Acta Psychologica 205 (2020) 103054

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Visual search inverts the classic Stroop asymmetry

Kenith V. Sobel[a,*], Amrita M. Puri[b], A. Kane York[c]

[a] *Department of Psychology and Counseling, University of Central Arkansas, United States of America*
[b] *Department of Biology, University of Central Arkansas, USA*
[c] *Neuroscience Graduate Program, University of Michigan, USA*

ARTICLE INFO

Keywords:
Visual search
Stroop effect
Reverse Stroop effect
Stroop asymmetry
Attention

ABSTRACT

The Stroop effect is typically much larger than the reverse Stroop effect. One explanation for this asymmetry asserts that interference between the attended feature and an incongruent unattended feature depends on which feature is more strongly associated with the processing typically needed to complete the task. Accordingly, because identification of the target's color or the target word (as in the traditional Stroop paradigm) is more strongly associated with verbal processing than visual processing, the target's meaning should interfere with identification of the target's color (Stroop) more than vice versa (reverse Stroop). In contrast, localization is more strongly associated with visual processing, so strength-of-association predicts that the target's color should interfere with localizing the target word (reverse Stroop) more than vice versa (Stroop). Experiments 1 and 2 supported the strength-of-association account: compared to Stroop, the reverse Stroop effect was smaller for an identification task, but larger for a localization task. Because overall responses were slower for the reverse Stroop condition than the Stroop condition in Experiment 2, we entertained two alternative explanations for the reverse Stroop effect being larger than the Stroop effect. Experiments 3 and 4 showed that the larger reverse Stroop effect could not have been due to scaling, and Experiment 5 showed that it could not have been due to covert translation. Taken together, these experiments demonstrate the role of strength of association in generating the classic Stroop asymmetry, and pave the way for future exploration of the reverse Stroop effect using localization tasks.

1. Visual search inverts the classic Stroop asymmetry

As Stroop (1935) demonstrated in his most famous experiment, reporting the color of a word for which the meaning is incongruent (e.g., the word *Red* written in blue) takes more time than reporting the color of abstract shapes. In another of his experiments, reporting the word rather than the ink color was no slower for incongruent words than for words written in a neutral color (black). This is the classic Stroop asymmetry: An incongruent word typically interferes with color naming much more than an incongruent color interferes with word reading, resulting in larger Stroop effects than reverse Stroop effects (Melara & Algom, 2003). There has been a lively debate as to why the classic Stroop asymmetry occurs. In the next sections, we briefly review this body of work, noting that there are three factors that seem to influence the relative size of Stroop and reverse Stroop effects. We then develop a hypothesis based on the third account, and describe a new set of tasks that aims to test this hypothesis.

1.1. Discriminability

According to one account, in the typical Stroop experiment the target word is presumed to be more discriminable than its color, so the more discriminable feature (word meaning) interferes with processing the less discriminable feature (naming the color) more than the less discriminable feature (color) interferes with processing the more discriminable feature (word reading). Accordingly, reducing the discriminability of the target word by presenting it in a tiny (Melara & Mounts, 1993) or an upside-down and backwards (Dunbar & MacLeod, 1984) font makes the target word less discriminable than its color, which enables the color to interfere with word reading: a reverse Stroop effect.

1.2. Translation

The translation account is based on the premise that visual and verbal information are encoded and processed in separate systems (Song & Hakoda, 2015; Virzi & Egeth, 1985). The traditional Stroop

* Corresponding author at: Department of Psychology and Counseling, University of Central Arkansas, 201 Donaghey Ave., Conway, AR 72035, USA.
E-mail address: ksobel@uca.edu (K.V. Sobel).

https://doi.org/10.1016/j.actpsy.2020.103054
Received 7 August 2019; Received in revised form 20 February 2020; Accepted 2 March 2020
0001-6918/

condition entails vocal identification of the target's color, so color must be translated from a visual code into a verbal code to generate the vocal response. However, in the traditional reverse Stroop condition, the attended feature (word meaning) and response (vocal identification) both rely on a verbal code, so no translation is required. Thus, the traditional Stroop condition requires the attended feature (color) to be translated into a verbal code, which allows the irrelevant dimension (word meaning) to interfere with naming the color, but the traditional reverse Stroop condition does not require the attended feature (word meaning) to be translated because it is already encoded verbally, so the irrelevant dimension (color) has no opportunity to interfere with reading the word. The translation theory implies that recasting the traditional task such that translation is *not* required in the Stroop condition but *is* required in the reverse Stroop condition should invert the classic asymmetry: the reverse Stroop effect should be larger than the Stroop effect.

