

Published in final edited form as:

J Exp Psychol Learn Mem Cogn. 2017 March 1; 43(3): 369–384. doi:10.1037/xlm0000311.

The Magic of Words Reconsidered: Investigating the Automaticity of Reading Color-Neutral Words in the Stroop Task

Sachiko Kinoshita,

Department of Psychology and ARC Centre of Excellence in Cognition and its Disorders,
Macquarie University

Bianca De Wit, and

Department of Cognitive Science, Macquarie University

Dennis Norris

MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom

Abstract

In 2 variants of the color-word Stroop task, we compared 5 types of color-neutral distractors—real words (e.g., *HAT*), pseudowords (e.g., *HIX*), consonant strings (e.g., *HDK*), symbol strings (e.g., # \$%), and a row of Xs (e.g., *XXX*)—as well as incongruent color words (e.g., *GREEN* displayed in red). When participants named the color, relative to a row of Xs, words and pseudowords interfered equally and more than the consonant strings, which in turn interfered more than the symbols. In contrast, when participants identified the color by manual key-press responses, all 5 types of neutral strings produced equal color response latencies. In both tasks, the incongruent color words produced robust interference relative to the color-neutral words. Reaction time (RT) distribution analyses showed that all interference effects (relative to the row of Xs) increased across the quantiles. We interpret these results in terms of an evidence accumulation process in which the interfering distractor reduces the effective rate of evidence accumulation for the color target. We take the results to argue that the task of reading, even when triggered unintentionally, is not an invariant process driven solely by the stimulus properties, and is instead guided by the task goal.

Keywords

Stroop; reading; RT distribution; automaticity; vocal versus manual Stroop tasks

“Words are magic. They are magic in the sense that when presented with a word or a wordlike letter string, a literate subject cannot help but read” (Prinzmetal, Hoffman, & Vest, 1991, p. 902). Perhaps the best known demonstration of the magic of words—the automaticity of reading—is the Stroop task (Stroop, 1935): When asked to name the color in which a letter string is displayed, it is much harder if that letter string is the name of a conflicting color word (e.g., *GREEN* displayed in red) than if it is a control string (e.g.,

XXXXX displayed in red). This Stroop interference effect is one of the most robust and well-studied cognitive phenomena, and is widely regarded as the gold standard in demonstrating the automaticity of reading (see review by MacLeod, 1991). The Stroop task would thus be a useful tool for investigating the processes underlying skilled reading, but surprisingly little is known about the visual word recognition processes that produce interference in a Stroop task. Specifically, from the perspective of visual word recognition, “reading” could mean reading aloud—the process of generating a speech response from the printed letter string (which may involve the application of grapheme-phoneme mappings or retrieval of stored phonology from lexical memory)—or retrieving the meaning of the word from lexical memory, or the process of lexical access—finding an item in the mental lexicon that matches the visual input. Which of these processes are engaged automatically despite the reader’s intention to ignore the word, and how do they interfere with responding to color?

It is important to note at the outset that our aim here is not to debate whether or not reading is automatic, according to any of the many-faceted criteria of automaticity (see Besner, 2001; Moors & De Houwer, 2006; Neely & Kahan, 2001). In particular, previous studies have shown that manipulations such as presenting the word distractor in a separate spatial location from the to-be-named color patch (e.g., Kahneman & Henik, 1981; Spieler, Balota, & Faust, 2000), or presenting just one letter within a word distractor in color (e.g., Augustinova & Ferrand, 2014; Besner, 2001; Besner, Stolz, & Boutilier, 1997; Labuschagne & Besner, 2015; Risko, Stolz, & Besner, 2005; Robidoux, Rauwerda, & Besner, 2014), reduces, or even eliminates, the Stroop interference effect, indicating that the effect is dependent on spatial attention. It has also been shown that masked priming effects—another set of phenomena widely assumed to be automatic—disappear when the masked prime is presented in an unattended spatial location (Lachter, Forster, & Ruthruff, 2004; Lien, Ruthruff, Kouchi, & Lachter, 2010). Thus, we take it as given that reading requires spatial attention. What is currently not known, however, is which aspect of reading a word, presented in full view as in normal reading, is triggered against the intention not to read the word and causes interference in a Stroop task, and this is what we investigate in our study.

The starting point of our research is the “wordlikeness gradient” found with non-color-associated stimuli: More wordlike stimuli produce more interference in a Stroop task (Klein, 1964; Monsell, Taylor, & Murphy, 2001). Monsell et al. (2001) took the finding to suggest that “the detection of wordlike properties of the stimulus evokes exogenously in literate subjects the associated task set of reading” (p. 147). We build on this notion in three ways. First, we will investigate which properties of wordlikeness are critical. The results indicate that it is the sublexical properties related to the pronounceability of the letter string, but not the lexicality, reinforcing Monsell et al.’s conclusion that it is not the involuntary lexical access that causes the interference. Second, previous studies investigating the effects of wordlike properties have used only the vocal Stroop task in which the target color name must be spoken. Here, we will contrast the manual Stroop task in which the colors are identified by means of a key press with the vocal Stroop task. We will show that the patterns of interference observed are qualitatively different in the two Stroop tasks, with the wordlikeness gradient being completely absent in the manual task. However, the incongruent color words produce substantial interference in both tasks, indicating that reading is not

absent in the manual task, and we interpret the difference in terms of the task goal of generating a speech response in the vocal, but not in the manual, task. Third, we examine the response time (RT) distributions. Although the vast majority of Stroop studies report only the mean RTs and error rates, analysis of RT distributions can provide greater insights into the dynamics of the underlying processes. Based on the pattern of RT distributions, we will argue that all of the interference effects observed here reflect task conflict, rather than a late-occurring response conflict. Taken together, we argue that the pattern of Stroop interference effects indicates that the task of reading, even when triggered automatically, is governed by the goal of the task required to the target.

Wordlikeness Gradient in Stroop Color Naming

Klein (1964) was the first to investigate the properties of distractor words with no color associations, and reported that the interference was greater in the following descending order: common words, rare words, consonant strings, and a string of asterisks. The rare words used by Klein (*SOL*, *HELOT*, *EFT*, *ABJURE*) would have been unfamiliar to most participants and hence may be considered to be pseudowords. The effects of pronounceability and lexicality were subsequently replicated by Bakan and Alperson (1967) and Fox, Schor, and Steinman (1971).

More recent studies have, however, called into question the effects of lexicality of the distractor on interference with color naming. Monsell et al. (2001) pointed out a number of methodological shortcomings with these early studies. First, they used only a small set (four or five) of words repeated many times, which is likely to counteract the effect of (extraexperimental) familiarity. (In addition, we note that these studies did not test whether the effects are generalizable across items. Moreover, recent studies have shown that the repetition of a small set of distractors produces contingency effects that complicate the interpretation of observed interference effects—see, e.g., Lorentz et al., 2016; Schmidt, Crump, Cheesman, & Besner, 2007.) Second, naming latency was measured by presenting different types of stimuli on different cards and measuring the total time taken to read each list. Blocking stimulus types increases the likelihood of adopting a different response criterion in different conditions. Third, list reading allows preview: Monsell et al. suggested that an overlap between early stages of processing of each item with later stages of processing of the previous item may conceal interference effects. To avoid these shortcomings, Monsell et al. (2001) measured color naming response latencies individually for each item, with the different types of distractor conditions mixed randomly, as is now standard in modern Stroop studies. Monsell et al. replicated the faster color naming latencies observed with consonant string distractors than with words and pseudowords, but observed little difference between words versus pseudowords (and between high- and low-frequency words). Using a method similar to Monsell et al. (2001), Burt (2002, Experiment 5) reported that pseudowords interfered with color naming more than words. However, that experiment also included a repetition manipulation (which interacted with familiarity/lexicality, confirming Monsell et al.'s concern with the early studies), and for unrepeated items, color naming latencies for nonwords (648 ms) did not differ from that for low-frequency words (644 ms).

In sum, previous Stroop color naming studies investigating the interference produced by color-neutral stimuli have found that pronounceable letter strings—words or pseudowords—produce a greater amount of interference than consonant strings, relative to a homogeneous string of letters (e.g., XXX) or a string of pseudofonts (Monsell et al., 2001). The effect of lexicality is more mixed. Although earlier studies (e.g., Bakan & Alperson, 1967; Klein, 1964) reported that words produce more interference than pronounceable pseudowords, more recent studies (Burt, 2002; Monsell et al., 2001) using a larger stimulus set presented only once, and a better method for measuring the color naming latency, have found little difference in interference caused by word and pseudoword distractors. Note that the absence of lexicality effect is inconsistent with the widely held view that the automatic lexical activation of the distractor causes interference (e.g., Augustinova & Ferrand, 2014; Brown, 2011; Neely & Kahan, 2001).

Monsell et al. (2001) took the absence of the lexicality effect (and the frequency effect) to suggest that there are two types of conflict in the Stroop color-naming task: Response conflict and task set conflict. Response conflict refers to the conflict between the specific response to the stimulus: For example, whether to say “hat” or “red” to the word *HAT* presented in red. Stroop interference effects have been traditionally interpreted as reflecting response conflict (Dyer, 1973; MacLeod, 1991). Monsell et al. argued that the absence of lexicality effect is incompatible with the idea that the interference in color naming reflects response conflict, as this would have predicted more interference from a familiar word that has a stronger connection between the orthographic input and the phonological output than a novel letter string. (Monsell et al. verified this assumption empirically by showing that the words were read aloud faster than the pseudowords.) The notion of task set has its origins in the task-switching literature (e.g., Allport, Styles, & Hsieh, 1994; Rogers & Monsell, 1995), and refers to the configuration of cognitive processes necessary to perform the task. A word displayed in color affords two task sets: the task of naming the color and the task of reading. A colored row of Xs, on the other hand, affords only one task set: to name the color. To explain the finding of wordlikeness gradient together with the absence of lexicality effect, Monsell et al. suggested that the interference effects found with color-neutral distractors are due to the conflict between task sets: In a Stroop task, “a control bias is applied endogenously to enable the appropriate task set of color naming and suppress the task set of reading,” but “the wordlike properties of the stimulus evoke exogenously in literate subjects the associated task set of reading” (p. 147). Although Monsell et al. did not specify exactly what they meant by “the task of reading”—it could mean reading aloud (generation of pronunciation), or lexical access, or semantic activation—both the presence of an effect of pronounceability (greater interference produced by words and pseudowords than the consonant strings) and the absence of lexicality effect suggest it is unlikely to be the latter two.

