

Tracking Eye Movements to Localize Stroop Interference in Naming: Word Planning Versus Articulatory Buffering

Ardi Roelofs
Radboud University Nijmegen

Investigators have found no agreement on the functional locus of Stroop interference in vocal naming. Whereas it has long been assumed that the interference arises during spoken word planning, more recently some investigators have revived an account from the 1960s and 1970s holding that the interference occurs in an articulatory buffer after word planning. Here, 2 color-word Stroop experiments are reported that tested between these accounts using eye tracking. Previous research has indicated that the shifting of eye gaze from a stimulus to another occurs before the articulatory buffer is reached in spoken word planning. In the present experiments, participants were presented with color-word Stroop stimuli and left- or right-pointing arrows on different sides of a computer screen. They named the color attribute and shifted their gaze to the arrow to manually indicate its direction. If Stroop interference arises in the articulatory buffer, the interference should be present in the color-naming latencies but not in the gaze shift and manual response latencies. Contrary to these predictions, Stroop interference was present in all 3 behavioral measures. These results indicate that Stroop interference arises during spoken word planning rather than in articulatory buffering.

Keywords: articulatory buffer, eye movements, interference, naming, word planning

An important tool in studying naming performance is the color-word Stroop task (Stroop, 1935), one of the most widely employed tasks in academic and applied psychology. In a commonly used version of this task, participants name the color attribute of colored incongruent color words (e.g., the printed word *red* in green color; say “green”) or neutral series of Xs or neutral unrelated words (e.g., *house* in green color). Mean response time (RT) is longer on incongruent than neutral color-naming trials, descriptively called *Stroop interference* (for reviews, see MacLeod, 1991; Roelofs, 2003). According to a prominent account, colors (e.g., the color green) activate a set of color names, not only the correct name (*green*) but also related ones (e.g., *red* and *blue*), which compete for selection. On incongruent trials (e.g., the word *red* in green color), the written word (i.e., *red*) boosts the activation of one of these competing names (*red*), which prolongs the duration of selecting the target name (*green*) compared with neutral trials, surfacing as Stroop interference in the RTs. This competition account of Stroop interference has been computationally implemented in a number of models, including WEAVER++ (Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992, 1997, 2003, 2007, 2008a, 2008b, 2008c).

According to the WEAVER++ model, planning the name of a color involves conceptualizing, lemma retrieval, and word-form

encoding, which is followed by articulation, as outlined in Figure 1. Lemma retrieval and word-form encoding make up lexical access, with word-form encoding comprising morphological, phonological, and phonetic encoding. Information about words is assumed to be stored in a large associative network, which is accessed by spreading activation. In naming a color, activation spreads from nodes for color concepts to lemmas, morphemes, phonemes, and articulatory programs. For example, naming the color green involves the activation and selection of the corresponding color concept during conceptualizing, the lemma of the color name *green* (e.g., specifying that the word is an adjective) during lemma retrieval (this process is also called *lexical selection*), the morpheme <green> during morphological encoding, the phonemes /g/, /r/, /i:/, and /n/ during phonological encoding, and the articulatory program [gri:n] during phonetic encoding. According to the model, lemmas and articulatory programs are selected by competition (Levelt et al., 1999; Roelofs, 1992, 1997, 2003, 2010a, 2010b). Relative to neutral Xs, competition prolongs the duration of word planning on incongruent trials (e.g., the word *red* in green color), when the distractor word activates a competing lemma and articulatory program. This yields Stroop interference in the naming RTs. In the model, most interference arises in lemma retrieval, whereas competition in selecting articulatory programs only makes a small contribution to the total amount of Stroop interference (Roelofs, 2003).

Evidence suggesting a contribution of lexical response selection (i.e., lemma retrieval) to the Stroop effect comes from, among others, the finding that a considerable part of the interference caused by incongruent stimuli is specific to the members of the response set. In a classic study, Klein (1964) observed that color words that were eligible responses produced approximately 2 times more interference on color naming than color words that were not used as responses in the experiment. For example, when

This article was published Online First April 21, 2014.

I thank Bicoor Bolla-Bong for his help in running the experiments and the members of the Attention and Language Performance Lab at the Donders Centre for Cognition for helpful comments.

Correspondence concerning this article should be addressed to Ardi Roelofs, Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Spinoza Building B.02.34, Montessorilaan 3, 6525 HR Nijmegen, the Netherlands. E-mail: A.Roelofs@donders.ru.nl

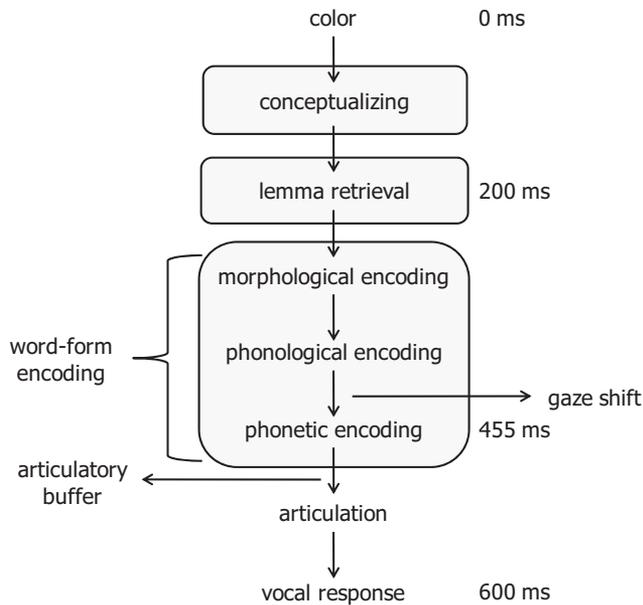


Figure 1. Word planning stages in color naming according to the WEAVER++ model (e.g., Roelofs, 2003). The numbers to the right indicate onset estimates by Indefrey and Levelt (2004; Indefrey, 2011).

the ink colors are red and green, color-naming RTs are much longer for the word *green* in red ink than for the word *blue* in red ink (cf. W. R. Glaser & Glaser, 1989; Lamers, Roelofs, & Rabeling-Keus, 2010). Evidence suggesting a contribution of word-form encoding to the Stroop effect comes from, among others, phonological effects in the Stroop task (e.g., Coltheart, Woollams, Kinoshita, & Perry, 1999; Dennis & Newstead, 1981; Roelofs, 2003). For example, Coltheart et al. (1999) observed that color-naming RTs are shorter for phonologically related words, such as the rhyming word *skew* in blue ink, than phonologically unrelated words (e.g., *fist* in blue). Moreover, RTs are longer for words that are phonologically related to other color names (e.g., *skew* in red ink) than unrelated words (e.g., Roelofs, 2003). It has been suggested that part of the Stroop interference effect arises during conceptualizing (e.g., Augustinova & Ferrand, 2012; Klein, 1964; Seymour, 1977), but the amount of interference arising at this level is much smaller than during word planning. Moreover, it is unclear whether these semantic contributions arise at the conceptual level or are only mediated by it (Roelofs, 2003). Either way, the available evidence suggests that Stroop interference arises, for the most part, during the word planning process.

Recently, however, some researchers have proposed an alternative explanation of the Stroop interference effect, one that does not assume lexical competition and a locus in word planning (Mahon, Costa, Peterson, Vargas, & Caramazza, 2007; Mahon, Garcea, & Navarrete, 2012), called the *response-exclusion* account (e.g., Dhooge, De Baene, & Hartsuiker, 2013; Dhooge & Hartsuiker, 2010, 2011; Finkbeiner & Caramazza, 2006a, 2006b; Janssen, Schirm, Mahon, & Caramazza, 2008; Miozzo & Caramazza, 2003). This alternative explanation revives an account of Stroop interference proposed in the 1960s and 1970s (e.g., Dyer, 1973; Klein, 1964; Morton & Chambers, 1973). The account

proposes that the word component of a Stroop stimulus is automatically processed up to the activation of its articulatory program. This activation competes against the articulatory program of the picture- or color-naming response for control of the response execution. The Stroop inhibition is explained as an extra amount of time needed to suppress the distractor-related response or to clear a response buffer from its program. (W. R. Glaser & Dünghoff, 1984, p. 641)

According to the response-exclusion account (e.g., Finkbeiner & Caramazza, 2006a, 2006b; Janssen et al., 2008; Mahon et al., 2007, 2012), a word is selected if its activation exceeds some threshold, but selection is taken to be independent of the activation state of other words. As argued by Mahon et al. (2012) for the Stroop task, interference in color naming arises in an articulatory output buffer after the target color name has been planned, reflecting the difficulty of excluding an articulatory response to the distractor word from the output buffer. Mahon et al. (2007) stated,

In the case of the Stroop and picture-word interference tasks, printed words, compared with colors or pictures, have privileged access to the articulators. . . . On this account, the target response (the picture or color name) can be produced only if the single-channel output buffer is not occupied by a representation corresponding to the distractor word. (p. 524)

According to Finkbeiner and Caramazza (2006a), “interference may arise at the point of deciding which of two articulatory programs should be excluded from the output buffer in order that the correct response may be produced” (p. 1033). This exclusion process is stipulated to take longer when the distractor is semantically related to the target (e.g., the word *red* in green color; red and green are members of the semantic category color) than when it is unrelated (e.g., the word *house*) or a series of Xs. This delays responding on incongruent relative to neutral trials. Whereas the competitive planning account assumes that a semantic relation between target and distractor (e.g., *red* and *green*) increases competition in lexical selection, the response-exclusion account assumes that a semantic relation facilitates lexical selection (Finkbeiner & Caramazza, 2006a, 2006b; Mahon et al., 2012). Because the word *red* activates the target name *green* whereas the word *house* and a series of Xs do not, the target name *green* exceeds the lexical selection threshold earlier with the distractor word *red* than the word *house* or a series of Xs. Still, under the response-exclusion account, Stroop interference rather than facilitation is observed in the naming RTs on incongruent relative to neutral trials because the facilitation of lexical selection is offset by a larger interference effect arising during the removal of the articulatory program for the distractor from the articulatory buffer (Finkbeiner & Caramazza, 2006b; Mahon et al., 2012).

The response-exclusion account concerns a claim about the locus of selective attention in Stroop task performance, assuming that the target rather than the distractor gains control over vocal responding at the level of the articulatory buffer. Stroop interference arises and is resolved at this postplanning level, close to articulation onset. In contrast, attentional mechanisms operate during spoken word planning in a model like WEAVER++ (Roelofs, 2003). According to the competitive planning account, Stroop interference arises and is resolved during word planning rather than at the postplanning level of the articulatory buffer. During the past several years, researchers have tried to adjudicate between the

competitive planning and response-exclusion accounts using a picture-word analogue of the color-word Stroop task (e.g., [Abdel Rahman & Aristei, 2010](#); [Dhooge et al., 2013](#); [Dhooge & Hart-suiker, 2010, 2011](#); [Finkbeiner & Caramazza, 2006a, 2006b](#); [Hutson, Damian, & Spalek, 2013](#); [Janssen, 2013](#); [Janssen et al., 2008](#); [La Heij, Kuipers, & Starreveld, 2006](#); [Mädebach, Oppermann, Hantsch, Curda, & Jescheniak, 2011](#); [Mahon et al., 2007](#); [Miozzo & Caramazza, 2003](#); [Mulatti & Coltheart, 2012](#); [Piai, Roelofs, & Schriefers, 2011, 2012](#); [Roelofs, Piai, & Schriefers, 2011, 2013a, 2013b](#)). This debate is still going on. In the picture-word task, participants name pictures while trying to ignore superimposed distractor words. Mean naming RTs are typically longer in a semantic condition (e.g., a pictured cat combined with the word *dog*) than an unrelated condition (e.g., a pictured cat combined with the word *house*), longer in the semantic condition than a neutral control condition (e.g., a pictured cat combined with a series of Xs), and longer in the semantic condition than an identity condition (e.g., a pictured cat combined with the word *cat*). Note that the semantic condition corresponds to the incongruent condition of the color-word Stroop task (i.e., cat and dog are from the same semantic category, just like red and green are) and that the identity condition corresponds to the congruent condition of the Stroop task (cf. [Piai, Roelofs, & Van der Meij, 2012](#); [Roelofs, 2003](#)).