Durgin (2000; replicated by Miller, Kubicki, Caffier, Kolski, & Naveteur, 2016) developed a matching task to test this implication of the translation theory. In each trial, a word cue appeared in a color that was either congruent or incongruent with its meaning, surrounded by four color patches. Participants were instructed to attend to either the cue's color (Stroop) or meaning (reverse Stroop), and in both conditions to select the patch that had the matching color. Thus, Durgin's Stroop condition required no translation to match the cue's color to the target's color, but his reverse Stroop condition *did* require translation to match the cue's meaning with the target's color. Consistent with the translation account, interference was greater when translation was required in the reverse Stroop condition than when no translation was required in the Stroop condition.

1.3. Strength of association

The strength-of-association account (Blais & Besner, 2006) emphasizes the compatibility between the attended feature and the task, such that the feature that is more strongly associated with a given task interferes with the processing of a weakly associated feature more than vice versa. Because Durgin's matching task required visual processing to localize the target, Blais and Besner (2007) argued that Durgin's matching task should have been sufficient to induce an advantage for the visual feature (color) over the verbal feature (word), leading to a reverse Stroop effect even if translation had not been required. To test this prediction, Blais and Besner (2007; and a replication by Yamamoto, Incera, & McLennan, 2016) eliminated the need for translation in Durgin's procedure by replacing the four colored patches with words printed in a neutral color, and instructed participants to select the target word that matched the cue word. Even though no translation was required between the cue word and the target word, incongruent cue color interfered with matching, resulting in a reverse Stroop effect.

These results show that translation is not a necessary condition for reverse Stroop interference, and yet they do not imply that the translation and strength-of-association accounts are mutually exclusive. Indeed, unless the experimenter makes an explicit effort to control for the roles of discriminability, translation, and strength of association, the three factors could very well act in concert to influence the relative sizes of Stroop and reverse Stroop effects. Moreover, an additional factor that could possibly affect the relative sizes of Stroop and reverse Stroop effects is response modality. Grégoire, Poulin-Charronnat, and Perruchet (2019) noted that recent studies that have revealed a reverse Stroop effect without manipulating the discriminability of the target words required participants to provide a manual response (Blais & Besner, 2006, 2007; Durgin, 2000; Miller et al., 2016; Song & Hakoda, 2015; Yamamoto, Incera, & McLennan, 2016), whereas traditional Stroop tasks require a vocal response.

Although Grégoire et al. (2019) declined to speculate about why response modality (i.e., manual versus vocal) would influence the relative sizes of Stroop and reverse Stroop effects, one possibility is that the typical manual responses elicited in these experiments entail more visuospatial engagement than a vocal response would. That is, some experiments require participants to move a mouse cursor or other object towards a target location, and other experiments require participants to respond by pressing different keys, with each response key having a unique spatial location. Thus, similar to the strength-of-association account, the response modality account implies that vocal responses are more strongly associated with verbal than visual processing, whereas manual responses are more strongly associated with visual than verbal processing. That could explain why a vocal response induces more interference from incongruent word meaning (i.e., Stroop effect) than a manual response does (Kinoshita, Mills, & Norris, 2018).