Before accepting the null lexicality effect, there is one methodological issue that should be addressed. This relates to matching words and nonwords on the subword components. Woollams, Silani, Okada, Patterson, and Price (2011) pointed out that words and nonwords differ in what they termed “orthographic typicality”—the frequency of subword components as indexed by bigram (and trigram) frequency, which, unless specifically controlled, will naturally tend to be higher (i.e., be more wordlike) for words than nonwords. It is relevant to

point out that Monsell et al. (2001) noted a link between the wordlikeness gradient and the neuroimaging research that has identified an area of cortex that is activated equally by words and pseudowords, and much more so than by consonant or false font strings (e.g., Price, Wise, & Frackowiak, 1996), suggesting this area as “potential neural correlates of detection of wordlikeness in the sense that we need” (Monsell et al., 2001, p. 147). The area has been dubbed the *visual word form area* (VWFA) by Dehaene and colleagues (e.g., Dehaene, Le Clec’H, Poline, Le Bihan, & Cohen 2002), and subsequent neuroimaging studies (see, e.g., review by Price, 2012; Taylor, Rastle, & Davis, 2013) have indicated that definitive evidence for a preference for real words over pseudowords in the VWFA remains elusive. In contrast, bigram frequency of nonword stimuli has been found to modulate the amount of activation in this area (Binder, Medler, Westbury, Liebenthal, & Buchanan, 2006). Accordingly, in the present study, we matched the words and pseudowords on bigram frequency, and hence they differed only on the familiarity of the whole string. An absence of lexicality effect here would reinforce Monsell et al.’s conclusion that the Stroop interference is not caused by involuntary lexical access.

Vocal Versus Manual Response

The traditional view of the Stroop task is that it involves four major stages—an input process, a decision process, a response selection process, and a response output/execution process (e.g., Lupker & Katz, 1981)—with the general consensus regarding the locus of the Stroop interference being late, either the response output process, or a response selection and a response output process (e.g., Dyer, 1973; MacLeod, 1991). Within this framework, the modality of the response to the color—whether it is vocal or manual—should affect only the last stage or the last two stages. If the wordlike properties of a distractor trigger the task of reading automatically, it would be expected that the wordlikeness gradient should be present, irrespective of whether the required response to color is vocal or manual. In the Stroop literature, there are only few relevant studies and the results do not provide a clear answer to the question.

Before reviewing these studies, it is important to note that unlike the vocal Stroop task, the mapping of color to a manual key press response (e.g., press the “Z” key for red) is arbitrary, and practice is needed to learn the stimulus–response mapping contingency. Some manual Stroop studies (e.g., Sugg & McDonald, 1994) used only two responses. Although this makes the learning of colorresponse mapping easy, in such studies, the Stroop congruence effect is greatly diminished, and with practice, it may even be eliminated (see MacLeod, 1991; Magen & Cohen, 2002). To maintain the comparability with the vocal Stroop studies (which typically use four or more colors), more than two colors should be used, with each color mapped onto a separate response.

A study reported by Keele (1972) meets these criteria. He presented color names (e.g., *BLUE*), noncolor words (e.g., *BIRD*), “scrambled letters” generated from the color words (e.g., *BELU*), and nonletter symbol strings (Gibson forms), presented in one of four colors. Responses were slowest to the color names (604 ms), and the other conditions did not differ from each other (noncolor words = 559 ms; scrambled letters = 560 ms; nonletter symbols = 553 ms). There are several methodological issues with this study that complicate the

interpretation of the results, however. First, it used only four stimuli in each condition, presented repeatedly. Second, the noncolor words and scrambled letters shared the initial letter with the color names. Third, it is unclear whether the word-like stimuli were presented in a color that was onset-congruent or incongruent with the initial letter (e.g., *BELU* presented in blue vs. red), and whether this was controlled across the distractor type. Fourth, some of the scrambled letter stimuli were pronounceable (like *BELU*), but others were not (e.g., *RDE*).

More recently, Sharma and McKenna (1998) compared manual and vocal responses with color-neutral words and a row of Xs, and reported that the difference, which they called the lexical component of the Stroop effect, was present in the vocal version of the task but not in the manual version. However, like Keele (1972), Sharma and McKenna used only four color-neutral words (*TOP*, *CHIEF*, *CLUB*, and *STAGE*), and they presented the different types of items in separate blocks, thus raising the concerns noted by Monsell et al. (2001) earlier regarding adopting different response criteria for different blocks. Also, Sharma and McKenna did not include unpronounceable letter strings or pseudowords. In sum, there is a suggestion in the extant literature that, in contrast to a vocal Stroop task, in a manual Stroop task, color-neutral word distractors are no more interfering than a row of Xs (or nonlinguistic symbols). However, there are methodological concerns associated with the finding, and currently it is not known whether there is graded interference due to the presence of letters, pronounceability, and lexicality.

RT Distribution and Task Versus Response Conflict

The vast majority of Stroop studies report only the analysis of mean RTs and error rates. However, analysis of RT distributions provides richer information and can lead to valuable insights into the dynamics of the cognitive processes underlying the Stroop effect (e.g., Heathcote, Popiel, & Mewhort, 1991; Pratte, Rouder, Morey, & Feng, 2010; Spieler et al., 2000; Steinhauser & Hübner, 2009).

There are different methods of RT distribution analysis (see Balota & Yap, 2011, for a review), and here we describe the approach using a graphical exploratory method: the delta plot. The delta plot is based on the analysis of quantiles. In this analysis, for each participant for each condition, RT data are ordered from the fastest to the slowest and then are divided into equal-sized portions (RT bins), for example, the fastest 10%, the next 10%, and so on, called *quantiles*. In the delta plot, the difference between the conditions is plotted as a function of the quantiles.

In conflict tasks like the Stroop task, flanker task, and Simon task (this task will be described shortly), an interference effect due to a distractor stimulus has been manifested in three different RT distribution patterns (Kinoshita & Aji, 2014; Pratte et al., 2010; Spieler et al., 2000; Steinhauser & Hübner, 2009). In one pattern, the delta plot is a horizontal line with an intercept reflecting the cost, indicating that the effect remains the same size across the quantiles (i.e., the distributions are shifted). In another, the delta plot shows a positive slope indicating a monotonic increase across the quantiles, which captures an effect that is small

for the fastest responses and increases as RT slows. In the third pattern, the delta plot shows a negative slope, indicating an effect that is greatest early on and decreases thereafter.

Pratte et al. (2010) pointed out that the positive delta slope pattern has been observed with a number of “strength” variables (e.g., the intensity or duration of to-be-detected light source), and is concordant with many information/evidence accumulation models (e.g., the diffusion model of Ratcliff, 1978). In these models, information about a decision slowly accrues until a criterion is reached, and the positive slope of the delta plot is typically explained in terms of a difference in the rate of evidence accumulation between the conditions.¹ The positive delta slope pattern is ubiquitous across a wide range of strength manipulations that it may be considered a default pattern (Wagenmakers & Brown, 2007)—that is, most manipulations that impact on the strength of the stimulus modulate the rate of evidence accumulation. This is also the case with the Stroop task: Pratte et al. (2010) noted that in all of the previous Stroop experiments they examined, as well as in their own Stroop experiments, the slope of the delta plot of the Stroop effect was positive, and suggested that this may be explained by the assumption that the information from the distractor influences the effective rate of evidence accumulation. In brief, the idea is that in a Stroop stimulus in which the color and word are integrated into a single object, sampling of information (evidence accumulation) cannot be restricted to the relevant color dimension, and information is sampled from the distractor dimension along with it. (See also Spieler et al., 2000, for a similar suggestion to explain the Stroop interference effect reflected in the τ (tau) parameter in the ex-Gaussian). In contrast, the delta plot slope in the Simon task is negative. In the Simon task, a stimulus (e.g., a green or red circle) is presented to the left or right, and participants are asked to respond to the color of the stimulus by pressing one of two spatially defined keys (e.g., the right key for red and the left key for green). The Simon effect refers to the faster response when the stimulus is presented at the congruent spatial location as the required response (e.g., when a red circle is presented on the right in the example just described). Pratte et al. suggested that the negative delta slope pattern found in this task may be explained by “quick, automatic motor activation from the distracting information that passively decays in time” (p. 2023), and also that the occasional finding of reversal of the effect for the slowest responses may reflect the active inhibition of motor responses.

To recap, all previous Stroop RT distribution studies (Kinoshita & Aji, 2014; Pratte et al., 2010; Spieler et al., 2000; Steinhauser & Hübner, 2009) showed a positive delta slope in the interference observed with the incongruent color words, and this can be interpreted in terms of a reduction in the effective rate of evidence accumulation for the target color. We suggest that this positive delta slope reflects a *task conflict*, based on the argument that any task is a process of accumulating evidence for a task goal, be it generating a speech response to the color, or categorizing the color, and reducing the rate of evidence accumulation amounts to interfering with the task. It follows from this that task conflict—that is, interfering with the task of color naming (in the vocal task) or the task of color categorization (in the manual

¹Pratte et al. (2010) noted that although not all previous studies used the delta plot, all positive delta plots are compatible with either drift rate or bound changes in the diffusion model (Ratcliff, 1978), or with changes in the τ (tau) parameter in the ex-Gaussian (Andrews & Heathcote, 2001). In addition, we note that changes in the μ (mu) parameter alone in the ex-Gaussian is consistent with a shift in RT distribution and a flat delta slope.

task)—would be manifested as a positive delta slope. In contrast to the task conflict, we expect the *response conflict*—in the sense of competing motor response tendencies—to be manifested as a negative delta slope. As noted above, this pattern has been found with the Simon task, in which the (task-irrelevant) spatial location of the stimulus cues the spatially defined response. A negative delta slope has also been observed in the category congruence effect with masked primes in a number judgment task (“Is the number bigger/smaller than 5?”) when the prime had been responded to as a target, and this was interpreted in terms of learned stimulus–response mapping (Kinoshita & Hunt, 2008). The idea is that the motor response (e.g., press the right or left key) activated by the prime or distractor decays over time (or that it is actively inhibited to prevent an error response); hence, the effect of the distractor should be smaller for the slowest responses. Thus, we expected that the conflict between opposite response tendencies activated to the motor level should be revealed as a negatively sloped delta plot.

The Present Study

Although the Stroop interference effect is widely regarded as the gold standard in the automaticity of reading, what aspect of reading process is triggered automatically to produce interference is not well-specified. Specifically, what is meant by “reading” is unclear: Is it the process of generating pronunciation (as in reading aloud), finding a lexical entry that matches the visual input (lexical access), or retrieving meaning, that interferes with responding to the color? The present study aims to provide answers to these questions.