More recently, proponents of the response-exclusion hypothesis have extended the debate to the color-word Stroop task and have examined Stroop performance to adjudicate between their hypothesis and the competition account ([Mahon et al., 2012](#)). Subsequently, [Mulatti and Coltheart \(2014\)](#) and [Roelofs and Piai \(2013\)](#) have also evaluated the response-exclusion and competition hypotheses using color-word Stroop task data. The present article follows the lead of proponents of the response-exclusion hypothesis (i.e., [Mahon et al., 2012](#)) to use the Stroop task to adjudicate between the two accounts. In particular, two experiments are reported that tested predictions derived from the response-exclusion and competition accounts concerning color-word Stroop interference.

Although the picture-word interference task and the color-word Stroop task may seem very similar at first sight, the apparent similarity does not mean that the underlying locus is the same. In particular, [Dell'Acqua, Job, Peressotti, and Pascali \(2007\)](#) argued that the functional locus of picture-word interference and color-word interference differs. However, this claim of Dell'Acqua et al. was based on a comparison between their own picture-word interference data and earlier data from the color-word Stroop task of [Fagot and Pashler \(1992\)](#), differing in many methodological respects. In recent picture-word interference studies, [Schnur and Martin \(2012, two experiments\)](#) and [Piai, Roelofs, and Schriefers \(2014, six experiments\)](#) failed to replicate the findings of Dell'Acqua et al. Moreover, a direct experimental comparison of color-word Stroop and picture-word interference by [Piai et al.](#) yielded exactly the same results for the two tasks, replicating the Stroop data of [Fagot and Pashler](#) but not the picture-word interference data of [Dell'Acqua et al.](#) Furthermore, [Piai et al.](#) systematically manipulated several major methodological differences between picture-word interference and color-word Stroop, including number of stimuli, repetitions, and so forth, and observed that none of these differences mattered, that is, the findings for Stroop and picture-word interference remained equivalent. To conclude, there

is good evidence that the picture-word and color-word tasks are equivalent in respects that are relevant for the present discussion.

To summarize, investigators have found no agreement on whether spoken word planning involves competition and on the functional locus of the Stroop interference effect in naming, that is, whether the effect arises during or after spoken word planning. The aim of the experiments reported in the present article was to assess the relative merits of the competition and response-exclusion accounts of Stroop interference.

Time Course of Stroop Interference

A critical difference between the competition and response-exclusion accounts concerns the time course of the Stroop interference effect. Whereas the response-exclusion account maintains that Stroop interference arises in an articulatory buffer after word planning and close to articulation onset, the competition account claims that the interference arises earlier, during the word planning process. According to an influential estimation of the relative timing of processing stages underlying spoken word planning by [Indefrey and Levelt \(2004; Indefrey, 2011\)](#), lemma retrieval begins about 200 ms after target color onset, whereas phonetic encoding starts about 145 ms before articulation onset, as indicated in [Figure 1](#). During phonetic encoding, articulatory programs are accessed and may be placed into the articulatory buffer. Accessing the programs will take some time. Thus, the articulatory buffer is reached somewhat later than the onset of phonetic encoding (i.e., later than 145 ms before articulation onset). Nevertheless, in the remainder of the article, 145 ms will be used as a rough estimate of the moment that the articulatory buffer is reached. Later, I make clear that the conclusions drawn from the results of the present experiments hold regardless of whether the articulatory buffer is reached 145 ms before articulation onset or somewhat later. In sum, whereas the response-exclusion account claims that Stroop interference should emerge no earlier than about 145 ms before articulation onset, the competition account maintains that the interference should start to arise much earlier, about 200 ms after color-word onset.

The functional locus of an experimental effect may be assessed by making multiple measurements during the course of response planning and execution (cf. [Meyer, Osman, Irwin, & Yantis, 1988](#)). This was also done in the experiments reported in the present article, which exploited the finding that in dual-task performance with spatially separated task stimuli, a shifting of eye gaze from a color-word Stroop stimulus to another stimulus happens before the articulatory buffer is reached in spoken word planning ([Roelofs, 2011](#)), as is explained further below. If Stroop interference is present in both the gaze shifts and vocal responses, then the interference must arise during processes that are prior to the articulatory buffer, in particular, during spoken word planning, as maintained by the competition account. In contrast, if Stroop interference is present in the vocal responses but not in the gaze shifts, then the interference must have arisen during processes after the gaze shifts, in particular, in the articulatory buffer, as maintained by the response-exclusion account.

Although eye movements have long been measured in the history of experimental psychology (e.g., [Buswell, 1920](#); [Quantz, 1897](#); for reviews of the early literature, see [Woodworth, 1938](#); [Woodworth & Schlosberg, 1954](#)) and eye tracking predates the

inception of the Stroop task (Stroop, 1935), only a very few studies have measured eye movements during color-word Stroop task performance (e.g., Olk, 2013; Roelofs, 2011). Olk (2013) examined Stroop task performance using manual responding and observed that the Stroop effect was reflected in gaze durations. However, given that manual rather than vocal responding was used, these findings do not bear on the issue of the locus of Stroop interference in naming. In two eye-tracking experiments (Roelofs, 2011), I examined vocal responding, as in the classic version of the Stroop task (Stroop, 1935), which allows for an evaluation of the two accounts. Participants were presented with color-word Stroop stimuli and left- or right-pointing arrows on different sides of a computer screen, and they named the color attribute and shifted their gaze to the arrow to manually indicate its direction, as illustrated in Figure 2. The results showed that gaze shifts from the Stroop stimulus to the arrow were initiated about 215 ms before articulation onset (Roelofs, 2011). The distance that the eyes are ahead of speech is often referred to as the *eye-voice span* (e.g., Buswell, 1920; Levin & Buckler-Addis, 1979). This distance is not due to a difference in time needed for generating and initiating the execution of the corresponding motor programs, which takes some 150–175 ms for both saccadic eye movements and articulatory movements (Indefrey, 2011; Indefrey & Levelt, 2004; Rayner, 1998; Reichle, Pollatsek, Fisher, & Rayner, 1998; Roelofs, 2007). That is, it is not the case that saccadic eye movements are programmed and initiated 215 ms faster than articulatory movements. Given the evidence that word planning arrives at the level of the output buffer no earlier than 145 ms before articulation onset (Indefrey, 2011; Indefrey & Levelt, 2004), the gaze shifts in the study of Roelofs (2011) must have been initiated before the output

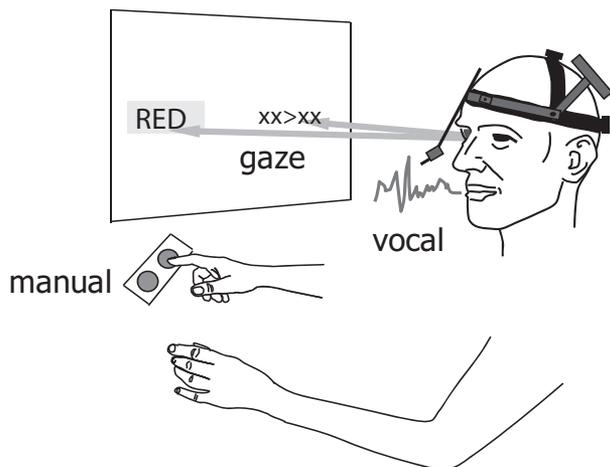


Figure 2. Schematic illustration of the dual-task procedure with spatially separated task stimuli. On each trial, participants name the color attribute of a color-word Stroop stimulus and shift their gaze to a left- or right-pointing arrow stimulus (i.e., < or > flanked by two Xs on each side) to manually indicate its direction. The latencies of the naming responses, gaze shifts, and manual responses are recorded. Adapted from “Attention, Exposure Duration, and Gaze Shifting in Naming Performance,” by A. Roelofs, 2011, *Journal of Experimental Psychology: Human Perception and Performance*, 37, p. 862. Copyright 2011 by the American Psychological Association.

buffer was reached. This allows for an evaluation of the competition and response-exclusion accounts of Stroop interference.

If gaze shifts happen before the articulatory buffer in naming is reached, the response-exclusion account predicts that Stroop interference should be present in the color-naming RTs on incongruent trials but not in the gaze shift latencies. In the gaze shifts, facilitation (i.e., shorter latencies on incongruent than neutral trials) should be obtained because a semantic relation between distractor word and target (e.g., the word *red* in green color) speeds up lexical selection, according to the response-exclusion account (Finkbeiner & Caramazza, 2006a, 2006b; Mahon et al., 2012). Recall that the facilitation of lexical selection by semantic relatedness (between *green* and *red*) is assumed to be offset by a larger interference effect in removing the articulatory program for the incongruent word from the output buffer, resulting in Stroop interference in the naming RTs. Thus, the response-exclusion account predicts Stroop interference in the color-naming RTs and facilitation in the gaze shift latencies. In contrast, the competition account predicts Stroop interference in both the naming RTs and gaze shift latencies (Roelofs, 2003, 2010a, 2011).

Contrary to the predictions by the response-exclusion account, Roelofs (2011) observed that Stroop interference was present in both the naming RTs and gaze shift latencies (manual RTs were not analyzed). This agrees with the predictions by the competition account. However, the magnitude of Stroop interference was somewhat smaller in the gaze shift latencies than the naming RTs (respectively, 92 ms vs. 141 ms in Experiment 1 and 102 ms vs. 115 ms in Experiment 2), which suggests that gaze shifts were initiated before the competition underlying the Stroop interference was fully resolved. A difference in Stroop effect between the gaze shifts and vocal responses is in agreement with the response-exclusion account, although the presence of a substantial amount of interference in the gaze shifts challenges the account (i.e., it predicts facilitation for the gaze shifts). The empirical finding concerning the different magnitude of Stroop interference in gaze shifting and naming may be related to an unusual feature of the experiments reported in Roelofs, which is that the color attribute was removed 100 ms after Stroop stimulus onset on half the trials. That is, the color attribute (e.g., green) was changed into neutral white color on a black background shortly after stimulus presentation onset. The removal of the color attribute on half the trials may have led to earlier gaze shifts than normal (i.e., before the competition underlying the Stroop interference was completely resolved), which may explain why Stroop interference was smaller in the gaze shift latencies than the naming RTs (see Roelofs, 2011, for an extensive discussion). According to the WEAVER++ model, Stroop interference arises because of competition in selecting lemmas (in lemma retrieval) and articulatory programs (in phonetic encoding). If gaze shifting happens before phonetic encoding, some of the Stroop interference will not be reflected in the gaze shifts. Consequently, the magnitude of Stroop interference will be smaller in the gaze shifts than the vocal responses, as observed (Roelofs, 2011).

To conclude, given the unusual duration of color exposure and the different magnitude of Stroop interference in the gazes and naming responses, the findings from the eye-tracking experiments of Roelofs (2011) do not provide conclusive evidence regarding the relative merits of the competition and response-exclusion accounts of the locus of the Stroop effect. It should be noted that

these experiments were designed to examine the effect of color exposure duration on Stroop interference but not to adjudicate between different accounts of the functional locus of Stroop interference. The aim of the experiments reported in the present article was to examine whether Stroop interference is present in the naming RTs, gaze shift latencies, and manual RTs when the colors are presented throughout a trial, as is typically the case in Stroop experiments. The experiments were specifically designed to test between the response-exclusion and competition accounts.

