(2), 163-203.

Marmurek, H. H. C., Proctor, C. & Javor, A. (2006). Stroop-like serial position effects in colour naming of words and nonwords. *Experimental Psychology, 53*(2), 105-110.

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word frequency effects in naming. *Journal of Experimental Psychology: Human Perception & Performance, 13*, 14-24

McClain, L. (1983). Effects of response type and set size on Stroop color-word performance. *Perceptual* & *Motor Skills*, 56, 735-743.

Nilipour, R., Bakhtiar, M., Momenian, M., & Weekes, B. S. (2017). Object and action picture naming in brain-damaged Persian speakers with aphasia. *Aphasiology, 31*(4), 388-405.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*, 327– 357.

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*, *137*, 434 – 455.

Paap, K. R., Noel, R. W., & Johansen, L.S. (1992). Dual route models of print to sound: Read herrings and real horses. In R. Frost and L. Katz (Eds.) *Orthography, Phonology, Morphology and Meaning.* North Holland Press, C.A.

Parris, B. A., Hasshim, N., Wadsley, M., Augustinova, M. & Ferrand, L. (submitted). Intra-Stroop conflict: A review of varieties of conflict (and facilitation) and their measurement in the colour-word Stroop task.

Patterson, K., Suzuki, T., & Wydell, T.N. (1996). Interpreting a case of Japanese phonological alexia: the key is in phonology. *Cognitive Neuropsychology, 13*, 803 822.

Redding, G. M., & Gerjets, D. A. (1977). Stroop effects: Interference and facilitation with verbal and manual responses. *Perceptual & Motor Skills, 45*, 11-17.

Regan, J. (1978). Involuntary automatic processing in colour-naming tasks. *Perception & Psychophysics, 24*(2), 130-136.

Reynolds, M. G. & Besner, D. (2005). Basic processes in reading: a critical review of pseudohomophone effects in reading aloud and a new computational account. *Psychonomic Bulletin and Review, 12*, 622-646.

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.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review, 110*(1), 88.

Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition, 26*(5), 1033-1040.

Singer, M. H., Lappin, J. S., & Moore, L. P. (1975). The interference of various word parts on colour naming in the Stroop test. *Perception & Psychophysics, 18*(3), 191-193.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology,18*(6), 643.

Sugg, M. J., & McDonald, J. E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 20*(3), 647-675.

Tree, J. J., Longmore, C., Majerus, S. &, Evans, N. (2011). SKOOL versus ZOOL: Effects of orthographic and phonological long term memory on nonword immediate serial recall. *Memory, 19,* 487-501.

Tree, J. J., & Playfoot, D. (2015). Strategic effects on pseudohomophone reading in phonological dyslexics with and without phonological impairment. *Journal of Neurolinguistics, 35*, 39-54.

Turken, A. U., & Swick,. D. (1999) Response selection in the human anterior cingulate cortex. *Nature Neuroscience, 2,* 920-924.

Tzelgov, J., Henik, A., Sneg, R., & Baruch, O. (1996). Unintentional word reading via the phonological route: The Stroop effect with cross-script homophones. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(2), 336-349.

Van Voorhis, B. A., & Dark, V. J. (1995). Semantic Matching, Response-Mode, and Response Mapping as Contributors to Retroactive and Proactive Priming. *Journal of Experimental Psychology-Learning Memory and Cognition, 21*(4), 913-932.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition, 13*(4), 304-319.

Zahedi, A., Abdel Rahman, R., Stürmer, B., & Sommer, W. (2019). Common and specific loci of Stroop effects in vocal and manual asks, revealed by event-related brain potentials and posthypnotic suggestions. *Journal of Experimental Psychology: General*. Advance online publication. http://dx.doi.org/10.1037/xge0000574

Table 1: Mean reaction times (ms), standard errors (in parentheses), and magnitudes of Stroop facilitation when identifying colour as a function of the position of phonological overlap (Initial- or End-letter) between the irrelevant word and the colour name.

		VOCAL		MANUAL	
		Position of overlap		*Position of overlap*	
Condition		*Initial-letter*	*End-letter*	*Initial-letter*	*End-letter*
Control (no overlap)	RT SE	614 (9.1)	609 (8.1)	533 (9.5)	529 (10.3)
Critical (overlap)	RT SE	587 (8.2)	608 (3.7)	524 (9.5)	533 (9.5)
Stroop Facilitation		27	1	9	-4

Table 2. Summary of the mixed and random effects models

	Estimate	**Std. error**	**t value**	**P***
Fixed effects				
Position (Initial)	0.007	0.002	2.74	0.006
Word Type (Critical)	0.006	0.002	2.48	0.013
Response Type (Vocal)	0.065	0.007	9.22	0.001
Colour (Green)	0.013	0.002	6.33	0.001
Colour (Red)	-0.019	0.001	-10.08	0.001
Position*Word Type	-0.016	0.003	-5.16	0.001
Word Type*Response Type	-0.008	0.002	-2.75	0.001
Position*Response Type	-0.009	0.002	-3.08	0.001
Random effects	**Variance**	**SD**	**Correlation**	
Item (Intercept)	0.00001	0.003		

Subject (Intercept)	0.00339	0.582		
Subject (Response Type slope)	0.00192	0.438	-0.68	
Residual	0.00820	0.905		

* Calculated with lmerTest in R.