The literature reviewed has indicated that in a vocal Stroop task using color-neutral stimuli, the wordlikeness properties—the presence of letters and the pronounceability of the letter string but not necessarily lexicality—are key factors that produce interference with color naming. According to Monsell et al. (2001), in a Stroop color naming task, participants endogenously apply control bias to suppress the task of reading, but the wordlike properties of the distractor exogenously turn it back on. We will take advantage of this finding to investigate the nature of the reading task being triggered automatically in two versions of the Stroop task: one using the vocal color naming response and the other using the manual color-categorization response.

The traditional “response competition” view of the Stroop task leads one to expect that the modality of response to the color should affect only the later response selection and response execution processes, and hence there is no reason to expect the response modality to impact on the wordlikeness gradient. In this framework, the underlying reading task that produces interference in a Stroop task is assumed to be invariant. In contrast, in the Bayesian Reader (Norris, 2006; Norris & Kinoshita, 2008) framework, the task is critical. There is no single invariant reading process; rather, the reader is viewed as a Bayesian decision maker accumulating noisy evidence from the percept, with the task goal guiding the nature of evidence to be accumulated. In our work with masked priming—another phenomenon that is widely regarded as reflecting the automaticity of reading—we have found that the pattern of masked priming effects depend critically on the task goal (see Kinoshita & Norris, 2012, for a review). Of particular relevance, the role of phonology is much more prominent when the task requires a speech response: For example, in reading aloud, but not in lexical decision, a

mere overlap in the onset between the masked prime and the target (e.g., *save*–*SINK*) produces priming (termed the *masked onset priming effect*; Forster & Davis, 1991; see Kinoshita, 2003, for a review), as expected from the fact that speech production involves a left-to-right serial phonological encoding process (cf. Levelt, Roelofs, & Meyer, 1999). This task dissociation is as expected from the Bayesian Reader framework (Norris & Kinoshita, 2008; see Kinoshita & Norris, 2012, for a summary of task-dependent masked priming effects), and argues against the view that the process of reading is invariant, and the required response—whether it is to read aloud the word, classify it as a word or a nonword, classify the word into a semantic category, and so forth—is simply tacked on at the end of the process. In the vocal Stroop task, the task goal is to *name* the color, and in the manual Stroop task, it is to *categorize* the color. Given this, we expected the reading process that competes with responding to color in these two tasks to be also different.²

To investigate this, we presented four colors in two versions of the Stroop task: In Experiment 1, we required the participants to name the color, and in Experiment 2, to categorize the color by means of a manual key-press response. We compared five types of neutral strings: real words (e.g., *HAT*), pronounceable pseudowords (e.g., *HIX*), consonant strings (e.g., *HDK*), nonalphabetic symbol strings (e.g., *#\$%*), and a row of Xs. String length (three, four, five, and six letters) was manipulated within each of the string types. In addition to the neutral strings, color words corresponding to the four colors used were presented in incongruent colors (e.g., *RED* in green) as a manipulation check. (The extant literature indicates that when more than two colors are used, interference to incongruent color words is a robust finding in the manual as well as vocal Stroop task). Note that the color-congruent condition (e.g., *RED* presented in red) was absent. Given this, there was no incentive to attend to the carrier stimulus (cf. Melara & Algom, 2003), and thus any interference with responding to the color was assumed to reflect the nature of reading process triggered involuntarily by the stimulus.

As well as analyzing the mean RTs (and error rates), we analyzed RT distributions. As we explained above, we expected the task conflict to be manifested as a positive delta slope, and response conflict as a negative delta slope.

Experiment 1: Vocal Stroop Task

Method

Participants—Twenty Macquarie University undergraduates participated in Experiment 1 for course credit. All participants were fluent English speakers and had normal color vision and normal or corrected-to-normal vision. Participants were tested individually in a quiet room.

²We hasten to add that the view that phonology plays a more prominent role in reading tasks that require a speech output than in silent reading tasks like lexical decision or semantic categorization is not specific to the Bayesian Reader. However, in a Stroop task, a speech response is not required to the *word*; in fact, the reader is told to ignore the word and respond to the color. The issue at question here is whether or not the task of reading, triggered against the reader's intention, is dependent on the goal of the task required to the *color*.

Materials—The stimuli consisted of 80 real words (RW), 80 pseudowords (PW), 80 consonant strings (CS), 80 symbol strings (Symbols), and 80 strings of Xs (XXX), as well as four color words, presented in one of four colours (incongruent color name; *red, pink, green, and blue*). All strings except the color words were three, four, five, or six letters long (20 of each length). The words were medium frequency (mean = 40.9, range = 11.0 to 89.5 per million subtitle frequency; Brysbaert & New, 2009), and none had a meaning associated with color. The pseudowords were all pronounceable and matched on position-dependent bigram (type) frequency to the words (words = 35.3; pseudowords = 34.8) using the MC-Word database (Medler & Binder, 2005). Consonant strings were all nonpronounceable. None of the word, pseudoword, and consonant string stimuli started with R, P, G, or B (the initial letters of the color names), because interference to color naming is reduced when the distractor shares the onset with the color name (Coltheart, Woollams, Kinoshita & Perry, 1999).³ Symbol strings were generated by replacing the letters in the consonant string stimuli with the nonalphabetic characters !, #, \$, %, & *,), <, >, /, +, and ?, thus mimicking the consonant strings in the heterogeneity and occasional repetition of characters. In addition, there were 16 real words, 16 pseudowords, 16 consonant strings, and 16 symbol strings selected according to the same criteria that were used as practice items, along with the row of Xs and the incongruent color words. The practice trials were not included in the analysis.

Each neutral string (except the Xs) of each length was presented once, in one of four colors: red, pink, green, or blue; Xs of each length were presented in one of four colors equally often. Each color word was presented in an incongruent color 5 times (e.g., the word *red* was presented in pink 5 times, green 5 times, and blue 5 times). Hence in total, there were 400 neutral trials and 60 incongruent trials, with each color occurring equally often.

Apparatus and procedure—Participants were tested individually in a quiet room, seated approximately 60 cm in front of a flat screen monitor upon which the stimuli were presented. Each participant completed five blocks of 92 trials, preceded by a block of 92 practice trials, with a short self-paced break between blocks. Each block contained an equal number of trials of each carrier string type of each length, and an equal number of incongruent trials, presented in one of the four colors equally often. A different random order of trials was generated for each block.

Participants were instructed at the outset of the experiment that on each trial they would be presented with a color name or string of letters or symbols presented in one of four colors: red, pink, green, and blue. Participants were told to ignore the word and to name the color by speaking into a head-worn microphone. The color naming latencies were measured by means of a voice key. Naming errors were recorded by the experimenter, who sat beside the participant, with an error record sheet with a preordered sequence of trials, generated randomly per block.

³Due to experimenter error, seven of the 80 pseudowords started with one of these letters (see Appendix). Only one was presented in a congruent color (PAME in pink). Excluding these items from the data made little difference (mean RT 640 ms with or without the seven items).

Stimulus presentation and data collection were achieved using the DMDX display system developed by K. I. Forster and J. C. Forster at the University of Arizona (Forster & Forster, 2003). Stimulus display was synchronized to the screen refresh rate (10.1 ms).

Each trial started with the presentation of a fixation sign (+) for 500 ms in the center of the screen. This was followed immediately by a test stimulus presented in one of four colors. The stimulus remained on the screen for 2,000 ms or until the participant's response, whichever occurred sooner. All stimuli were presented in Courier New, size 12 font, against a black background.

Results and Discussion

The dependent variables were color naming latency and error rate. Of the 460 test trials, those in which a wrong response was given or no response was made within the 2,000-ms timeout period were treated as errors and excluded from the RT analyses. The correct mean RTs and error rates for the six carrier string types are summarized in Table 1.

In this and a subsequent experiment, we report the analyses of correct RT and error rates for the six carrier string types, using linear mixed effects model with subjects and items as crossed random factors (Baayen, 2008). We then report the analysis of RT distributions for correct trials.

Correct RT—The preliminary treatment of RT data for this analysis was as follows. First, we examined the shape of the RT distribution for correct trials, and applied a log transformation (which best approximated a normal distribution after excluding 27 data points faster than 250 ms as outliers) in order to meet the distributional assumption of the linear mixed effects model.

The data were submitted to a linear mixed effects model using the lme4 package (Bates, Maechler, & Bolker, 2013, Version 1.1–5) implemented in R 3.0.3 (R Core Team, 2014). Degrees of freedom (estimated using Satterthwaite's approximation) and p values were estimated using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2013, Version 2.0–11). In line with the recommendation to keep the random effect structure maximal (Barr, Levy, Scheepers & Tily, 2013), the initial model included random slopes (on carrier string type), but, as the models did not converge, the model we report included only the subject and item intercepts. Using R syntax, the model we report was: $\log RT \sim \text{stringtype} + (1 | \text{subject}) + (1 | \text{string})$, with 20 subjects and 328 strings. The string type factor was referenced to the row of X condition. The model's estimates of the effect of each string (distractor) type, the associated standard error, estimated degrees of freedom, and t and p values are shown in Table 2.

The model showed that referenced to the row of Xs, all other distractor types interfered with color naming (for the consonant strings, the effect was marginal), except the symbol strings: INC, $t = 7.34$, $p < .001$; words, $t = 4.347$, $p < .001$; PW, $t = 3.814$, $p < .001$; consonant strings, $t = 1.729$, $p = .086$; and symbols, $t = .111$, $p = .91$. In addition, real words and pseudowords produced more interference than the consonant strings: words, $t = 6.354$, $p < .$

001; pseudowords, $t = 5.056$, $p < .001$. The real words and pseudowords did not differ from each other, $t = 1.29$, $p = .198$.

To further investigate these differences, we examined the color naming latencies as a function of string type and string length (recall that all of the neutral strings were either three, four, five, or six letters in length). Mean (untransformed) color naming latencies as a function of string length and string type are shown in Table 3. The pattern shown by Table 3 is that the pronounceable strings—real words and pseudowords—showed a string length effect (with the color naming being slower for longer strings), but the nonpronounceable strings (consonants, symbols, and row of Xs) did not. This was supported by the analysis of logRT as a function of string length: A significant effect of string length was found for the pronounceable strings (3,037 observations, 20 subjects, 160 strings), $t = 2.321$, $p < .03$, but not for the unpronounceable strings (4,612 observations, 20 subjects, 164 strings), $t = -.602$, $p = .548$.

Error rate—The error data (a total of 9,200 observations) were analyzed using a logit mixed model (Jaeger, 2008) using the same model as for correct RT. The model's estimates, standard error, and z and p values are shown in Table 4.

Relative to the row of Xs, only the incongruent color words produced a greater error rate, $z = 4.91$, $p < .001$. There were no significant differences among the neutral string types.