Plan of the Present Study

In the present article, two eye-tracking experiments are reported that used the dual-task procedure with spatially separated task stimuli that is illustrated in Figure 2. Naming RTs, gaze shift latencies, and manual RTs were recorded. The response-exclusion account predicts that Stroop interference should be present in the color-naming RTs on incongruent trials but that facilitation should be obtained for the gaze shift latencies and manual RTs (cf. Finkbeiner & Caramazza, 2006b; Mahon et al., 2012). In contrast, the competition account predicts that Stroop interference should be present in all three behavioral measures (cf. Roelofs, 2011). The latter is predicted because gaze shifts are initiated before the completion of word planning (i.e., before the end of lexical access), and word planning is the locus of Stroop interference under the competition account.

Moreover, according to the response-exclusion account, whether Stroop interference is obtained on incongruent trials critically depends on the relative duration of color and word processing. Previous research showed that words are read some 100–150 ms quicker than colors are named (e.g., M. O. Glaser & Glaser, 1982). Consequently, the motor program for the word will typically have reached the articulatory buffer before color-name planning reaches the buffer. Thus, according to the response-exclusion account, the motor program for the word has to be removed, which yields Stroop interference. Allowing word processing to start 200 or 300 ms earlier than color processing (i.e., further increasing the difference in timing between word and color processing) should eliminate Stroop interference because there would then be sufficient time to remove the articulatory program for the word from the articulatory buffer (which takes 200 ms, according to Dhooze & Hartsuiker, 2010) before the color name arrives. According to Dhooze and Hartsuiker (2010), “the control process operating over the response buffer takes no more than 200 ms to completely remove the response from the buffer” (p. 886). This estimate should hold regardless of whether the target is a picture or a color (i.e., in the Stroop task). Incongruent words will only interfere with color naming if the exclusion of the articulatory program for the word is still in progress when color-name planning reaches the articulatory buffer (i.e., later than 145 ms before articulation onset). If color naming takes 600 ms and the word is presented 300 ms before the color, Stroop interference will only be obtained if the removal process is still not completed after 755 ms (i.e., 455 ms + 300 ms) post word onset, which does not correspond to the estimates of Dhooze and Hartsuiker. Thus, according to the response-exclusion account, allowing word processing to start 200 or 300 ms earlier than color processing should eliminate Stroop interference. Interference should be absent in the vocal responses, gaze shifts, and manual responses.

Moreover, when the word is presented 200 or 300 ms later than the color, the motor program for the word will reach the buffer later than the color name. “Because of the privileged relationship of words with the articulators, any response that is already in the buffer upon the presentation of the distractor will be overwritten” (Dhooze & Hartsuiker, 2011, p. 117). Thus, at word postexposure stimulus onset asynchronies (SOAs), the motor program for the word will overwrite the motor program for the color name, which yields Stroop interference if this overwriting happens before articulation onset. Note that presenting the word 200 or 300 ms later than the color will further increase the time interval between gaze shifting and the moment that Stroop interference arises in the color-naming process. This should further decrease the possibility that Stroop interference is obtained in the gaze shifts and manual responses.

Experiment 1 tested these predictions derived from the response-exclusion account about the relative timing of word and color presentation by manipulating the SOA. To this end, colors were presented as colored rectangles, and the distractor words were superimposed onto these rectangles with a certain SOA (cf. M. O. Glaser & Glaser, 1982). According to the response-exclusion account, presenting the word 200 or 300 ms before color onset should eliminate Stroop interference in the naming RTs. The response-exclusion account predicts that Stroop interference should be absent in all three measures at the long distractor preexposure SOAs. Moreover, presenting the word 200 or 300 ms after color onset may yield Stroop interference in the naming RTs but should minimize the chance that the interference is observed in the gaze shifts and manual responses. In contrast, Stroop interference may be present in all three measures at these SOAs under the competition account. According to the WEAVER++ model (Roelofs, 2003, 2007, 2008a, 2008b), activation of nodes in the lexical network builds up and decays over time. Interference will be maximal at short distractor SOAs, when activations of target and distractor are both high, but some interference may still be obtained at long distractor preexposure and postexposure SOAs if activation of target and distractor overlaps. Different from the response-exclusion account, the exact relative timing of color and distractor word processing is not important for interference to occur, as long as there is some activation of the distractor word when the target is presented, which may be the case at long distractor preexposure and postexposure SOAs. Moreover, at these SOAs, activation of the distractor will still be boosted by the target color, contributing to the Stroop interference effect.

The aim of Experiment 2 was to replicate and extend the findings of the first experiment, creating conditions in which gaze shifts occur before as well as after articulation onset using the dual-task paradigm with spatially separated task stimuli (see Figure 2). Like Experiment 1, the second experiment tested the predictions derived from the response-exclusion and competition accounts concerning the presence of Stroop effects in the gaze shift latencies and manual RTs. However, rather than manipulating the SOA between color and distractor word presentation, henceforth referred to as the *distractor SOA*, the SOA between Stroop stimulus and arrow was manipulated, henceforth the *task SOA*, thereby creating a classic psychological refractory period (PRP) situation with spatially separated task stimuli (cf. Meyer & Kieras, 1997a, 1997b). In this second experiment, the color and word of the Stroop stimuli were presented simultaneously (i.e., only the dis-

tractor SOA of 0 ms was used). Evidence from Roelofs (2008a) suggests that in a situation with short and long task SOAs, gaze shifts are initiated before the onset of articulation at short task SOAs (i.e., when targets, here the Stroop stimulus and arrow, are presented close in time, i.e., at task SOAs of 0–300 ms) but after the onset of articulation at long task SOAs (i.e., when the presentation of targets, here the color-word stimulus and arrow, is separated by 1 s). Note that distractor SOAs of 200 or 300 ms are considered to be long, whereas task SOAs of 200 or 300 ms are considered to be short. In the literature on distractor SOAs in Stroop task performance, the long SOAs typically are 300 or 400 ms (e.g., M. O. Glaser & Glaser, 1982), whereas long task SOAs in the PRP literature are typically 1 s or more (e.g., Meyer & Kieras, 1997a, 1997b).

At long task SOAs of 1 s or more, the Stroop stimulus is presented at trial onset, but the spatial position of the arrow on the screen remains empty until the arrow is presented. In a picture-naming study of Roelofs (2008a), participants postponed the initiation of gaze shifts away from pictures (rather than Stroop stimuli) until close before arrow presentation onset at these long task SOAs, thereby apparently trying to avoid long fixations of an empty position in space. Thus, at long task SOAs, gaze shifts are initiated after word planning has reached the articulatory buffer. The response-exclusion account predicts that Stroop interference should be present in the color-naming RTs but absent in the gaze shift latencies at these long task SOAs. This is because gaze shifts happen well after the articulatory buffering stage at long task SOAs, leaving sufficient time to exclude the motor program for the distractor from the buffer before gaze shift onset. For the short task SOAs, the account predicts Stroop interference in the color-naming RTs but facilitation in the gaze shift latencies and the manual RTs. In contrast, the competition account predicts that Stroop interference should be present in all three measures at all task SOAs, except for the manual responses at long task SOAs. Elsewhere (Roelofs, 2008a), I argued that in order to avoid long eye fixations of an empty position in space, gaze shifts are delayed by a constant amount of time relative to the normal shift onset. As a consequence, Stroop interference will be reflected in the vocal responses and gaze shift latencies at all task SOAs but in the manual RTs only at short SOAs. At the long task SOAs, the eyes will fixate the spatial position of the arrow 100 ms or more before the arrow appears on both incongruent and neutral trials, which will absorb the Stroop interference (cf. Roelofs, 2008a).

Experiment 1

In the first experiment, participants were presented with color-word Stroop stimuli displayed on the left side of a computer screen and left- or right-pointing arrows (i.e., < or > flanked by two Xs) displayed on the right side of the screen (cf. Roelofs, 2007, 2008a, 2008b, 2011), as illustrated in Figure 2. Colors were presented as colored rectangles, and incongruent words or Xs were superimposed onto these rectangles. The participants' tasks were to vocally name the color rectangles and to manually indicate the direction in which the arrow was pointing by pressing a left or right button. The speed and accuracy of color naming and manual responding were recorded. In addition, eye movements were tracked in order to determine the onset of the shift of gaze from the color to the arrow. The SOA between color rectangle and word or Xs was

manipulated. The word or Xs were presented 300, 200, or 100 ms before the onset of the color rectangle (henceforth referred to as distractor preexposure SOAs, indicated by a minus sign, e.g., –300 ms), simultaneously with the color rectangle (i.e., zero SOA), or 100, 200, or 300 ms after the color rectangle was presented (henceforth referred to as distractor postexposure SOAs, e.g., 300 ms). Whereas the response-exclusion account predicts that Stroop interference should be present in the color-naming RTs but not in the gaze shift latencies and the manual RTs to the arrows, the competition account predicts that Stroop interference should be present in all three measures. Moreover, the response-exclusion account predicts that Stroop interference in the naming RTs should be present at the short distractor SOAs (i.e., –100, 0, and 100 ms) but not at the long distractor preexposure SOAs (i.e., –300 and –200 ms). In contrast, according to the competition account, Stroop interference in the naming RTs may be present at all distractor SOAs.

Method

Participants. The experiment was carried out with a group of 21 participants, who were young adult students at Radboud University Nijmegen. All participants were native speakers of Dutch. They were paid five euros for their participation.

Materials and design. The stimuli consisted of the Dutch color words *rood* (red), *groen* (green), and *blauw* (blue) and corresponding color rectangles. The rectangles were 3.0 cm high and 6.0 cm wide. On average, the rectangles subtended 3.4° vertically and 8.0° horizontally at a viewing distance of 50 cm. The distractor words were presented in 36-point lowercase Arial font. In addition, a row of five Xs served as distractor stimulus in the neutral condition. On average, the words and Xs subtended 2.3° vertically and 6.3° horizontally at a viewing distance of 50 cm. The color words and Xs were superimposed in the center of the color rectangles. The arrow stimuli consisted of < or > flanked by two Xs on each side (yielding XX<XX and XX>XX as stimuli). The Xs were used to demand that the arrows were foveated and to minimize the chance that participants could identify the direction of the arrows by their peripheral vision (cf. Lamers & Roelofs, 2011a; Roelofs, 2007, 2008a, 2008b, 2011). The arrow stimuli XX<XX and XX>XX were presented in 28-point uppercase Arial font, subtending 1.1° vertically and 4.6° horizontally. The horizontal distance between the middle of the color-word stimuli and the arrow stimuli was 24°. The stimuli were presented on a black background. The color of the words, Xs, and arrows was white.

The experiment included incongruent and congruent combinations of colors and words. The incongruent stimuli served to test the predictions about the locus of Stroop interference, whereas the congruent stimuli were included to boost the magnitude of the interference effects. Stroop interference is often larger when congruent stimuli are present in the experiment (cf. Lamers & Roelofs, 2011b). Congruent trials were not analyzed. With three colors, three words, and Xs, there are six possible color–distractor combinations in the incongruent condition but only three in the neutral condition. In order to have an equal number of stimuli in each of the distractor conditions, incongruent trials were therefore constructed by repeatedly pairing one color word with one color rectangle (i.e., *rood*–blue, *groen*–red, *blauw*–green). Thus, there were three incongruent pairings (*rood*–blue, *groen*–red, *blauw*–

green) and three neutral pairings (XXXXX on a red, green, or blue color rectangle). Roelofs (2010a) observed that the yoking of colors with words does not affect the time course of Stroop effects compared with fully crossing colors and words.