Here we aimed to find evidence supporting a key prediction of the strength-of-association account: For a task that is more strongly associated with verbal processing than visual processing, the Stroop effect should be larger than the reverse Stroop effect, but for a task that is more strongly associated with visual processing than verbal processing, the Stroop effect should be smaller than the reverse Stroop effect. At the same time, we aimed to control for discriminability, translation, and response modality. To do so, we manipulated the task such that one was more strongly associated with verbal processing and the other with visual processing. Both tasks required manual keypress responses. The first task instructed participants to identify either the target's color (Stroop) or meaning (reverse Stroop). A larger Stroop effect than the reverse in this identification task would be consistent with: discriminability, because the target word could be more discriminable than its color; translation, because identification entails translation of the target's color in the Stroop condition, but not its meaning in the reverse Stroop condition, into a verbal code; and strength of association, because attending to a color (Stroop) is weakly associated with the verbal processing required for identification, but attending to a word (reverse Stroop) is strongly associated with verbal processing. However, a larger Stroop effect than the reverse Stroop effect would be inconsistent with the response modality account, because the visuospatial content of a manual keypress response should induce less interference when participants attend to a color (Stroop) than when they attend to a word (reverse Stroop). Whereas this first task cannot distinguish between discriminability, translation, and strength of association, it does distinguish between all three on one hand and the response modality account on the other, and lays the foundation for a second task that does distinguish between the three.

The second task used the same font size and colors as the first task condition, but required visual processing (as in Durgin, 2000), and eliminated the need for translation (as in Blais & Besner, 2007). By holding the target stimuli constant across both the identification and visual tasks, the relative discriminability of the words and colors was the same in both tasks, so the discriminability account predicts the same outcome for the visual task as the identification task: the Stroop effect should be larger than the reverse Stroop effect. Because neither the Stroop nor the reverse Stroop condition in the visual task required translation, the translation account predicts there should be no difference between the Stroop and reverse Stroop effects in the visual task. Also, because a task requiring visual processing is more strongly associated with attending to a color (Stroop) than attending to a word (reverse Stroop), strength of association predicts that the Stroop effect should be smaller than the reverse Stroop effect. And finally, because the visual task elicited a manual keypress response as in the identification task, the response modality account makes the same prediction for the visual task as for the identification task: the Stroop effect should be smaller than the reverse Stroop effect. The predictions of each account are summarized in Table 1. Next, we describe a new task designed to test the predictions of the strength-of-association account while controlling for the other factors.

1.4. Design of a new visual task

Although the reverse Stroop effect is relatively fragile (Dunbar &

Table 1
Predictions of each hypothesis.

	Task processing	
	Experiment 1: Verbal	Experiment 2: Visual
Hypothesis		
Discriminability	Stroop > reverse Stroop	Stroop > reverse Stroop
Translation	Stroop > reverse Stroop	Stroop = reverse Stroop
Strength-of-association	Stroop > reverse Stroop	Stroop < reverse Stroop
Response modality	Stroop < reverse Stroop	Stroop < reverse Stroop

MacLeod, 1984), the reverse Stroop effect has been shown to be robust in matching tasks (Blais & Besner, 2007; Durgin, 2000; Miller et al., 2015; Song & Hakoda, 2015; Yamamoto et al., 2016). While the methods varied widely across these studies, in each, participants viewed a display containing a cue word in a color that was either congruent or incongruent with the cue's meaning, along with multiple potential targets that were either words written in a neutral color or color patches. Participants attended to either the cue's meaning or color, and selected the target that matched the attended feature. Because all these matching tasks presented more than one possible target, they all required participants to search through the possibilities in order to select the target.

And yet, the fact that incongruity between the cue's meaning and color can affect visual search performance is surprising; whereas color is undoubtedly a guiding feature in visual search, an item's semantic associations are typically presumed not to guide search (Wolfe & Horowitz, 2004, 2017). Nevertheless, recent studies have shown that in a visual search for a numerical target, incongruity between numerical size and physical size (i.e., numerical Stroop; Dadon & Henik, 2017) can influence visual search performance (Krause, Bekkering, Pratt, & Lindemann, 2017; Sobel, Puri, & Faulkenberry, 2016). To explain how sensory-semantic incongruity can influence visual search performance even though semantic associations are probably not guiding features for visual search, the target's semantic feature (numerical size) is presumed not to influence the selection stage, during which the search items' visual feature (physical size) guides attention to select one item for further processing. Then, only after an item has been selected on the basis of its visual features, sensory-semantic incongruity influences the decision stage (Risko, Maloney, & Fugelsang, 2013; Sobel & Puri, 2018). To extend on experiments that have found a numerical Stroop effect in visual search (Krause et al., 2017; Sobel et al., 2016), here we sought to develop a visual search experiment in which incongruity between a target's meaning and color has the opportunity to influence the decision stage of processing. Next, we describe the rationale underlying each of our experimental design decisions.