Quantile analysis—RT distributions were analyzed with QMPE (Version 2.18; Cousineau, Brown, & Heathcote, 2004). To calculate the quantile estimates, correct RTs for each condition for each participant were sorted from the fastest to slowest and subsequently divided into five equal-sized bins (fastest 20%, next fastest 20%, etc.). The RT of the slowest trial of the lower bin and the fastest trial of the higher bin make up the four observed quintile estimates generated by QMPE. Only the first trial of the slowest quantile is used to calculate the quantile estimate for the last quantile, the quantile estimates are therefore not unduly affected by the extremely fast or slow outliers, and hence RT data were not trimmed for outliers in generating the quantiles.

The delta plot, referenced to the row of Xs condition, averaged over the participants per condition are presented in Figure 1. It is apparent from Figure 2 that the neutral strings produced different amounts of interference, and the differences increased across the quantiles. This observation is supported by the 4 (quantiles) \times 5 (neutral carrier string type: words, pseudowords, consonant strings, symbols, Xs) ANOVA. There was a robust main effect of neutral string type, $F(4, 76) = 49.49$, $MSE = 1,716.44$, $p < .001$, and an interaction between neutral string type and quantiles, $F(12, 228) = 9.81$, mean square error (MSE) = 234.04, $p < .001$. Simple effects analyses using the row of Xs as the reference condition showed that all other neutral string types except the symbol strings interfered with color naming: words, $F(1, 19) = 77.52$, $MSE = 5,221.23$, $p < .001$; pseudowords, $F(1, 19) = 118.424$, $MSE = 2,629.13$, $p < .001$; consonant strings, $F(1, 19) = 57.49$, $MSE = 1,018.46$, $p < .001$; and symbol strings, $F(1, 19) = 1.214$, $MSE = 1,467.18$, $p = .28$. Each of the string types that produced interference relative to the row of Xs (words, pseudowords, and consonant strings) increased linearly in the size of interference across the quantiles, as

indicated by the significant interaction with the linear trend of the quantiles factor: words, $F(1, 19) = 28.83$, $MSE = 983.05$, $p < .001$; pseudowords, $F(1, 19) = 27.64$, $MSE = 366.79$, $p < .001$; and consonant strings, $F(1, 19) = 10.65$, $MSE = 111.83$, $p < .005$. The (absence of) interference produced by the symbol strings remained constant across the quantiles, as indicated by the null interaction between the linear trend of the quantiles factor and the difference between the symbol string and the Xs, $F(1, 19) < 1.0$. In addition, the difference between the words and incongruent color conditions was highly significant, $F(1, 19) = 79.434$, $MSE = 5,475.344$, $p < .001$, and the difference increased linearly across the quantiles, $F(1, 19) = 48.761$, $MSE = 1,029.548$, $p < .001$.

Summary—In sum, in the vocal Stroop task, replicating Monsell et al. (2001), color naming latency increased with the pronounceability of the carrier string, with the words and pseudowords producing the slowest color naming latency (which did not differ from each other), followed by consonant strings, then the row of Xs and nonalphabetic symbols (which did not differ from each other). The RT distribution analysis showed that all interference effects (relative to a row of Xs), including the standard effect observed with incongruent color names increased across the quantiles, suggesting that the interference reflects the reduced rate of evidence accumulation. In addition, interference effects increased as a function of string length when the string was pronounceable (words and pseudowords), but not when it was unpronounceable (consonant strings, symbol strings, and row of Xs). We now turn to the manual Stroop task to see if these patterns are also seen when the task does not require a speech response.

Experiment 2: Manual Stroop Task

Method

Participants—An additional 20 Macquarie University undergraduates, none of whom had taken part in Experiment 1, participated in Experiment 2 for course credit. All participants were fluent English speakers and had normal color vision, and normal or corrected-to-normal vision. Participants were tested individually or in pairs in a quiet room.

Materials—The stimuli were identical to those of Experiment 1.

Apparatus and procedure—The apparatus and the general procedure were identical to those of Experiment 1, except that the required response was a manual color categorization response. As in Experiment 1, each participant completed five blocks of 92 trials, preceded by a block of 92 practice trials.

Participants were instructed at the outset of the experiment that on each trial they would be presented with a color name or a string of letters or symbols in one of four colors: red, pink, green, or blue. Participants were told to ignore the meaning of the word and to classify the color, as quickly and accurately as possible, by pressing one of the keys on the keyboard, specifically, the “Z” key for red, “X” for pink, “N” for green, and “M” for blue. The four keys were arranged on the bottom row of the QWERTY keyboard, and participants were instructed to place their left middle and index fingers on the Z and X keys, and their right index and middle fingers on the N and M keys, respectively.

Each trial started with the presentation of a fixation sign (+) for 500 ms in the center of the screen. This was followed immediately by a test stimulus presented in one of the four colors. The stimulus remained on the screen for 2,000 ms or until the participant's response, whichever occurred sooner. Following each response, participants were given accuracy feedback with the message "Correct," "Wrong," or "No response" (if no response was made within the 2,000-ms timeout period).

Results and Discussion

As in Experiment 1, we first report analyses of response latency on correct trials and error rate using linear mixed effects modeling treating subjects and items as crossed random factors. The correct mean RTs and error rates for the six carrier string types are summarized in Table 1.

Correct RT—The preliminary treatment of RT data was the same as for Experiment 1. In Experiment 2, 3 data points faster than 250 ms were excluded as outliers. To be consistent with Experiment 1, the correct RTs were log-transformed.⁴ The model's estimates, standard error, estimated degrees of freedom, and t and p values are shown in Table 5.

The model showed that, referenced to the row of Xs, the only condition that differed was the incongruent color words, $t = 13.075$, $p < .001$. Thus, although a standard Stroop interference effect to color-incongruent words was robust, no difference was observed among the neutral string types.

The color response latencies as a function of string length and string type are shown in Table 2. The analysis of the length effect for the logRT in the manual task showed that there was no effect for either the pronounceable strings (3,071 observations, 20 subjects, 160 strings), $t = -0.609$, $p = .543$, or the unpronounceable strings (4,595 observations, 20 subjects, 164 strings), $t = -1.307$, $p = .191$. The combined analysis testing the interaction between Task \times Length for the pronounceable strings (words and pseudowords) showed a significant interaction, $t = 2.723$, $p < .01$, consistent with the presence of length effect in the vocal task and its absence in the manual task for the pronounceable distractors.

Error rate—The error data (a total of 9,200 observations) were analyzed using a logit mixed model (Jaeger, 2008) using the same model as for correct RT, and subjects and item intercepts as crossed random factors. The model's estimates, standard error, and z and p values are shown in Table 6.

Mirroring the RT data, the incongruent color words produced greater error rate in color categorization than the row of Xs, $z = 4.831$, $p < .001$. In addition, relative to the row of Xs, symbol strings produced a significantly greater error rate, $z = 2.959$, $p < .01$, as well as the consonant strings, $z = 2.802$, $p < .01$, and pseudowords, $z = 1.977$, $p < .05$.

⁴We also analyzed an inverse-transformed RT ($-1,000/\text{RT}$), which is commonly used for speeded manual classification responses (see Kinoshita, Mozer, & Forster, 2011; Masson & Kliegl, 2013). The analysis produced the same pattern as logRT.

Combined analysis—To test whether the manual Stroop task and the vocal Stroop task produced different patterns of interference, we combined the logRT data from the two experiments, and tested whether task (vocal vs. manual) interacted with each of the interference effect produced by the different distractors (referenced to the row of Xs). Task was contrast-coded (manual vs. vocal: $-.5$ vs. $.5$). There was a significant main effect of task, $t = -2.393$, $p < .03$, indicating that overall responses were faster in the vocal task than the manual task. Significant interaction with task was observed with the interference effect produced by words, $t = 8.414$, $p < .001$; pseudowords, $t = 7.217$, $p < .001$; consonant strings, $t = 3.234$, $p < .002$; and the incongruent color names INC, $t = 8.941$, $p < .001$. These interactions indicated that the interference produced by words, pseudowords, consonant strings, and INC were all greater in the vocal task than the manual task. Over the two tasks, symbol strings did not interfere relative to the row of Xs, $t < 1$, $p = .407$, and it did not interact with Task, $t = -1.26$, $p = .207$.

In addition, we tested the interaction between the task and length for the pronounceable strings (words and pseudowords). The interaction was significant, $t = 2.305$, $p < .03$, consistent with the presence of the length effect for the pronounceable strings in the vocal task but not the manual task.

For the error data, the model testing task by string type interactions did not converge.

Quantile analysis—As in Experiment 1, the quantile estimates of correct RTs of color-neutral stimuli were analyzed as a 4 (quantiles) \times 5 (neutral carrier string type: words, pseudowords, consonant strings, symbols, Xs) ANOVA. The delta plot, referenced to the row of X condition, averaged over the participants per condition, is presented in Figure 2.

It is apparent from Figure 2 that although there is a substantial Stroop interference effect in the incongruent color condition that increases across the quantiles, there is little difference among the five neutral strings throughout the quantiles. The pattern apparent in the figure is supported by the analysis of RT distribution as a 4 (quantile) \times 5 (neutral string type: words, pseudowords, conditional stimulus [CS], symbols, Xs) ANOVA, with quantile and string type as within-subject factors. The main effect of neutral string type was nonsignificant, $F(4, 76) = 1.276$, $MSE = 1,257.92$, $p = .287$. The interaction between the string types and quantiles was nonsignificant, $F(12, 228) < 1.0$. Simple effects using the row of Xs as the reference condition showed no difference with any of the other strings, except for a marginal effect for the symbol strings: words, $F(1, 19) < 1.0$; pseudowords, $F(1, 19) < 1.0$; CS: $F(1, 19) < 1.0$; symbol string, $F(1, 19) = 3.242$, $MSE = 2,048.74$, $p = .088$. None of the effects increased across the quantiles, all $F_s < 1.0$.

The difference between the words and incongruent color conditions was highly significant, $F(1, 19) = 36.261$, $MSE = 873,207.25$, $p < .001$, and the interference effect increased linearly across the quantiles, $F(1, 19) = 18.209$, $MSE = 6,794.50$, $p < .001$.