There were two experimental factors. The first independent variable was *distractor type* (incongruent, neutral). The second independent variable was *distractor SOA* with seven levels: -300, -200, -100, 0, 100, 200, and 300 ms. Trials were blocked by SOA. The order of presenting the SOA blocks was counterbalanced across participants using a Latin square. There were three dependent variables concerning latencies, henceforth referred to as *measure*: naming response, gaze shift, and manual response. In addition, errors in naming and manual responding were recorded, which are defined below. The order of presenting the stimuli across trials was random, except that repetitions of Stroop stimuli on successive trials were not permitted. Each stimulus was repeated four times per SOA (as in M. O. Glaser & Glaser, 1982; W. R. Glaser & Glaser, 1989).

Apparatus. Materials were presented on a 39-cm ViewSonic 17PS screen. Eye movements were measured using an SMI EyeLink-HiSpeed 2D headband-mounted eye-tracking system (SensoMotoric Instruments GmbH, Teltow, Germany). The eye-tracker was controlled by a Pentium 90 MHz computer. The experiment was run under the Nijmegen Experiment Setup (NESU) with an NESU button box on a Pentium 400 MHz computer. The participants' utterances were recorded over a Sennheiser ME400 microphone to a SONY DTC55 digital audio tape recorder. Vocal response latencies were measured using an electronic voice key.

Procedure. The participants were tested individually. They were seated in front of the computer monitor, a panel with a left and a right push button, and the microphone. The distance between participant and screen was approximately 50 cm. Participants were given written instructions telling them how their eyes would be monitored and what the task was. The experimenter also orally described the eye-tracking equipment and restated the instructions. The participants were told that they had to name the color of rectangles with superimposed words or Xs presented on the left side of a computer screen and manually respond by pressing a left or right button in response to the arrows presented on the right side of the screen. The participants were asked to respond as quickly as possible without making mistakes. Following standard PRP instructions, it was required that the onset of the vocal response precede the onset of the manual response (cf. Meyer & Kieras, 1997a, 1997b).

When a participant had read the instructions, the headband of the eye-tracking system was placed on the participant's head, and the system was calibrated and validated. For pupil-to-gaze calibration, a grid of three by three positions had been defined. During a calibration trial, a fixation target appeared once, in random order, in each of these positions for 1 s. Participants were asked to fixate upon each target until the next target appeared. After the calibration trial, the estimated positions of the participant's fixations and the distances from the fixation targets were displayed to the experimenter. Calibration was considered adequate if there was at least one fixation within 1.5° of each fixation target. When calibration was inadequate, the procedure was repeated, sometimes after adjusting the eye cameras. Successful calibration was followed by a pupil-to-gaze validation trial. For the participants, this

trial did not differ from the calibration trial, but the data collected during the validation trial were used to estimate the participants' gaze positions, and the error (i.e., the distance between the estimated gaze position and the target position) was measured. Validation was considered completed if the average error was below 1.0° and the worst error below 1.5°. Depending on the result of the validation trial, the calibration and validation trials were repeated or testing began.

After successful calibration and validation, a block of 18 practice trials was administered. During the practice block, participants named all stimuli at zero distractor SOA. On each trial, they shifted gaze to the arrow to indicate its direction. The order of distractor types across trials was random. The practice trials were followed by the experimental trials. The structure of a trial was as follows. A trial started by the simultaneous presentation of the color and arrow stimuli, with the words or Xs presented at a certain SOA relative to color onset. The stimuli remained on the screen until the participant pushed one of the buttons in response to the arrow. The latencies of the naming and manual responses were measured from color and arrow presentation onset. Before the start of the next trial, there was a blank interval of 1.5 s. The position of the left and right eyes was determined every 4 ms. Drift correction occurred automatically after every eight trials.

Analyses. To determine the speakers' gaze shift latencies, their eye fixations were classified as falling within or on the outer contours of the Stroop stimulus or elsewhere. Although viewing was binocular and the positions of both eyes were tracked, only the position of the right eye was analyzed. The gaze shift latency was defined as the time interval between the onset of the Stroop stimulus and the end of the last eye fixation at the stimulus before a saccade to the arrow was initiated. Because the Stroop stimuli were always presented in the same position on the screen, there was no fixation point to indicate the position of the stimuli before trial onset. At the beginning of a trial, participants were virtually always fixating the position where the Stroop stimulus would come up. Gaze shifts latencies were measured from the onset of the Stroop stimulus.

A naming response was considered to be invalid when it included a speech error, when a wrong word was produced, or when the voice key was triggered incorrectly. A manual response was invalid when the wrong button was pressed. Invalid trials were discarded from the analyses of the naming latencies, gaze durations, and manual responses. The vocal response latencies, gaze shift latencies, and manual response latencies were submitted to repeated-measures analyses of variance with distractor type and distractor SOA as experimental factors. To correct for the difference in absolute latencies among measures, the comparisons of the magnitude of the factor effects were performed on standard scores (z) with zero mean and unit standard deviation (cf. Roelofs, 2007). Standard scores are commonly used to obtain comparability of observations obtained by different behavioral measures (Winer, Brown, & Michels, 1991). The errors in vocal and manual responding were submitted to logistic regression analyses (Jaeger, 2008). For all tests, an alpha level of .05 was adopted.

Results and Discussion

Figure 3 displays for each distractor type and distractor SOA the mean latencies for the vocal responses, gaze shifts, and manual

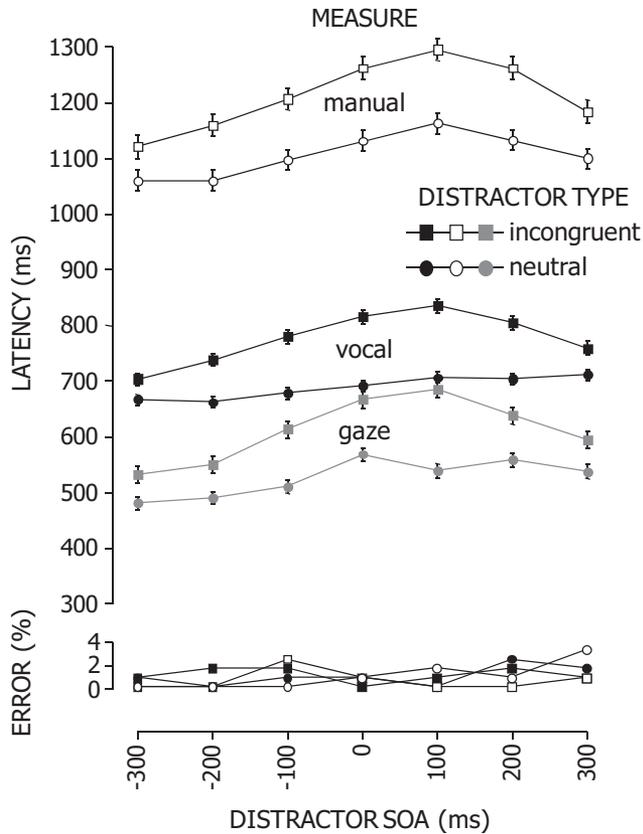


Figure 3. Mean latencies and error percentages for vocal responding, gaze shifting, and manual responding per distractor SOA and distractor type in Experiment 1. The error bars indicate one standard error. SOA = stimulus onset asynchrony.

responses. The figure shows that gaze shifts were initiated before articulation onset: the eye-voice span. Stroop interference effects were obtained in all three behavioral measures. The latencies were longer in the incongruent than the neutral condition at all distractor SOAs. The figure also indicates that the number of errors was very low (less than 1%, on average) and did not differ much among distractor types and distractor SOAs. The statistical analysis of the errors in vocal and manual responding yielded no significant results.

The onset of gaze shifting was 164 ms earlier than the onset of vocal responding, $F(1, 20) = 58.54, p < .001, \eta_p^2 = .75$. Analysis of the eye-voice span revealed no effect of distractor type, $F(1, 20) < 1, p = .68, \eta_p^2 = .01$; distractor SOA, $F(6, 120) = 1.16, p = .33, \eta_p^2 = .06$; or interaction between distractor type and distractor SOA, $F(6, 120) = 1.28, p = .27, \eta_p^2 = .06$. Thus, in all distractor type and distractor SOA conditions, gaze shifts were initiated earlier (i.e., 164 ms before articulation onset) than the articulatory buffer was reached (i.e., estimated to start no earlier than 145 ms before articulation onset).

The statistical analysis of the z scores comparing the latency effects among the naming responses, gaze shifts, and manual responses yielded effects of distractor type, $F(1, 20) = 66.79, p < .001, \eta_p^2 = .77$; distractor SOA, $F(6, 120) = 10.42, p < .001, \eta_p^2 = .34$; and an interaction of distractor type and distractor SOA, $F(6,$

$120) = 2.41, p = .03, \eta_p^2 = .11$. There was an interaction of measure and distractor type, $F(2, 40) = 7.76, p < .001, \eta_p^2 = .28$, but not of measure and distractor SOA, $F(12, 240) < 1, p = .86, \eta_p^2 = .03$. There was also an interaction of measure, distractor type, and distractor SOA, $F(12, 240) = 1.88, p = .04, \eta_p^2 = .09$.

To compare the magnitude of factor effects between the vocal responses and gaze shifts, statistical analyses were performed on the z scores restricted to these two measures. This yielded effects of distractor type, $F(1, 20) = 62.34, p < .001, \eta_p^2 = .76$; distractor SOA, $F(6, 120) = 9.64, p < .001, \eta_p^2 = .33$; and an interaction of distractor type and distractor SOA, $F(6, 120) = 2.89, p = .01, \eta_p^2 = .13$. There was an interaction of measure and distractor type, $F(1, 20) = 8.55, p = .008, \eta_p^2 = .30$, indicating that the magnitude of Stroop interference was somewhat smaller in the gaze shifts than vocal responses (i.e., 85 ms. vs. 89 ms across distractor SOAs; 98 ms. vs. 125 ms at zero SOA, which was used by Roelofs, 2011). There was no interaction of measure and distractor SOA, $F(6, 120) < 1, p = .79, \eta_p^2 = .03$, and also none of measure, distractor type, and distractor SOA, $F(6, 120) = 1.75, p = .12, \eta_p^2 = .08$. The smaller magnitude of the Stroop interference in the gaze shifts than vocal responses replicates Roelofs (2011).

To compare the magnitude of the factor effects between the gaze shifts and manual responses, statistical analyses were performed on the z scores restricted to these two measures. This yielded effects of distractor type, $F(1, 20) = 59.60, p < .001, \eta_p^2 = .75$, and distractor SOA, $F(6, 120) = 9.37, p < .001, \eta_p^2 = .32$, but not an interaction of distractor type and distractor SOA, $F(6, 120) = 1.62, p = .15, \eta_p^2 = .08$. There was also no interaction of measure and distractor type, $F(1, 20) = 1.47, p = .24, \eta_p^2 = .07$; measure and distractor SOA, $F(6, 120) < 1, p = .51, \eta_p^2 = .04$; or measure, distractor type, and distractor SOA, $F(6, 120) = 1.65, p = .14, \eta_p^2 = .08$. Thus, the magnitude of the factor effects did not differ between the gaze shifts and manual responses. The Stroop interference in the gaze shift latencies is fully reflected in the manual responses.

After back-transformation from the standard-score units, the statistical analysis of the vocal naming RTs yielded effects of distractor type, $F(1, 20) = 63.45, p < .001, \eta_p^2 = .76$; distractor SOA, $F(6, 120) = 6.98, p < .001, \eta_p^2 = .26$; and an interaction of distractor type and distractor SOA, $F(6, 120) = 3.41, p = .004, \eta_p^2 = .15$. Pairwise comparisons revealed that the Stroop interference was significant at all distractor SOAs except SOA = 300 ms. Similarly, the analysis of the gaze shift latencies yielded effects of distractor type, $F(1, 20) = 50.51, p < .001, \eta_p^2 = .72$, and distractor SOA, $F(6, 120) = 7.74, p < .001, \eta_p^2 = .28$, and a marginally significant interaction of distractor type and distractor SOA, $F(6, 120) = 2.07, p = .06, \eta_p^2 = .09$. Pairwise comparisons revealed that the Stroop interference was significant at all distractor SOAs except SOA = 300 ms. Finally, the analysis of the manual RTs yielded effects of distractor type, $F(1, 20) = 58.35, p < .001, \eta_p^2 = .75$, and distractor SOA, $F(6, 120) = 7.35, p < .001, \eta_p^2 = .27$, but no interaction of distractor type and distractor SOA, $F(6, 120) < 1, p = .43, \eta_p^2 = .05$.

At distractor preexposure SOAs of 200 or 300 ms, the distractor word may still be in the articulatory buffer when the color name arrives on some of the trials due to stochastic variability of color- and word-processing latencies. That is, if color processing is very fast and word processing is very slow on a particular trial, the word may still be in the buffer at color-name arrival and cause Stroop

interference. This entails that Stroop interference should be present in the fast color-naming responses but should be absent in the slow responses at the long distractor preexposure SOAs. This possibility was evaluated by examining the latency distributions for the incongruent and neutral conditions at the long distractor preexposure SOAs and, for comparison, at the long distractor postexposure SOAs. For these latter SOAs, the response-exclusion hypothesis predicts Stroop interference in the naming latencies but not in the gaze shift latencies.

To obtain the naming latency distributions, the rank-ordered latencies for each participant were divided into tertiles (i.e., three bins, each containing a third of the latencies), and mean naming latencies were computed for each tertile, separately for the latencies in the incongruent and neutral conditions for each of the long distractor preexposure SOAs and postexposure SOAs (i.e., -300 , -200 , 200 , and 300 ms). By averaging these means across participants, Vincentized cumulative distribution curves were obtained (Ratcliff, 1979). Vincentizing the latencies across individual participants provides a way of averaging data while preserving the shapes of the individual distributions. In a similar fashion, the latency distributions for the gaze shifts were obtained. Figure 4 shows the distributional plots.

Figure 4 shows that Stroop interference tended to be present throughout the entire latency range at the long distractor preexposure and postexposure SOAs for both the vocal responses and the gaze shifts. Statistical analysis revealed that for the vocal responses, the Stroop interference did not vary with tertile except for distractor SOA = 200 ms, where the interference increased with tertile. Similarly, for the gaze shifts, the Stroop interference did not vary with tertile except for distractor SOA = 200 ms, where the interference increased with tertile. These results indicate that the Stroop effects at the long distractor preexposure SOAs are not due

to stochastic variability of color- and word-processing latencies, although the effect at the long distractor postexposure SOA of 200 ms may be due to such variability (given that interference increases with tertile).

The results show that Stroop interference is present in all three behavioral measures, at both long distractor preexposure SOAs (i.e., -300 and -200 ms) and a long distractor postexposure SOA (i.e., 200 ms). The presence of Stroop interference in all three measures and at long distractor preexposure SOAs agrees with the competition account but not with the response-exclusion account.

Experiment 2

In the second experiment, participants were again presented with color-word Stroop stimuli displayed on the left side of a computer screen and left- or right-pointing arrows displayed on the right side of the screen, as in Experiment 1. However, rather than manipulating the SOA between color and distractor presentation (i.e., the distractor SOA), the SOA between Stroop stimulus and arrow was manipulated (i.e., the task SOA). The color and distractor of the Stroop stimuli were presented simultaneously. Previous research (Roelofs, 2008a) has indicated that gaze shifts are initiated before the onset of articulation at short task SOAs (i.e., 0 – 300 ms) but after the onset of articulation at long task SOAs (i.e., 1 s). Therefore, at long task SOAs, gaze shifts are initiated well after word planning has reached the articulatory buffer. Thus, the response-exclusion hypothesis predicts that Stroop interference should be present in the color-naming RTs but absent in the gaze shift latencies at these long SOAs. For the short task SOAs, the response-exclusion account predicts Stroop interference in the color-naming RTs and facilitation in the gaze shift latencies and the manual RTs. In contrast, the competition account predicts that Stroop interference should be present in all three measures at all

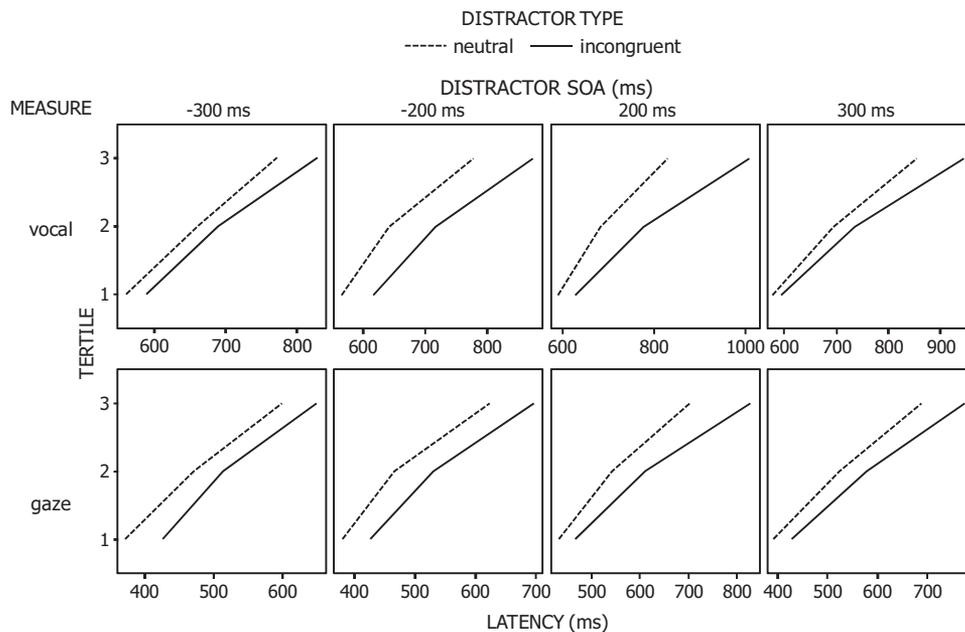


Figure 4. Mean latencies of vocal responding and gaze shifting in the incongruent and neutral conditions shown for three tertiles at long distractor preexposure and postexposure SOAs in Experiment 1. SOA = stimulus onset asynchrony.

task SOAs, except for the manual responses at long task SOAs (as explained earlier).

Method

Participants. The experiment was run with 14 new participants from the same subject population as in the first experiment. Fewer participants were tested in Experiment 2 than in Experiment 1. There is no deep reason for the difference in sample sizes between experiments. Experiment 1 already showed that effects were strong, hence fewer participants were tested for Experiment 2. As can be seen below, analysis of Experiment 2 confirmed that the experiment was powerful enough to detect the relevant effects. Stroop experiments typically do not need many participants to obtain robust effects (e.g., M. O. Glaser & Glaser, 1982; W. R. Glaser & Glaser, 1989). Note that Mahon et al. (2012) tested between the response-exclusion and lexical competition accounts using the color-word Stroop task with only eight participants.

Materials and design. The experiment used displays that were identical to those of Experiment 1 except that the SOA between Stroop stimulus and arrow was manipulated. The arrows were presented at task SOAs of 0, 100, 200, 300, 1000, 1200, and 1400 ms. Trials were blocked by task SOA. On all trials, the color and distractor of the Stroop stimuli were presented simultaneously (i.e., the distractor SOA was 0 ms).

Apparatus, procedure, and analyses. These were the same as in Experiment 1, except that the task SOA was now manipulated. The task SOA had seven levels: 0, 100, 200, 300, 1000, 1200, and 1400 ms.

Results and Discussion

Figure 5 displays for each distractor type and task SOA the mean latencies for the vocal responses, gaze shifts, and manual responses. The figure shows that Stroop interference was obtained in the vocal responses and gaze shifts at all task SOAs, but in the manual responses only at the short task SOAs (i.e., 0, 100, 200, and 300 ms). In line with previous research (Roelofs, 2008a), gaze shifts were initiated, on average, 137 ms before articulation onset at the short task SOAs (i.e., 0, 100, 200, and 300 ms), whereas gaze shifts were initiated, on average, 72 ms after articulation onset at the long task SOAs (i.e., 1000, 1200, and 1400 ms). This effect of task SOA on the moment of gaze shifting did not differ much between distractor types. The figure also indicates that the number of errors was low (less than 2%, on average) and did not differ much among distractor types and task SOAs. The statistical analysis of the errors in vocal and manual responding yielded no significant results.

Statistical analysis confirmed that there was a task SOA effect on the eye-voice span (i.e., the difference in onset between gaze shifting and vocal responding), $F(6, 78) = 25.26, p < .001, \eta_p^2 = .66$. The eye-voice span did not differ between distractor types, $F(1, 13) = 2.99, p = .11, \eta_p^2 = .19$, and there was also no interaction between distractor type and task SOA, $F(6, 78) < 1, p = .95, \eta_p^2 = .02$.

The statistical analysis of the z scores comparing the latency effects among the naming responses, gaze shifts, and manual responses yielded effects of distractor type, $F(1, 13) = 30.98, p < .001, \eta_p^2 = .70$, and task SOA, $F(6, 78) = 6.03, p < .001, \eta_p^2 = .32$,

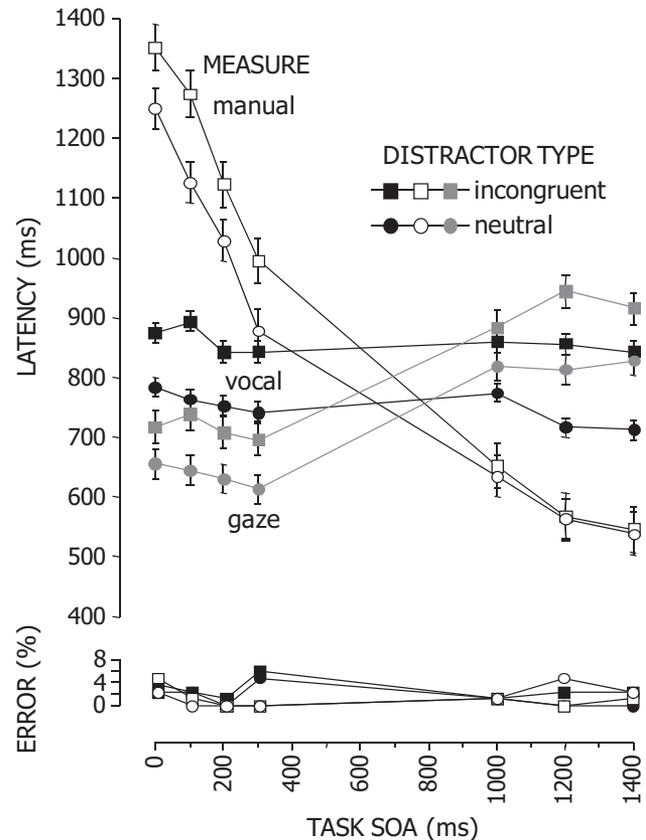


Figure 5. Mean latencies and error percentages for vocal responding, gaze shifting, and manual responding per task SOA and distractor type in Experiment 2. The error bars indicate one standard error. SOA = stimulus onset asynchrony.

but no interaction of distractor type and task SOA, $F(6, 78) < 1, p = .91, \eta_p^2 = .03$. There was an interaction of measure and distractor type, $F(2, 26) = 18.37, p < .001, \eta_p^2 = .59$, and also of measure and task SOA, $F(12, 156) = 70.60, p < .001, \eta_p^2 = .85$. Moreover, measure, distractor type, and task SOA interacted, $F(12, 156) = 1.87, p = .04, \eta_p^2 = .13$.