Combined analysis of quantiles—The RT data from the color-neutral distractors in Experiment 1 and 2 were combined and analyzed as a Task (vocal vs. manual) \times Distractor Type (words, PW, CS, symbols, row of Xs) \times Quantiles (1–4) factorial design. Omnibus

ANOVA showed that task interacted with distractor type, $F(12, 456) = 3.563$, $MSE = 1,068.383$, $p < .001$; as well, there was a significant triple interaction, $F(12, 456) = 4.689$, $MSE = 1,405.782$, $p < .001$. Consistent with the LME analysis of logRT, there was a significant interaction with task and each of the interference effect relative to the row of Xs—words, $F(1, 38) = 54.736$, $p < .001$; pseudowords, $F(1, 38) = 64.206$, $p < .001$; consonant strings, $F(1, 38) = 20.364$, $p < .001$ —but no interaction with the interference effect to symbols, $F(1, 38) < 1.0$. Furthermore, for the interference effects (relative to the row of Xs) for the words and pseudowords, there was a significant triple interaction with task and the linear trend of quantiles factor, indicating that the increase in the interference effect across the quantiles was greater in the vocal task than the manual task: for words, $F(1, 38) = 14.266$, $p < .001$; pseudowords, $F(1, 38) = 7.901$, $p < .01$. The triple interaction was marginal for consonant strings, $F(1, 38) = 3.254$, $p = .079$. The triple interaction was nonsignificant for symbols, $F(1, 38) < 1.0$.

Summary—In sum, the key finding of the manual Stroop task is that there was little difference in the amount of interference with color categorization among the five neutral string types. The absence of difference between the neutral string types with manual color categorization response contrasts sharply with the graded pattern of interference observed with the vocal Stroop task.

General Discussion

The goal of the present study was to provide a better specification of which aspect of the reading process is triggered automatically to produce interference in a color-word Stroop task. To this end, we compared five types of color-neutral distractors—real words (RW), pronounceable pseudowords (PW), consonant strings, nonalphabetic symbol strings and a row of Xs—as well as the incongruent color words (INC) using a vocal color naming response (Experiment 1) and a manual color categorization response (Experiment 2). The key findings are that (a) in the vocal Stroop task, color naming latencies showed a “wordlikeness gradient,” with the greatest interference observed with the real words and pseudowords (which did not differ from each other), then consonant strings, and the nonalphabetic symbol strings producing a null interference effect, relative to the baseline row of Xs condition; (b) in contrast, the type of color-neutral distractors had little effect on the color categorization latencies in the manual Stroop task (Experiment 2); (c) in the vocal task, but not in the manual task, interference increased with length for words and pseudowords, but not for nonpronounceable strings; (d) a robust interference effect was found with the incongruent color words (e.g., *RED* presented in green) in both the manual and vocal Stroop tasks; and (e) the delta plots showed that each interference effect (relative to the row of Xs) when significant for its overall mean increased across the quantiles (for both manual and vocal Stroop tasks). These results provide novel insights into the nature of the reading processes that produce interference in the Stroop task, as discussed below.

Task Dissociation

First and foremost, the dissociation between the manual and vocal Stroop tasks indicates that the task of reading is not invariant, even when it is triggered automatically. In the vocal

(color naming) task, the presence of (heterogeneous sequence of) letters and the pronounceability of the string have been found to be key properties that “evoke exogenously in literate subjects the associated task set of reading” (Monsell et al., 2001, p. 147). Consistent with most extant models of word recognition, the implicit assumption here is that there is one invariant task set of reading. The impact of these properties in the vocal Stroop task, but not in the manual Stroop task, argues against this view, and the associated view of the Stroop task in which the modality of response to the color affects only the late response selection and response execution process, and has little impact on the processes prior to these stages. The finding also challenges the conceptualization of “automaticity of reading” as purely stimulus-driven, in that the stimulus characteristics that trigger the reading task in the vocal Stroop task do not do so in the manual Stroop task. Instead, the observed dissociation is consistent with the assumption of the Bayesian Reader (Kinoshita & Norris, 2012; Norris, 2006; Norris & Kinoshita, 2008) in indicating that the task goal—what response is required to the color—is critical. Naming the color and categorizing the color involve different processes; consequently, the task of reading that competes with these processes is also different. What, then, are these differences?

Wordlikeness Gradient in the Vocal Task and the Manual Task

In the vocal Stroop task, the pattern of wordlikeness gradient replicated the results reported by Monsell et al. (2001), with the greatest interference observed with the real words and pseudowords, which did not differ from each other, then consonant strings, and the least interference produced by the nonalphabetic symbol strings (which did not differ from a row of Xs). The “wordlikeness gradient” is therefore better termed the “pronounceability gradient.” That letter strings (including consonant strings) produced more interference than nonalphabetic symbol string supports Monsell et al.’s suggestion that in literate subjects, the presence of letters (more specifically heterogeneous sequence of letters) triggers the task of reading involuntarily. Specifically, the task of “reading” here seems to be the generation of a speech code from a letter string.

It is important to emphasize there was no effect of lexicality—words and pseudowords that were matched on sublexical orthographic typicality (bigram frequency) produced equal amounts of interference. This means that the orthographic familiarity of the *whole* string does not modulate the interference effect, confirming Monsell et al.’s (2001) conclusion and arguing against the commonly held view that the Stroop effect reflects involuntary lexical access (e.g., Augustinova & Ferrand, 2014; Brown, 2011; Labuschagne & Besner, 2015; Neely & Kahan, 2001). (The absence of a lexicality effect also means that whether or not the letter string has a meaning did not modulate the Stroop interference effect, and we will return to this issue later.) This also has implications for the major models of the Stroop task (e.g., Cohen, Dunbar & McClelland, 1990; Roelofs, 2003). The absence of lexicality effect is at odds with the key assumption of the Cohen et al. (1990) model. According to Cohen et al., the Stroop interference effect, and more specifically, why word reading interferes with color naming, but not vice versa, stems from the fact that word reading is more practiced than color naming. However, this assumption cannot explain why word distractors did not interfere more with color naming than pseudoword distractors, because word reading, by definition, is more practiced than pseudoword reading.

In contrast, the absence of lexicality effect is consistent with Roelofs's (2003) model of Stroop naming task. The model, couched within his theory of speech production, WEAVER ++, attributes the involuntary nature of Stroop interference, and the asymmetry between the interference caused by word distractors on color naming and by color distractors on word reading, to an architectural difference in reading words and naming colors. Specifically, whereas color naming is conceptually driven, word reading typically is not, that is, "written words in alphabetical systems are intrinsically tied to their sounds, whereas colors are not" (Roelofs, 2003, p. 96), and in his model of speech production, word form perception can be linked directly to the encoding of phonological form in the speech production process without being mediated by the retrieval of word concept. This tight link between the alphabetic writing system—in which a letter or letter cluster maps onto a phoneme—and speech production could also explain why the consonant string distractors interfered with color naming more than a row of Xs. In sum, when investigating the interference caused by the letter-string distractors in the color naming Stroop task, due consideration needs to be given to the role of speech production processes.

Another result of note is the length effect observed with pronounceable strings in the vocal task. With words and pseudowords, the interference increased with length, but this did not increase for the nonpronounceable strings (row of Xs, nonalphabetic symbols, and consonant strings). Moreover, in the manual Stroop task, no length effect was observed with either the pronounceable or unpronounceable strings. As the length effect for the pronounceable strings in the vocal task was not very strong (and the difference was mainly observed with the six-letter strings), definitive interpretation must await a replication; nevertheless, the results have several interesting implications. First, the fact that the length effect was limited to pronounceable strings, and that the effect was not observed in the manual Stroop task, argues against visual encoding as the locus of interference and reinforces the interpretation that what competed with color naming in the vocal Stroop task is the phonology of the distractor. Second, the fact that interference is greater for longer strings implies that the competition is at the level of speech production. Two forms of competition could lead to a length effect. First, generation of the color name could be delayed while at least part of the interfering stimulus is processed. This would be expected to produce a simple shift in the RT distribution. Alternatively, the production of the color name and the interfering stimulus could compete for resources in such a way as to produce a slowing down of the naming process, leading to a change in the delta slope over quantiles. These possibilities should be investigated in future studies.

In contrast to the vocal task, in the manual Stroop task, no wordlikeness gradient was observed. It is important to note that a substantial interference effect was observed with the incongruent color names randomly intermixed with the color-neutral distractors. Because it could not have been known until a word is read that it was an incongruent color word, the total absence of wordlikeness gradient here, together with the interference effect observed with the incongruent color words, cannot be explained in terms of "the absence of reading." The difference between the vocal Stroop task versus the manual Stroop task is instead consistent with Burt's (2002) claim that "a factor likely to be important in the color-naming interference observed in the non-color-word Stroop task is the vocal response requirement" (p. 1033). As suggested above, it is the process of generating a speech code from the letter

string that produced interference in the vocal task: In the manual Stroop task, because the goal is to categorize the color and not to utter the color name, no interference due to the generation of speech code to the distractor is observed.

The task dependence observed with the wordlikeness gradient here has important implications for the choice of neutral baseline condition used to partition the overall Stroop effect into the facilitation and interference components (the difference between the congruent condition and a neutral condition, and the difference between the incongruent condition and a neutral condition, respectively). Relative size of facilitation and interference components is important to the theoretical interpretation of the Stroop effects (e.g., Brown, 2011; Goldfarb & Henik, 2007; Lorentz et al., 2016), but currently there is little consensus on what constitutes an appropriate baseline (see, e.g., Besner, 2001; Brown, 2011). Our findings reinforce the point that the choice of a neutral baseline requires due consideration of the mechanism responsible for interference. In the vocal task, speech production processes play a vital role, and hence factors like the overlap in the onset between the distractor and the color name are likely to be important (cf. Coltheart et al., 1999), but not in the manual task.

Interference to Incongruent Color Words

Finally, we turn to the interference observed with the incongruent color names (e.g., *RED* presented in blue). Responses to incongruent words were substantially slower than to color-neutral words, which is of course a standard finding, and it was observed here both in the vocal task and the manual task. This component of interference is widely assumed to be semantic in origin: The meaning of the word distractor *RED* presented in blue competes with the semantic features of the blue color. However, this raises a puzzle: The color-neutral words (e.g., *HAT*, *STORM*) are also semantically incongruent with the color. Why do the color-neutral words produce no more interference than the pseudowords (e.g., *HIX*, *STASE*), which do not have meaning and, hence, by definition, do not contain semantic conflict?

Our explanation is that the semantic features considered in the evidence accumulation process in the Stroop task are not general, but specific to color. That is, only the semantic features of the word distractor that are diagnostic of color compete with the target color. This idea was proposed by Norris (2006) earlier in explaining how the Bayesian Reader accounts for the pattern of interference effects produced by nonword neighbors that closely resemble a specific word (e.g., *TURPLE*–turtle; *TABRIC*–fabric). In semantic categorization tasks, these nonword neighbors interfere only if they are relevant to the target category, for example, *TURPLE* interferes more than *TABRIC* if the target category is “animal” (Forster & Hector, 2002).⁵ In contrast, in the lexical-decision task, Forster and Hector (2002) showed that the animal and nonanimal nonword neighbors interfere equally, and more than “nonneighbors” (nonwords that are not neighbors of any word, e.g., *GLIMON*). Norris suggested that this is because in semantic categorization tasks the reader does not wait until the form information is fully resolved (i.e., decide whether the input is turple or turtle), and

⁵More recently, Bell, Forster, and Drake (2015) extended this finding to masked primed semantic categorization, showing that when the task is to categorize the word as a vegetable or a city name for example, an orthographically related nonword (or a word) facilitated the categorization of a target word (e.g., *lucchibi* – *ZUCCHINI*; *capable* [cabbage] – *LETTUCE*).

the semantic assessment is made with reference to the information diagnostic of the category. A similar idea can be found in Roelofs's (2003) model of speech production. Roelofs views speaking as a goal-referenced action, and the idea that "goal concepts" (concepts/semantic features of what the speaker intends to say) are flagged early, during the earliest "conceptual preparation" stage of speech production, is a key assumption in his model of the Stroop color naming. Roelofs noted that "in making decisions, a cognitive system can, in principle, draw on all the information available, but the amount may be indefinitely large in that everything may potentially be relevant" (p. 120). Instead, an intelligent cognitive system could be selective and evaluate the evidence (here, the semantic features) that are directly relevant to the task goal which, in a Stroop task—vocal or manual—pertains to the color identity. Within this view, non-color-associated words like *HAT* produce no more interference than pseudowords like *HIX* because the semantic features (e.g., it is inanimate, it is an article of clothing) are not diagnostic of color. This explanation finds support in the finding that color-associated words (e.g., *SKY*, *BLOOD*) presented in an incongruent color, and other color words that are not in the response set produce greater interference than the color-neutral words (e.g., for the vocal task see Klein, 1964; for the manual task, see Sharma & McKenna, 1998; for both tasks see Risko, Schmidt, & Besner, 2006). Note that this explanation differs from the view that the semantically based Stroop interference effects reflect "automatic" semantic activation (e.g., Augustinova & Ferrand, 2014; Neely & Kahan, 2001), in that, here, these semantically based effects are assumed to be closely tied to the goal of the task.

The difference between the incongruent color condition and the color-neutral word condition is commonly interpreted in terms of response conflict (e.g., saying "red" vs. "blue" to the word *RED* presented in blue) occurring late in processing during the response selection stage (e.g., Augustinova & Ferrand, 2014; Monsell et al., 2001). For example, Monsell et al. (2001) suggested that the lexical access for the color words in the response set are primed, and hence the endogenous suppression of the reading task is not sufficient to prevent the "breakthrough" of these words to be activated to the level of response, resulting in "additional competition to be resolved by a response-selection process" (p. 148). This view has a parallel in the "response exclusion hypothesis" (e.g., Finkbeiner & Caramazza, 2006; Janssen, Schirm, Mahon, & Caramazza, 2008) proposed to explain the semantic interference effect in picture-word interference task. The effect refers to the finding that word distractors that are semantic coordinates of the target pictures (e.g., the word *CAT* presented with the picture of dog) interfere with the naming of pictures more than the semantically unrelated distractors (e.g., the word *PEN* presented with the picture of dog). The response exclusion hypothesis suggests that the effect arises late and reflects the exclusion of an articulatory response to the distractor word from an output buffer. Specifically, the mechanism of response exclusion is assumed to be sensitive to semantic information, such that if the articulatory response to the distractor shares semantic features with the picture name, the removal of the wrong articulatory program from the buffer is slowed, yielding the semantic interference effect. It would be expected from this late response-conflict view that the interference effect to incongruent color words relative to a color-neutral word not in the response set should be manifested as a negative delta slope: If the conflict was due to the motor response activated by the distractor word, this should decay over time, and hence the

effect should be smaller for the slowest responses. This is contradicted by the observed data. The delta plot shows clearly that the difference between the incongruent color name condition and the color-neutral word condition was present from the earliest quantile, and *increased* across the quantiles, both in the vocal task and the manual task. The RT distribution data thus argues against the response conflict view and instead indicates that the incongruent color words create greater difficulty for naming/categorizing the color than the color-neutral words because it reduces the rate of evidence accumulation, more specifically, the evidence pertaining to semantic features relevant to color.

Conclusion

Although there is a general consensus that Stroop interference reflects the automaticity of reading, which aspect of reading causes interference has not been well specified. The present study showed that a “wordlikeness gradient”—but with no difference between word and pseudoword distractors matched on bigram frequency—is found with the color-neutral distractors in the vocal Stroop task but not in the manual task, and that in the RT distribution analysis, all of the interference effects increased across the quantiles. Consistent with Monsell et al. (2001), we take the absence of lexicality effect as indicating that the Stroop interference effect is not caused by involuntary lexical access. The wordlikeness gradient—which would be better described as the pronounceability gradient—in the vocal task, but not in the manual task, is explained in terms of the close link between the alphabetic writing system and speech production processes in the former but not the latter: When the task goal is to name the color, the sublexically generated phonology interferes with the speech production process involved in naming the color, but not when the task is to categorize the color. In addition, in both the vocal task and the manual task, the interference effects observed with incongruent color words (e.g., “red” displayed in green) were much greater than that observed with color-neutral words (e.g., “hat”), but the latter did not produce more interference than pseudowords, which, by definition, do not have a meaning (e.g., “hix”). This result was taken to mean that in a Stroop task, it is not any semantic features but only the color-relevant semantic features of the word distractor that cause “semantic” interference. These findings converge to show that reading, even if it is triggered unintentionally, is not a purely stimulus-driven, invariant process; instead, the interference caused by the reading process in a Stroop task reflects the specific goal of the task required to the target.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research was supported by grants from the ARC Discovery project grant (DP 140101199) to Sachiko Kinoshita and Dennis Norris.

References

- Allport, D, Styles, EA, Hsieh, S. Shifting intentional set: Exploring the dynamic control of tasks. *Attention and performance XV: Conscious and unconscious information processing*. Umlita, C, Moscovitch, M, editors. Cambridge, MA: MIT Press; 1994. 421–452.
- Andrews S, Heathcote A. Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 2001; 27:514–544.
- Augustinova M, Ferrand L. Automaticity of word reading: Evidence from the semantic Stroop paradigm. *Current Directions in Psychological Science*. 2014; 23:343–348. DOI: 10.1177/0963721414540169
- Baayen, RH. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press; 2008.
- Bakan P, Alperson B. Pronounceability, attentivity, and interference in the color-word test. *The American Journal of Psychology*. 1967; 80:416–420. DOI: 10.2307/1420375
- Balota DA, Yap MJ. Moving beyond the mean in studies of mental chronometry: The power of response time distributional analysis. *Current Directions in Psychological Science*. 2011; 20:160–166.
- Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. 2013; 68:255–278. DOI: 10.1016/j.jml.2012.11.001
- Bates DM, Maechler M, Bolker B. *lme4: Linear mixed-effects models using Eigen and S4*. 2013; 2
- Bell D, Forster KI, Drake S. Early semantic activation in a semantic categorization task with masked primes: Cascaded or not? *Journal of Memory and Language*. 2015; 85:1–14. DOI: 10.1016/j.jml.2015.06.007
- Besner D. The myth of ballistic processing: Evidence from Stroop's paradigm. *Psychonomic Bulletin & Review*. 2001; 8:324–330. DOI: 10.3758/BF03196168 [PubMed: 11495121]
- Besner D, Stolz JA, Boutilier C. The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*. 1997; 4:221–225. DOI: 10.3758/BF03209396 [PubMed: 21331828]
- Binder JR, Medler DA, Westbury CF, Liebenthal E, Buchanan L. Tuning of the human left fusiform gyrus to sublexical orthographic structure. *NeuroImage*. 2006; 33:739–748. DOI: 10.1016/j.neuroimage.2006.06.053 [PubMed: 16956773]
- Brown TL. The relationship between Stroop interference and facilitation effects: Statistical artifacts, baselines, and a reassessment. *Journal of Experimental Psychology: Human Perception and Performance*. 2011; 37:85–99. DOI: 10.1037/a0019252 [PubMed: 20718567]
- Brysaert M, New B. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, Instruments & Computers*. 2009; 41:977–990. DOI: 10.3758/BRM.41.4.977
- Burt JS. Why do non-color words interfere with color naming? *Journal of Experimental Psychology: Human Perception and Performance*. 2002; 28:1019–1038. DOI: 10.1037/0096-1523.28.5.1019 [PubMed: 12421053]
- Cohen JD, Dunbar K, McClelland JL. On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*. 1990; 97:332–361. DOI: 10.1037/0033-295X.97.3.332 [PubMed: 2200075]
- Coltheart M, Woollams A, Kinoshita S, Perry C. A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. *Psychonomic Bulletin & Review*. 1999; 6:456–463. DOI: 10.3758/BF03210835 [PubMed: 12198784]
- Cousineau D, Brown S, Heathcote A. Fitting distributions using maximum likelihood: Methods and packages. *Behavior Research Methods, Instruments & Computers*. 2004; 36:742–756. DOI: 10.3758/BF03206555
- Dehaene S, Le Clec'H G, Poline JB, Le Bihan D, Cohen L. The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *NeuroReport: For Rapid Communication of Neuroscience Research*. 2002; 13:321–325. DOI: 10.1097/00001756-200203040-00015