To compare the magnitude of factor effects between the vocal responses and gaze shifts, statistical analyses were performed on the z scores restricted to these two measures. This yielded effects of distractor type, $F(1, 13) = 28.20, p < .001, \eta_p^2 = .68$, and task SOA, $F(6, 78) = 3.56, p < .004, \eta_p^2 = .22$, but no interaction of distractor type and task SOA, $F(6, 78) < 1, p = .82, \eta_p^2 = .04$. There was an interaction of measure and distractor type, $F(1, 13) = 21.28, p = .001, \eta_p^2 = .62$, indicating that the magnitude of Stroop interference was somewhat smaller in the gaze shifts than vocal responses (i.e., 86 ms vs. 111 ms, respectively). There was also an interaction of measure and task SOA, $F(6, 78) = 19.86, p < .001, \eta_p^2 = .60$, but not of measure, distractor type, and task SOA, $F(6, 78) < 1, p = .93, \eta_p^2 = .02$. The smaller magnitude of the Stroop interference in the gaze shifts than vocal responses replicates Experiment 1 and Roelofs (2011).

To compare the magnitude of the factor effects between the gaze shifts and manual responses, statistical analyses were performed

on the z scores restricted to these two measures. This yielded effects of distractor type, $F(1, 13) = 23.89, p < .001, \eta_p^2 = .65$, and task SOA, $F(6, 78) = 9.78, p < .001, \eta_p^2 = .43$, but not an interaction of distractor type and task SOA, $F(6, 78) < 1, p = .77, \eta_p^2 = .04$. There was a marginally significant interaction of measure and distractor type, $F(1, 13) = 4.22, p = .06, \eta_p^2 = .25$; an interaction of measure and task SOA, $F(6, 78) = 157.42, p < .001, \eta_p^2 = .92$; and an interaction of measure, distractor type, and task SOA, $F(6, 78) = 3.96, p < .002, \eta_p^2 = .23$.

After back-transformation from the standard-score units, the statistical analysis of the vocal naming RTs yielded effects of distractor type, $F(1, 13) = 34.74, p < .001, \eta_p^2 = .73$, but not of task SOA, $F(6, 78) = 1.07, p = .39, \eta_p^2 = .08$. Distractor type and task SOA did not interact, $F(6, 78) < 1, p = .83, \eta_p^2 = .04$. The analysis of the gaze shift latencies yielded effects of distractor type, $F(1, 13) = 15.90, p < .002, \eta_p^2 = .55$, and task SOA, $F(6, 78) = 15.20, p < .001, \eta_p^2 = .54$, but no interaction of distractor type and task SOA, $F(6, 78) < 1, p = .86, \eta_p^2 = .03$. Finally, the analysis of the manual RTs yielded effects of distractor type, $F(1, 13) = 27.72, p < .001, \eta_p^2 = .68$, and task SOA, $F(6, 78) = 83.94, p < .001, \eta_p^2 = .87$, and an interaction of distractor type and task SOA, $F(6, 78) = 2.97, p = .01, \eta_p^2 = .19$.

In the experiment, gaze shifts were initiated before the onset of articulation at short task SOAs but after the onset of articulation at long task SOAs. Therefore, at long task SOAs, gaze shift were initiated after word planning had reached the articulatory buffer. The response-exclusion hypothesis predicts that Stroop interference should be present in the vocal responses but absent in the gaze shift latencies at the long task SOAs. For the short task SOAs, the response-exclusion account predicts Stroop interference in the color-naming RTs and Stroop facilitation in the gaze shift latencies and the manual RTs. In contrast, the competition account predicts that Stroop interference should be present in all three measures. The results of the experiment support the predictions of the competition account.

Note that Experiment 2 is not just a replication of Experiment 1. Although the short task SOAs of Experiment 2 replicate Experiment 1 (even though Experiment 1 had distractor SOAs, not task SOAs), Experiment 2 explicitly shows that Stroop inference is obtained in the gaze shifts regardless of whether they occur *before* or *after* articulation onset.

General Discussion

As outlined previously, whereas it has long been assumed that Stroop interference arises during spoken word planning (e.g., Roelofs, 2003), more recently investigators have argued that the interference arises in an articulatory buffer after the word planning process (Mahon et al., 2012). Above, the results of two color-word Stroop experiments are reported that tested between these accounts using eye-tracking and dual-task procedures with spatially separated task stimuli. Previous research (Roelofs, 2011) indicated that the shifting of eye gaze from one stimulus to another in dual-task performance occurs before the articulatory buffer is reached in spoken word planning. In the present Experiments 1 and 2, participants were presented with color-word Stroop stimuli and left- or right-pointing arrows on different sides of a computer screen. They named the color attribute and shifted their gaze to the arrow to manually indicate its direction. If Stroop interference arises in

the articulatory buffer, the interference should be present in the color-naming latencies but not in the gaze shift and manual response latencies, which should reveal facilitation (cf. Finkbeiner & Caramazza, 2006b; Mahon et al., 2012). Moreover, Stroop interference should be absent in the naming RTs and gaze shift latencies at long distractor preexposure SOAs and absent in the gaze shift latencies at long distractor postexposure SOAs and be present in the naming RTs but absent in the gaze shift latencies at long task SOAs. Contrary to these predictions, Stroop interference was present in all three behavioral measures. Gaze shifts did not reveal facilitation. Moreover, Stroop interference was present at long distractor preexposure and postexposure SOAs, and interference was obtained regardless of whether gaze shifts were initiated before (i.e., at short task SOAs) or after articulation onset (i.e., at long task SOAs). These results indicate that Stroop interference arises during spoken word planning rather than in an articulatory buffer after planning.

In what follows, I first discuss the critical assumption that the articulatory buffer in naming is reached no earlier than about 145 ms before articulation onset. Next, the present eye-tracking findings are compared with prior observations in the literature, and the source of the Stroop conflict (i.e., task or response) is discussed. Finally, a theoretical account of the present data patterns is given, assuming that the locus of Stroop interference is in word planning.

Estimating the Onset of Articulatory Buffering

An important assumption underlying the present study is that the articulatory buffer in naming is reached no earlier than about 145 ms before articulation onset. This assumption was based on estimates by Indefrey and Levelt (2004; Indefrey, 2011). Most of the studies in the meta-analysis of Indefrey and Levelt had a different methodology than the one used here (i.e., the studies relied on covert responding, picture naming, etc.), and hence, it is not necessarily the case that their estimates generalize in a straightforward manner. However, elsewhere (Roelofs, 2003), I have indicated that the estimates of Indefrey and Levelt yield excellent fits to Stroop color-naming and word-reading RTs. So, despite the fact that the estimates of Indefrey and Levelt were based on picture naming and other tasks, they seem to generalize well to the Stroop task.

According to Indefrey and Levelt (2004; Indefrey, 2011), phonetic encoding starts approximately 145 ms before articulation onset, and the articulatory buffer is reached somewhat later. This estimate for the onset of phonetic encoding and reaching the articulatory buffer was based on a total naming RT of 600 ms. In the present experiments, the color-naming RTs were somewhat longer, namely, 688 ms and 762 ms in the neutral condition of Experiments 1 and 2, respectively. Does this have consequences for the conclusions drawn from the experiments reported above?

As extensively discussed by Indefrey (2011), if the RTs in an experiment differ from the mean RT of 600 ms assumed by Indefrey and Levelt (2004; Indefrey, 2011), then there are a number of possibilities for rescaling the estimates for the stage durations. The most straightforward procedure would be a linear rescaling of all stage durations. Applying this procedure to the present study would yield estimates of 166 ms (Experiment 1) and 184 ms (Experiment 2) rather than 145 ms for the onset of phonetic encoding and reaching the articulatory buffer. In Experiment 1,

gaze shifts occurred 164 ms before articulation onset. Thus, gaze shifts were initiated at approximately the estimated onset of phonetic encoding and before the articulatory buffer was reached. Still, Stroop interference was present in both the vocal responses and the gaze shifts, which does not agree with the response-exclusion account. In Experiment 2, gaze shifts occurred 137 ms before articulation onset at the short task SOAs, whereas phonetic encoding was estimated to start 184 ms before articulation. Thus, in the second experiment, gaze shifts happened during phonetic encoding. Still, the magnitude of Stroop interference did not differ between the short task SOAs (when gaze shifts occurred during phonetic encoding, before articulation onset) and long task SOAs (when gaze shifts occurred after phonetic encoding and articulation onset). This suggests that the articulatory buffer is not the locus of the Stroop interference.

Another option for rescaling the estimates for the stage durations would be to rescale only some rather than all of the stage durations, based on assumptions about the details of the experimental situation. As Indefrey (2011) stated, "it can be said that a linear rescaling of the duration of all processing stages can only be the last resort and is inadequate whenever the reason for shorter or longer naming latencies can be identified" (p. 3). For Stroop task performance, it seems plausible to assume that early visual processing is somewhat more difficult for color-word combinations than for colors only (cf. MacLeod, 1998). This would entail that only the duration of early visual lead-in processes has to be rescaled, henceforth lead-in rescaling, rather than the duration of all processing stages (i.e., linear rescaling). Adjusting the duration of earlier stages has no implications for the estimated duration of later stages, including the moment at which the articulatory buffer is reached. That is, the estimate of the onset of articulatory buffering would remain 145 ms before articulation onset, even though the naming RTs in the present experiments were longer than the 600 ms assumed by Indefrey and Levelt (2004; Indefrey, 2011). In Experiment 1, gaze shifts occurred 164 ms before articulation onset, which is before the estimated onset of phonetic encoding and reaching the articulatory buffer of 145 ms before articulation onset. Still, Stroop interference was present in both the vocal responses and gaze shifts, in disagreement with the response-exclusion account. In Experiment 2, gaze shifts occurred 137 ms before articulation onset at the short task SOAs, whereas phonetic encoding is estimated to start 145 ms before articulation. Thus, gaze shifts were initiated almost at the onset of phonetic encoding and before reaching the articulatory buffer. Still, Stroop interference was present in both the vocal responses and gaze shifts. Moreover, the magnitude of Stroop interference did not differ between the short task SOAs (when gaze shifts occurred at approximately the onset of phonetic encoding, before reaching the articulatory buffer) and long task SOAs (when gaze shifts occurred after completion of phonetic encoding and reaching the articulatory buffer). Thus, also with rescaling only some rather than all of the stage durations, the evidence suggests that the articulatory buffer is not the locus of the Stroop interference.

Moreover, there are reasons to assume that the estimate of 145 ms for the onset of articulatory buffering is too large. The response-exclusion account assumes that the output buffer contains "phonologically well-formed responses" (Finkbeiner & Caramazza, 2006b, p. 791), "production-ready representations" (Janssen et al., 2008, p. 250), or "articulatory programs" (Finkbeiner &

Caramazza, 2006a, p. 1033). The assumption that the representations in the output buffer are phonologically well formed implies that the phonological encoding stage (see Figure 1) must have been completed (otherwise, the representations in the output buffer would not have been phonologically well formed). Moreover, the assumption that the representations in the output buffer are production-ready representations or articulatory programs implies that phonetic encoding also must have been completed. This implies that the estimated onset of the articulatory buffering stage of 145 ms before articulation onset is, in fact, an overestimation because, at the onset of phonetic encoding, the articulatory program still has to be constructed. This strengthens the conclusion that the gaze shifts in the present experiments (except those at the long task SOAs in Experiment 2) were initiated before the articulatory buffer was reached in spoken word planning.

To conclude, an important assumption underlying the present study is that the articulatory buffer in naming is reached no earlier than about 145 ms before articulation onset. However, there are good reasons to assume that this estimate is, in fact, an overestimation. The articulatory buffering stage must be reached even closer to articulation onset than assumed in the present study. The mean naming RTs in the present study are longer than assumed for the estimate of 145 ms. However, even after rescaling the overestimated onset of phonetic encoding, the results of the present experiments are not in agreement with the response-exclusion account.

The present experiments rely on a rather complex methodology. Specifically, participants are required to name color rectangles with superimposed incongruent words or Xs while also having to manually indicate the direction of an arrow. The tacit assumption is that the manual response task and the naming task do not interfere. However, it is possible that participants verbally mediate (i.e., by covertly saying "left" or "right") their manual responses to the arrows, which could then interfere with the color-naming process. Consequently, this could affect the timing of the relative stages of processing in the naming task by delaying the point in time at which articulation starts. In short, this may undermine the fundamental assumption that the articulatory buffer is reached no earlier than about 145 ms prior to vocal onset.

However, the Stroop stimuli and arrows were separated by 24° of visual angle. Moreover, the arrows < and > were flanked by two Xs on each side. Pretests for the present experiments and previous work using the same paradigm (Roelofs, 2007, 2008a, 2008b, 2011) revealed that participants cannot discriminate the arrow while fixating the Stroop stimulus. This is also demonstrated by Figure 2, where fixating the Stroop stimulus precludes discrimination of the arrow, even though Stroop stimulus and arrow are much closer to each other than 24°. Still, gaze shifts happened before articulation onset in the present experiments (except at the long task SOAs in Experiment 2), so perhaps covert verbalization of the arrow occurs while the arrow is fixated but the planning of the vocal response is still in progress. However, this scenario does not fit the situation in Experiment 2, where Stroop interference is present in the vocal responses and gaze shift latencies regardless of whether the gaze shifts occurred before or after the vocal responses and regardless of whether the manual responses occurred more than 1 or 2 s after color onset (i.e., at short or long task SOAs). To conclude, it is unlikely that processing of the arrow stimuli in the present experiments interfered with color-naming planning. There

is no good reason to assume that the arrow task undermines the assumption that the articulatory buffer is reached no earlier than about 145 ms before articulation onset.

Comparison With Previous Findings

Whereas the present experiments provide evidence that Stroop interference arises during spoken word planning rather than in an articulatory buffer after word planning, previous eye-tracking studies obtained results that were less clear cut. For example, *Roelofs (2007)* examined gaze shifting using the picture-word interference analogue of the color-word Stroop task. Participants were presented with picture-word combinations, displayed on the left side of a computer screen, and left- or right-pointing arrows, displayed on the right side of the screen. The tasks were to name the picture and to shift gaze to the arrow to indicate its direction by pressing a left or right response key.

Mean latencies for the naming responses, gaze shifts, and manual responses were longer in a semantic condition (e.g., a pictured cat combined with the word *dog*) than an unrelated condition (e.g., a pictured cat combined with the word *house*), longer in the semantic condition than a neutral control condition (e.g., a pictured cat combined with a series of Xs), and longer in the semantic condition than an identity condition (e.g., a pictured cat combined with the word *cat*). The magnitude of the distractor effects did not differ among measures (i.e., vocal, gaze, manual). Gaze shifts occurred about 66 ms before articulation onset, regardless of distractor type. According to *Indefrey and Levelt (2004; Indefrey, 2011)*, the articulatory buffer in naming is reached no earlier than about 145 ms before articulation onset. Given that gaze shifts in picture naming were initiated 66 ms before the onset of articulation, the observation of distractor effects in all three measures is compatible with both the hypothesis of a locus of the effects in lexical selection and the hypothesis of a locus in the articulatory buffer. That is, the data of the picture-word interference experiment do not adjudicate between the competition and response-exclusion hypotheses but are compatible with both accounts.

In the picture-word interference experiment of *Roelofs (2007)*, the number of different responses was larger than in the present Stroop experiments, and the responses were repeated less often. Perhaps for this reason, participants were slightly more conservative in the moment of gaze shifting in the earlier picture-word interference experiment than in the present Stroop experiments. However, this account is speculative. Unless the same group of participants is tested on both the picture-word interference task and the Stroop task, it remains unclear whether the difference in outcome between experiments is due to the tasks or the participant groups (cf. *Piai et al., 2014*). Regardless of the cause of the difference, whereas the data of the picture-word interference experiment do not adjudicate between the competition and response-exclusion hypotheses, the present data from the Stroop task support the competition account.

As described earlier, *Roelofs (2011)* reported two eye-tracking experiments in which participants were presented with color-word Stroop stimuli and left- or right-pointing arrows on different sides of a computer screen, as in the present experiments. Participants named the color attribute and shifted their gaze to the arrow to manually indicate its direction. The results showed that Stroop interference was present in the gaze shift latencies even though the

shifts were initiated about 215 ms before articulation onset. However, the magnitude of Stroop interference was somewhat smaller in the gaze shift latencies than the naming RTs, which suggests that part of the Stroop interference occurred after gaze shift onset. An unusual feature of these experiments was that the color attribute was removed 100 ms after Stroop stimulus onset on half the trials. The removal of the color attribute may have led to earlier gaze shifts than normal, which may explain why distractor type had a smaller effect on the gaze shifts than the vocal responses.

In the present two experiments, the color attribute was presented until trial offset on all trials. In both experiments, the magnitude of the Stroop interference effect was slightly larger in the naming responses than the gaze shifts. This suggests that the color removal in the experiments of *Roelofs (2011)* may not have been the critical factor causing the reduced magnitude of the Stroop interference in the gaze shift latencies. Instead, the evidence suggests that gaze shifts occur before Stroop interference is fully resolved regardless of the color presentation duration. In WEAVER++, Stroop interference arises because of competition among lemmas in lemma retrieval and among articulatory programs in phonetic encoding, with lemma competition generating the largest amount of interference (*Roelofs, 2003*). If gaze shifts are initiated before the onset of phonetic encoding, some of the interference in the vocal responses (i.e., interference arising during phonetic encoding) will not be reflected in the gaze shift latencies. Consequently, the magnitude of Stroop interference will be smaller in the gaze shifts than vocal responses. This corresponds to what has been observed in *Roelofs (2011)* and the present experiments.

To conclude, whereas the results of previous eye-tracking experiments agreed, at least partly, with both the competition and response-exclusion accounts, the results of the present experiments indicate that Stroop interference arises during spoken word planning (i.e., in lexical access) rather than in an articulatory buffer after word planning.

Source of the Stroop Conflict

The number of stimuli is typically much smaller, and hence the stimuli and responses are repeated more often, in the color-word Stroop task than the picture-word interference task. This could lead to a stronger conflict in Stroop than picture-word interference and, more importantly, different sources of conflict. *Monsell, Taylor, and Murphy (2001)* argued that the Stroop task has two sources of competition: task sets (that of naming the word vs. naming the color) and response tendencies (due to breakthrough of strongly activated responses). From this perspective, it could be argued that in the picture-word interference task, the source of conflict is mainly in the task set. This would imply that the sources of Stroop and picture-word interferences are different, contrary to what is assumed in the debate about the locus of interference (i.e., word planning vs. articulatory buffering).

According to *Monsell et al. (2001)*, task set competition is triggered by wordlike constituents in a letter string. However, in the picture-word interference task, semantically related and unrelated distractors are all words, which are, moreover, often exactly the same in the semantically related and unrelated conditions (i.e., when these conditions are created by re-pairing pictures and words). To explain the semantic interference effect at the level of task set competition, one has to assume that semantically related

words cause more competition at the level of the task set than unrelated words, which is not what the data of Monsell et al. suggest. Moreover, attributing semantic interference to task set competition is not supported by other evidence in the literature. Several studies in the literature (e.g., Aarts, Roelofs, & Van Turenout, 2009; Roelofs, 2012; Steinhauser & Hübner, 2009) suggest that effects of task competition are present in the tail of an RT distribution only (i.e., manifested as distributional skewing). However, Piai et al. (2011) observed that semantic interference in the picture-word interference task is present in the normal part (i.e., manifested as distributional shifting) of the RT distribution but not in the tail, suggesting that the effect is due to response competition rather than task set competition. Moreover, the RT distributional analyses of Stroop color naming in the present Experiments 1 and 2 showed that the Stroop interference was present throughout the RT distribution (i.e., manifested as distributional shifting), which also suggests that the effect is mainly due to response competition. To conclude, existing evidence in the literature and the present findings agree with the assumption that the source of conflict is the same in picture-word and color-word interference.

Accounting for the Present Data Patterns

The results of the present eye-tracking experiments on color-word Stroop interference suggest that the effect arises during spoken word planning (i.e., in lexical access) rather than in an articulatory buffer after planning. To account for performance in the dual-task situation with spatially separated task stimuli (see Figure 2), the WEAVER++ model (Roelofs, 2007, 2008a) assumes that to maintain acceptable levels of speed and accuracy, to minimize resource consumption and crosstalk between tasks, and to satisfy instructions about task priorities, participants set a criterion for when the gaze shift between the stimuli for the vocal and manual tasks should occur (cf. Meyer & Kieras, 1997a, 1997b). In the present experiments, gaze shifts from Stroop stimulus to arrow were initiated during phonological encoding (cf. Figure 1), which corresponds to what has been observed in previous experiments (Roelofs, 2008a, 2008b). In Experiment 1, distractors were presented at preexposure and postexposure SOAs. Stroop interference was present in all three behavioral measures, at long preexposure as well as postexposure SOAs. This is to be expected if the initiation of gaze shifts depends on the completion of critical aspects of word planning. In the model, Stroop interference arises predominantly during lemma retrieval, whereas competition in selecting articulatory programs during phonetic encoding only makes a small contribution (Roelofs, 2003). If gaze shifts are initiated during phonological encoding, interference arising during phonetic encoding will not be reflected in the gaze shift latencies, which explains why the magnitude of Stroop interference was somewhat larger in the vocal responses than the gaze shifts in the present Experiments 1 and 2.

In Experiment 2, the Stroop and arrow stimuli were presented both with short and long task SOAs. At long task SOAs of 1 s or more, the Stroop stimulus is presented at trial onset (on the left side of a computer screen), but the spatial position of the arrow remains empty until the arrow is presented (on the right side of the screen). To avoid long eye fixations of an empty position in space, gaze shifts may be delayed by a constant amount of time after meeting

the shift criterion (cf. Roelofs, 2008a). As a consequence, Stroop interference will be reflected in the vocal responses and gaze shift latencies at all task SOAs but in the manual RTs only at short SOAs. At the long task SOAs (i.e., 1000, 1200, and 1400 ms), the eyes will fixate the spatial position of the arrow 100 ms or more before the arrow appears on both incongruent and neutral trials, which will absorb the Stroop interference (cf. Roelofs, 2008a). This corresponds to the empirical findings of Experiment 2.

To conclude, Stroop interference was present in the vocal responses, gaze shifts, and manual responses at both distractor preexposure and postexposure SOAs. The magnitude of the interference was somewhat larger in the vocal responses than the gaze shifts. Under the competition account, this may be explained by assuming that participants adopt a shift criterion that is met during phonological encoding. Moreover, Stroop interference was present in the vocal responses, gaze shifts, and manual responses at short task SOAs but only in the vocal responses and gaze shifts at long task SOAs. This may be explained by assuming that gaze shifts are postponed by a constant amount of time after meeting the shift criterion at long task SOAs but are completed before the arrow is presented.