- Dyer FN. The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*. 1973; 1:106–120. DOI: 10.3758/BF03198078 [PubMed: 24214501]
- Finkbeiner M, Caramazza A. Lexical selection is not a competitive process: A reply to La Heij et al. *Cortex*. 2006; 42:1032–1035.
- Forster KI, Davis C. The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*. 1991; 30:1–25. DOI: 10.1016/0749-596X(91)90008-8
- Forster KI, Forster JC. DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*. 2003; 35:116–124. DOI: 10.3758/BF03195503
- Forster KI, Hector J. Cascaded versus noncascaded models of lexical and semantic processing: The turtle effect. *Memory & Cognition*. 2002; 30:1106–1117. [PubMed: 12507375]
- Fox LA, Schor RE, Steinman RJ. Semantic gradients and interference in naming color, spatial direction, and numerosity. *Journal of Experimental Psychology*. 1971; 91:59–65. DOI: 10.1037/h0031850
- Goldfarb L, Henik A. Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology: Human Perception and Performance*. 2007; 33:1170–1176. DOI: 10.1037/0096-1523.33.5.1170 [PubMed: 17924815]
- Heathcote A, Popiel SJ, Mewhort D. Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*. 1991; 109:340–347. DOI: 10.1037/0033-2909.109.2.340
- Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*. 2008; 59:434–446. DOI: 10.1016/j.jml.2007.11.007 [PubMed: 19884961]
- Janssen N, Schirm W, Mahon BZ, Caramazza A. Semantic interference in a delayed naming task: Evidence for the response exclusion hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2008; 34:249–256.
- Kahneman, D, Henik, A. Perceptual organization and attention Perceptual organization. Kubovy, M, Pomerantz, JR, editors. Hillsdale, NJ: Erlbaum; 1981. 181–211.
- Keele SW. Attention demands of memory retrieval. *Journal of Experimental Psychology*. 1972; 93:245–248. DOI: 10.1037/h0032460 [PubMed: 5025731]
- Kinoshita, S. The nature of masked onset priming effects in naming: A review Masked priming: The state of the art. Kinoshita, S, Lupker, SJ, editors. New York, NY: Psychology Press; 2003. 223–240.
- Kinoshita, S; Aji, M. Stroop, priming and attentional control. Paper presented at the Annual meeting of the Psychonomic Society; Long Beach, CA. 2014 Nov.
- Kinoshita S, Hunt L. RT distribution analysis of category congruence effects with masked primes. *Memory & Cognition*. 2008; 36:1324–1334. DOI: 10.3758/MC.36.7.1324 [PubMed: 18927046]
- Kinoshita S, Mozer MC, Forster KI. Dynamic adaptation to history of trial difficulty explains the effect of congruency proportion on masked priming. *Journal of Experimental Psychology: General*. 2011; 140:622–636. [PubMed: 21707205]
- Kinoshita S, Norris D. Task-dependent masked priming effects in visual word recognition. *Frontiers in Psychology*. 2012; 3:178. doi: 10.3389/fpsyg2012.00178 [PubMed: 22675316]
- Klein GS. Semantic power measured through the interference of words with color-naming. *The American Journal of Psychology*. 1964; 77:576–588. DOI: 10.2307/1420768 [PubMed: 14255565]
- Kuznetsova, A; Brockhoff, PB; Christensen, RHB. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package; Version 2.0–11) [Computer software]. 2013. Retrieved from <http://CRAN.R-project.org/package=lmerTest>
- Labuschagne EM, Besner D. Automaticity revisited: When print doesn't activate semantics. *Frontiers in Psychology*. 2015; 6:117. [PubMed: 25713553]
- Lachter J, Forster KI, Ruthruff E. Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review*. 2004; 111:880–913. DOI: 10.1037/0033-295X.111.4.880 [PubMed: 15482066]
- Levelt WJ, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences*. 1999; 22:1–38. DOI: 10.1017/S0140525X99001776 [PubMed: 11301520]

- Lien MC, Ruthruff E, Kouchi S, Lachter J. Even frequent and expected words are not identified without spatial attention. *Attention, Perception, & Psychophysics*. 2010; 72:973–988. DOI: 10.3758/APP.72.4.973
- Lorentz E, McKibben T, Ekstrand C, Gould L, Anton K, Borowsky R. Disentangling genuine semantic Stroop effects in reading from contingency effects: On the need for two neutral baselines. *Frontiers in Psychology*. 2016; 7:386. [PubMed: 27014177]
- Lupker SJ, Katz AN. Input, decision, and response factors in picture-word interference. *Journal of Experimental Psychology: Human Learning and Memory*. 1981; 7:269–282. DOI: 10.1037/0278-7393.7.4.269
- MacLeod CM. Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*. 1991; 109:163–203. DOI: 10.1037/0033-2909.109.2.163 [PubMed: 2034749]
- Magen H, Cohen A. Action-based and vision-based selection of input: Two sources of control. *Psychological Research*. 2002; 66:247–259. DOI: 10.1007/s00426-002-0099-0 [PubMed: 12466923]
- Masson ME, Kliegl R. Modulation of additive and interactive effects in lexical decision by trial history. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2013; 39:898–914. DOI: 10.1037/a0029180
- Medler, DA; Binder, JR. MCWord: An on-line orthographic database of the English language. 2005. Retrieved from <http://www.neuro.mcw.edu/mcword/>
- Melara RD, Algom D. Driven by information: A tectonic theory of Stroop effects. *Psychological Review*. 2003; 110:422–471. DOI: 10.1037/0033-295X.110.3.422 [PubMed: 12885110]
- Monsell S, Taylor TJ, Murphy K. Naming the color of a word: Is it responses or task sets that compete? *Memory & Cognition*. 2001; 29:137–151. DOI: 10.3758/BF03195748 [PubMed: 11277457]
- Moors A, De Houwer J. Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*. 2006; 132:297–326. DOI: 10.1037/0033-2909.132.2.297 [PubMed: 16536645]
- Neely, JH, Kahan, TA. Is semantic activation automatic? The nature of remembering: Essays in honor of Robert G. Crowder. Roediger, HL, , III Nairne, JS, Neath, I, Surprenant, AM, editors. Washington, DC: American Psychological Association; 2001. 69–93. III
- Norris D. The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*. 2006; 113:327–357. [PubMed: 16637764]
- Norris D, Kinoshita S. Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*. 2008; 137:434–455. DOI: 10.1037/a0012799 [PubMed: 18729709]
- Pratte MS, Rouder JN, Morey RD, Feng C. Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophysics*. 2010; 72:2013–2025. DOI: 10.3758/APP.72.7.2013
- Price CJ. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*. 2012; 62:816–847. DOI: 10.1016/j.neuroimage.2012.04.062 [PubMed: 22584224]
- Price CJ, Wise RJS, Frackowiak RSJ. Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex*. 1996; 6:62–70. [PubMed: 8670639]
- Prinzmetal W, Hoffman H, Vest K. Automatic processes in word perception: An analysis from illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*. 1991; 17:902–923. DOI: 10.1037/0096-1523.17.4.902 [PubMed: 1837303]
- Ratcliff R. A theory of memory retrieval. *Psychological Review*. 1978; 85:59–108. DOI: 10.1037/0033-295X.85.2.59
- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. Retrieved from <http://www.R-project.org/>
- Risko EF, Schmidt JR, Besner D. Filling a gap in the semantic gradient: Color associates and response set effects in the Stroop task. *Psychonomic Bulletin & Review*. 2006; 13:310–315. DOI: 10.3758/BF03193849 [PubMed: 16893000]

- Risko EF, Stolz JA, Besner D. Basic processes in reading: Is visual word recognition obligatory? *Psychonomic Bulletin & Review*. 2005; 12:119–124. DOI: 10.3758/BF03196356 [PubMed: 15945205]
- Robidoux S, Rauwerda D, Besner D. Basic processes in reading aloud and colour naming: Towards a better understanding of the role of spatial attention. *Quarterly Journal of Experimental Psychology* (2006). 2014; 67:979–990. DOI: 10.1080/17470218.2013.838686 [PubMed: 24224499]
- Roelofs A. Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*. 2003; 110:88–125. DOI: 10.1037/0033-295X.110.1.88 [PubMed: 12529058]
- Rogers RD, Monsell S. The costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*. 1995; 124:207–231. DOI: 10.1037/0096-3445.124.2.207
- Schmidt JR, Crump MJC, Cheesman J, Besner D. Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition: An International Journal*. 2007; 16:421–435. DOI: 10.1016/j.concog.2006.06.010
- Sharma D, McKenna FP. Differential components of the manual and vocal Stroop tasks. *Memory & Cognition*. 1998; 26:1033–1040. DOI: 10.3758/BF03201181 [PubMed: 9796234]
- Spieler DH, Balota DA, Faust ME. Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance*. 2000; 26:506–526. DOI: 10.1037/0096-1523.26.2.506 [PubMed: 10811160]
- Steinhauser M, Hübner R. Distinguishing response conflict and task conflict in the Stroop task: Evidence from ex-Gaussian distribution analysis. *Journal of Experimental Psychology: Human Perception and Performance*. 2009; 35:1398–1412. DOI: 10.1037/a0016467 [PubMed: 19803645]
- Stroop JR. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*. 1935; 18:643–662. DOI: 10.1037/h0054651
- Sugg MJ, McDonald JE. Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*. 1994; 20:647–675. DOI: 10.1037/0096-1523.20.3.647 [PubMed: 8027716]
- Taylor JSH, Rastle K, Davis MH. Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychological Bulletin*. 2013; 139:766–791. DOI: 10.1037/a0030266 [PubMed: 23046391]
- Wagenmakers EJ, Brown S. On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*. 2007; 114:830–841. DOI: 10.1037/0033-295X.114.3.830 [PubMed: 17638508]
- Woollams AM, Silani G, Okada K, Patterson K, Price CJ. Word or word-like? Dissociating orthographic typicality from lexicality in the left occipito-temporal cortex. *Journal of Cognitive Neuroscience*. 2011; 23:992–1002. DOI: 10.1162/jocn.2010.21502 [PubMed: 20429854]