Conclusions

In dual-task performance with spatially separated task stimuli, Stroop interference was present in the naming RTs, gaze shift latencies, and manual RTs. Moreover, Stroop interference was present at long distractor preexposure and postexposure SOAs, and interference was obtained regardless of whether gaze shifts were initiated before (i.e., at short task SOAs) or after articulation onset (i.e., at long task SOAs). These results provide evidence that Stroop interference arises during spoken word planning rather than later in an articulatory buffer.

References

- Aarts, E., Roelofs, A., & Van Turenout, M. (2009). Attentional control of task and response in lateral and medial frontal cortex: Brain activity and reaction time distributions. *Neuropsychologia*, *47*, 2089–2099. doi:10.1016/j.neuropsychologia.2009.03.019
- Abdel Rahman, R., & Aristei, S. (2010). Now you see it . . . and now again: Semantic interference reflects lexical competition with and without articulation. *Psychonomic Bulletin & Review*, *17*, 657–661. doi:10.3758/PBR.17.5.657
- Augustinova, M., & Ferrand, L. (2012). Suggestion does not de-automatize word reading: Evidence from the semantically based Stroop task. *Psychonomic Bulletin & Review*, *19*, 521–527. doi:10.3758/s13423-012-0217-y
- Buswell, G. T. (1920). *An experimental study of the eye-voice span in reading*. Chicago, IL: University of Chicago.
- 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*, 456–463. doi:10.3758/BF03210835
- Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review*, *14*, 717–722. doi:10.3758/BF03196827
- Dennis, I., & Newstead, S. E. (1981). Is phonological recoding under strategic control? *Memory & Cognition*, *9*, 472–477. doi:10.3758/BF03202341
- Dhooge, E., De Baene, W., & Hartsuiker, R. J. (2013). A late locus of the distractor frequency effect in picture-word interference: Evidence from

- event-related potentials. *Brain and Language*, 124, 232–237. doi:10.1016/j.bandl.2012.12.005
- Dhooze, E., & Hartsuiker, R. J. (2010). The distractor frequency effect in picture-word interference: Evidence for response exclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 878–891. doi:10.1037/a0019128
- Dhooze, E., & Hartsuiker, R. J. (2011). The distractor frequency effect in a delayed picture-word interference task: Further evidence for a late locus of distractor exclusion. *Psychonomic Bulletin & Review*, 18, 116–122. doi:10.3758/s13423-010-0026-0
- Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, 1, 106–120. doi:10.3758/BF03198078
- Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Implications for the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 1058–1079. doi:10.1037/0096-1523.18.4.1058
- Finkbeiner, M., & Caramazza, A. (2006a). Lexical selection is not a competitive process: A reply to La Heij et al. (2006). *Cortex*, 42, 1031–1036. doi:10.1016/S0010-9452(08)70210-7
- Finkbeiner, M., & Caramazza, A. (2006b). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, 42, 790–796. doi:10.1016/S0010-9452(08)70419-2
- Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 875–894. doi:10.1037/0096-1523.8.6.875
- Glaser, W. R., & Dünghoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 640–654. doi:10.1037/0096-1523.10.5.640
- Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, 118, 13–42. doi:10.1037/0096-3445.118.1.13
- Hutson, J., Damian, M. F., & Spalek, K. (2013). Distractor frequency effects in picture-word interference task with vocal and manual responses. *Language and Cognitive Processes*, 28, 615–632. doi:10.1080/01690965.2011.605599
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, 2, Article 255. doi:10.3389/fpsyg.2011.00255
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101–144. doi:10.1016/j.cognition.2002.06.001
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446. doi:10.1016/j.jml.2007.11.007
- Janssen, N. (2013). Response exclusion in word-word tasks: A comment on Roelofs, Piai, and Schriefers. *Language and Cognitive Processes*, 28, 672–678. doi:10.1080/01690965.2012.746715
- Janssen, N., Schirm, W., Mahon, B. Z., & Caramazza, A. (2008). Semantic interference in a delayed naming task: Evidence for the response exclusion hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 249–256. doi:10.1037/0278-7393.34.1.249
- Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, 77, 576–588. doi:10.2307/1420768
- La Heij, W., Kuipers, J., & Starreveld, P. (2006). In defense of the lexical-competition account of picture-word interference: A comment on Finkbeiner and Caramazza (2006). *Cortex*, 42, 1028–1031. doi:10.1016/S0010-9452(08)70209-0
- Lamers, M. J. M., & Roelofs, A. (2011a). Attention and gaze shifting in dual-task and go/no-go performance with vocal responding. *Acta Psychologica*, 137, 261–268. doi:10.1016/j.actpsy.2010.12.005
- Lamers, M. J. M., & Roelofs, A. (2011b). Attentional control adjustments in Stroop and Eriksen task performance can be independent of response conflict. *Quarterly Journal of Experimental Psychology*, 64, 1056–1081. doi:10.1080/17470218.2010.523792
- Lamers, M. J. M., Roelofs, A., & Rabeling-Keus, I. M. (2010). Selective attention and response set in the Stroop task. *Memory & Cognition*, 38, 893–904. doi:10.3758/MC.38.7.893
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38. doi:10.1017/S0140525X99001776
- Levin, H., & Buckler-Addis, A. (1979). *The eye-voice span*. Cambridge, MA: MIT Press.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163–203. doi:10.1037/0033-2909.109.2.163
- MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition*, 26, 201–211. doi:10.3758/BF03201133
- Mädebach, A., Oppermann, F., Hantsch, A., Curda, C., & Jescheniak, J. D. (2011). Is there semantic interference in delayed naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 522–538. doi:10.1037/a0021970
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 503–535. doi:10.1037/0278-7393.33.3.503
- Mahon, B. Z., Garcea, F. E., & Navarrete, E. (2012). Picture-word interference and the response-exclusion hypothesis: A response to Mulatti and Coltheart. *Cortex*, 48, 373–377. doi:10.1016/j.cortex.2011.10.008
- Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, 104, 3–65. doi:10.1037/0033-295X.104.1.3
- Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review*, 104, 749–791. doi:10.1037/0033-295X.104.4.749
- Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26, 3–67.
- Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General*, 132, 228–252. doi:10.1037/0096-3445.132.2.228
- Monsell, S., Taylor, T. J., & Murphy, K. (2001). Naming the color of a word: Is it responses or task sets that compete? *Memory & Cognition*, 29, 137–151. doi:10.3758/BF03195748
- Morton, J., & Chambers, S. M. (1973). Selective attention to words and colours. *Quarterly Journal of Experimental Psychology*, 25, 387–397. doi:10.1080/14640747308400360
- Mulatti, C., & Coltheart, M. (2012). Picture-word interference and the response-exclusion hypothesis. *Cortex*, 48, 363–372. doi:10.1016/j.cortex.2011.04.025
- Mulatti, C., & Coltheart, M. (2014). Color naming of colored non-color words and the response-exclusion hypothesis: A comment on Mahon et al. and on Roelofs and Piai. *Cortex*, 52, 120–122. doi:10.1016/j.cortex.2013.08.018
- Olk, B. (2013). Measuring the allocation of attention in the Stroop task: Evidence from eye movement patterns. *Psychological Research*, 77, 106–115. doi:10.1007/s00426-011-0405-9
- Piai, V., Roelofs, A., & Schriefers, H. (2011). Semantic interference in immediate and delayed naming and reading: Attention and task decisions. *Journal of Memory and Language*, 64, 404–423. doi:10.1016/j.jml.2011.01.004

- Piai, V., Roelofs, A., & Schriefers, H. (2012). Distractor strength and selective attention in picture naming performance. *Memory & Cognition*, *40*, 614–627. doi:10.3758/s13421-011-0171-3
- Piai, V., Roelofs, A., & Schriefers, H. (2014). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 147–165. doi:10.1037/a0033745
- Piai, V., Roelofs, A., & Van der Meij, R. (2012). Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. *Brain Research*, *1450*, 87–101. doi:10.1016/j.brainres.2012.02.050
- Quantz, J. O. (1897). Problems in the psychology of reading. *Psychological Review: Monograph Supplements*, *2*, 1–51.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446–461. doi:10.1037/0033-2909.86.3.446
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. doi:10.1037/0033-2909.124.3.372
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157. doi:10.1037/0033-295X.105.1.125
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142. doi:10.1016/0010-0277(92)90041-F
- Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249–284. doi:10.1016/S0010-0277(97)00027-9
- Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88–125. doi:10.1037/0033-295X.110.1.88
- Roelofs, A. (2007). Attention and gaze control in picture naming, word reading, and word categorizing. *Journal of Memory and Language*, *57*, 232–251. doi:10.1016/j.jml.2006.10.001
- Roelofs, A. (2008a). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1580–1598. doi:10.1037/a0012476
- Roelofs, A. (2008b). Dynamics of the attentional control of word retrieval: Analyses of response time distributions. *Journal of Experimental Psychology: General*, *137*, 303–323. doi:10.1037/0096-3445.137.2.303
- Roelofs, A. (2008c). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 353–368. doi:10.1037/0278-7393.34.2.353
- Roelofs, A. (2010a). Attention and facilitation: Converging information versus inadvertent reading in Stroop task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 411–422. doi:10.1037/a0018523
- Roelofs, A. (2010b). Attention, temporal predictability, and the time course of context effects in naming performance. *Acta Psychologica*, *133*, 146–153. doi:10.1016/j.actpsy.2009.11.003
- Roelofs, A. (2011). Attention, exposure duration, and gaze shifting in naming performance. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 860–873. doi:10.1037/a0021929
- Roelofs, A. (2012). Attention, spatial integration, and the tail of response time distributions in Stroop task performance. *Quarterly Journal of Experimental Psychology*, *65*, 135–150. doi:10.1080/17470218.2011.605152
- Roelofs, A., & Piai, V. (2013). Associative facilitation in the Stroop task: Comment on Mahon et al. *Cortex*, *49*, 1767–1769. doi:10.1016/j.cortex.2013.03.001
- Roelofs, A., Piai, V., & Schriefers, H. (2011). Selective attention and distractor frequency in naming performance: Comment on Dhooge and Hartsuiker (2010). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1032–1038. doi:10.1037/a0023328
- Roelofs, A., Piai, V., & Schriefers, H. (2013a). Context effects and selective attention in picture naming and word reading: Competition versus response exclusion. *Language and Cognitive Processes*, *28*, 655–671. doi:10.1080/01690965.2011.615663
- Roelofs, A., Piai, V., & Schriefers, H. (2013b). Selection by competition in word production: Rejoinder to Janssen (2013). *Language and Cognitive Processes*, *28*, 679–683. doi:10.1080/01690965.2013.770890
- Schnur, T. T., & Martin, R. (2012). Semantic picture–word interference is a post-perceptual effect. *Psychonomic Bulletin & Review*, *19*, 301–308.
- Seymour, P. H. K. (1977). Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology*, *29*, 245–265. doi:10.1080/14640747708400601
- Steinhauser, M., & Hübner, R. (2009). Distinguishing response conflict and task conflict in the Stroop task: Evidence from ex-Gaussian distribution analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1398–1412. doi:10.1037/a0016467
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662. doi:10.1037/h0054651
- Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design*. New York, NY: McGraw-Hill.
- Woodworth, R. S. (1938). *Experimental psychology*. New York, NY: Holt.
- Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology*. New York, NY: Holt.

Received March 27, 2013

Revision received February 13, 2014

Accepted February 17, 2014 ■