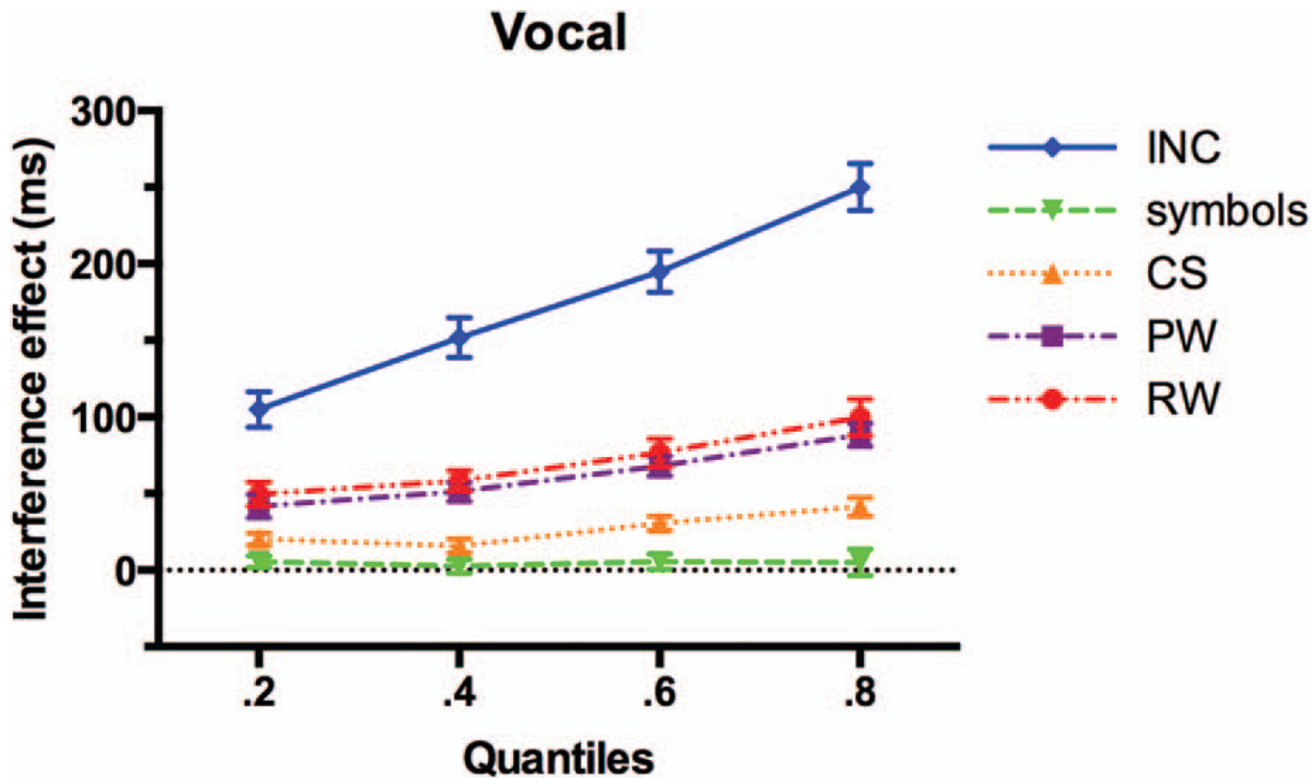


Figure 1.

Delta plots of Experiment 1 (vocal Stroop task). Interference effects in the vocal Stroop interference in Experiment 1, referenced to the row of Xs condition. The error bars are standard errors of the mean. INC = incongruent color word; symbols = nonalphabetic symbols; CS = consonant string; PW = pseudoword; RW = real word. See the online article for the color version of this figure.

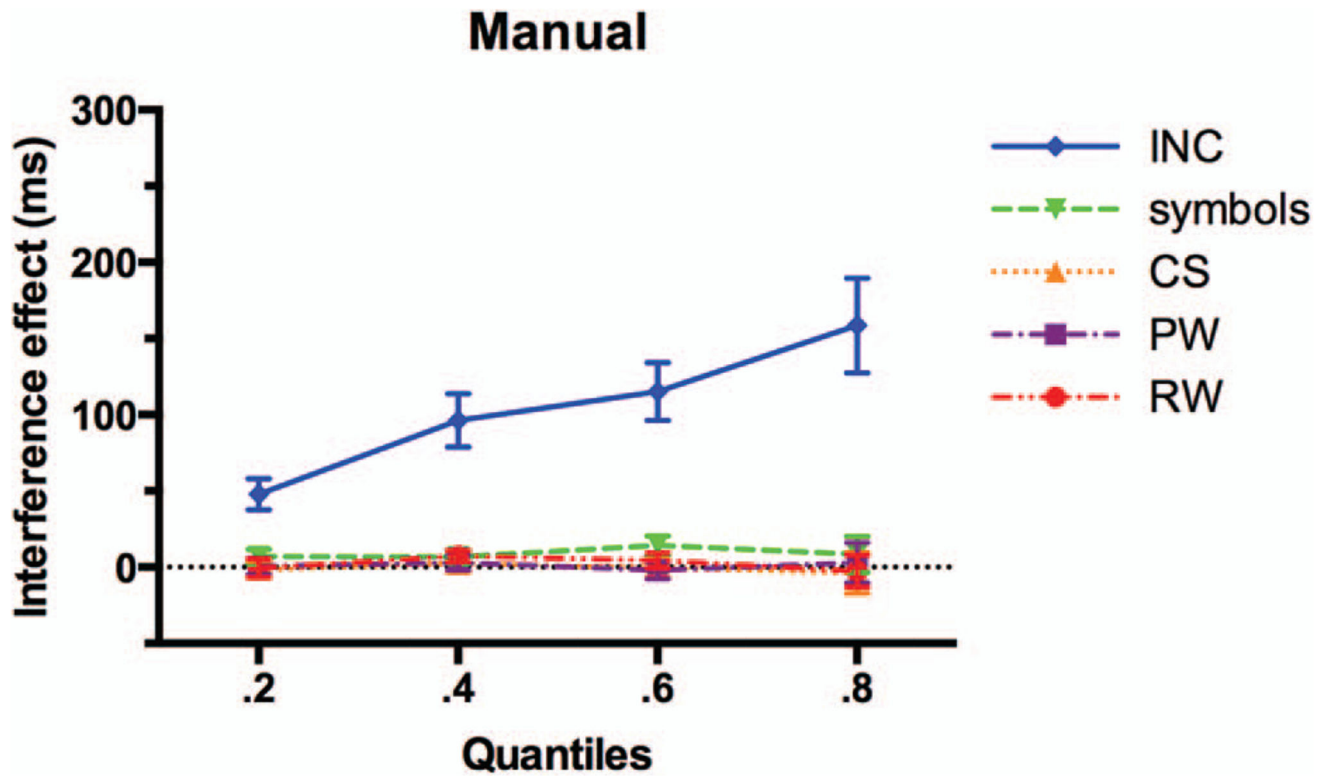


Figure 2.

Delta plots of Experiment 2 (manual Stroop task). Interference effects in the manual Stroop task in Experiment 2, referenced to the row of Xs condition. Error bars are standard errors of the mean. NC = incongruent color word; symbols = nonalphabetic symbols; CS = consonant string; PW = pseudoword; RW = real word. See the online article for the color version of this figure.

Table 1
Means and Standard Deviations (in Parentheses) Color Response Latencies (RT, in ms),
and Percent Error Rates in Experiment 1 (Vocal) and Experiment 2 (Manual)

Task type	String type					
	RW	PW	CS	Symbol	XXX	INC
Example	hat	hix	hdk	#\$%	XXX	red
Vocal	651 (87)	641 (75)	607 (73)	580 (68)	581 (74)	758 (94)
Manual	646 (69)	649 (77)	644 (71)	651 (67)	639 (77)	750 (120)
Error rate (%)						
Vocal	4.5 (3.9)	5.2 (3.6)	4.0 (3.7)	3.1 (4.1)	3.6 (3.0)	10.3 (8.2)
Manual	3.6 (4.4)	4.4 (4.9)	5.1 (4.3)	5.2 (5.4)	3.1 (3.0)	7.1 (7.6)

Note. RW = real word; PW = pseudoword; CS = consonant string; symbol = nonalphabetic symbols; XXX = a row of Xs; INC = incongruent color name.

Table 2
The Model's Estimate, Standard Error (Std. Error), Degrees of Freedom (df), *t* Value, and *p* Values of Fixed Effects for the Correct RT Data (Log RT) in Experiment 1

Task type	Estimate	Std. error	df	<i>t</i> value	Pr(> <i>t</i>)
(Intercept)	6.338103	.034763	60.25	182.324	<.001
stringtypeINC	.252791	.034516	99.18	7.324	<.001
stringtypeCS	.043912	.025390	99.03	1.729	.086836
stringtypePW	.096868	.025398	99.16	3.814	<.001
stringtypeRW	.110394	.025393	99.08	4.347	<.001
stringtypesymbol	.002812	.025387	98.99	.111	.912027

Note. String type factor was referenced to the row of Xs. *df* = degrees of freedom; INC = incongruent color name; CS = consonant string; PW = pseudoword; RW = real word; symbol = nonalphabetic symbols.

Table 3
Length Effect for Color-Neutral Strings in Experiment 1 (Vocal) and Experiment 2 (Manual)

Task type	String length (number of letters)			
	3	4	5	6
Vocal				
RW	644	643	640	668
PW	633	632	642	658
CS	619	596	606	608
symbol	589	563	586	577
XXX	583	577	578	584
Manual				
RW	658	638	653	640
PW	646	650	652	650
CS	650	637	643	650
symbol	666	653	649	642
XXX	647	638	638	639
PRON-Vocal	639	638	641	664
PRON-Manual	652	644	652	645

Note. RW = real word; PW = pseudoword; CS = consonant string; symbol = nonalphabetic symbols; XXX = a row of Xs; PRON = pronounceable strings (average of RW and PW).

Table 4
The Model's Estimate, Standard Error (Std. Error), z Value, and p Values of Fixed Effects for the Error Data in Experiment 1

Task type	Estimate	Std. error	z value	Pr(> z)
(Intercept)	-3.6067	.2638	-13.673	<.001
stringtypeINC	1.1653	.2373	4.91	<.001
stringtypeCS	.1158	.2225	.52	.6028
stringtypePW	.3954	.2146	1.843	.0654
stringtypeRW	.2419	.2187	1.106	.2687
stringtypesymbols	-.1453	.2318	-.627	.5306

Note. String type factor was referenced to the row of Xs. INC = incongruent color name; CS = consonant string; PW = pseudoword; RW = real word; symbol = nonalphabetic symbols.

Table 5
The Model's Estimate, Standard Error (Std. Error), Degrees of Freedom (df), *t* Value, and *p* Values of Fixed Effects for the Correct RT Data (LogRT) in Experiment 2

Task type	Estimate	Std. error	df	<i>t</i> value	Pr(> <i>t</i>)
(Intercept)	6.423	.0247	22	260.061	<.001
stringtypeINC	.1351	.01033	8744	13.075	<.001
stringtypeCS	.005102	.009498	8744	.537	.591
stringtypePW	.009977	.009481	8744	1.052	.293
stringtypeRW	.009575	.009461	8744	1.012	.312
stringtypesymbol	.017730	.009500	8744	1.867	.062

Note. String type factor was referenced to the row of Xs. *df* = degrees of freedom; INC = incongruent color name; CS = consonant string; PW = pseudoword; RW = real word; symbol = nonalphabetic symbols.

Table 6
The Model's Estimate, Standard Error (Std. Error), z Value, and p Values of Fixed Effects for the Error Data in Experiment 2

Task type	Estimate	Std. error	z value	Pr(> z)
(Intercept)	-3.7617	.2437	-15.433	<.0001
stringtypeINC	.8878	.1838	4.831	<.0001
stringtypeCS	.5162	.1843	2.802	<.001
stringtypePW	.3734	.1888	1.977	<.05
stringtypeRW	.1571	.1968	.798	.4249
stringtypesymbol	.5428	.1835	2.959	<.01

Note. String type factor was referenced to the row of Xs. INC = incongruent color name; CS = consonant string; PW = pseudoword; RW = real word; symbol = nonalphabetic symbols.