1.4.1. Word-color Stroop stimuli as search items rather than cues

Given that in numerical Stroop, the sensory-semantic incongruity affects the decision stage only after a search item has been selected for further processing, we wanted to instill incongruity in the search items rather than the cues. Doing so switches the roles of cues and search items from the previously cited matching experiments (Blais & Besner, 2007; Durgin, 2000; Miller et al., 2015; Song & Hakoda, 2015; Yamamoto et al., 2016); in these experiments, the cues were words presented in congruent or incongruent colors, and search items were unambiguously sensory (i.e., color patches) or semantic (i.e., words written in neutral colors). In our visual search experiment, the cues were color patches or words in neutral colors, and the search items were words presented in color.

Two pieces of evidence suggest that using words presented in color as search items can induce a larger reverse Stroop effect than the Stroop effect. The first piece of evidence comes from studies looking at the numerical Stroop effect in visual search. When participants searched for a number that was a conjunction of numerical and physical size, the target's visual feature (physical size) interfered with search for a target's semantic feature (numerical size: reverse Stroop) more than the target's meaning (numerical size) interfered with searching for target's visual feature (physical size: Stroop), so the reverse Stroop effect was larger than the Stroop effect (Sobel et al., 2016; Sobel & Puri, 2018; Wilson & Sobel, 2018). Second, one previous study cued participants with either a word or color, then instructed them to scan a card containing numerous word-color Stroop stimuli, and to write a checkmark next to the word-color items that matched the cue (Uleman & Reeves, 1971). Because this study was carried out before computers were widely available to measure trial-by-trial response times, the dependent variable was the overall time required to scan the card. The overall response time (RT) difference between incongruent and neutral cards was larger when participants were cued to attend to the word (reverse Stroop) than when they were cued to attend to color (Stroop).

1.4.2. No translation needed

Many of the cited matching tasks required translation in the reverse Stroop condition (Durgin, 2000; Miller et al., 2015; Song & Hakoda, 2015). No translation was required in Blais and Besner (2007) and Yamamoto et al. (2016), because participants were instructed to attend to the cue word, and to select the target word that matched it. We eliminated the need for translation between cue and search item by using cue features that matched the target features. Thus, in the Stroop condition, participants were presented with color cues and selected the search item with the same color, while in the reverse Stroop condition, participants were presented with word cues and selected the search item with the same meaning.

1.4.3. Localization

In the typical visual search experiment (reviewed in Wolfe, 1998), participants view displays that may or may not contain a target among several nontarget distractors, and make a binary decision about the target. Experiments typically require one of three commonly used decisions: Participants may be asked to detect the target by reporting whether it is present in, or absent from, the display; to identify the target by reporting which of two visual features it contains; or to localize the target by reporting which of two regions it occupies (Dukewich & Klein, 2009). Of these three widely used visual search tasks, localization is the one that is most closely analogous with matching tasks. Thus, we asked participants to localize the target to either the right or left side of the display and report via keypress which side of the display contained the target.

1.4.4. Congruent or neutral

To achieve our primary goal of revealing an interaction between attended feature (Stroop versus reverse Stroop) and task processing (verbal versus visual), we wanted to manipulate the attended feature and task processing conditions while using the same display items in both the Stroop and reverse Stroop conditions. This presents no problem for incongruent display items, which are the same in the Stroop and reverse Stroop conditions (e.g., *Red* appearing in blue is incongruent for both the Stroop and reverse Stroop conditions). However, neutral display items entail different stimuli for Stroop and reverse Stroop conditions (e.g., colored blocks for the Stroop condition, and words in a neutral color for the reverse Stroop condition), whereas congruent display items are the same in Stroop and reverse Stroop conditions (e.g., *Red* appearing in red). Thus, we defined Stroop and reverse Stroop effects as the difference between the RTs elicited by incongruent targets and the RTs elicited by and congruent targets.

1.5. Summary of experimental design

In Experiments 1 and 2, we aimed to find evidence that supports the strength-of-association hypothesis that could not also be explained in terms of the discriminability, translation, or response modality accounts. In Experiment 1, participants were presented with one word-color Stroop target in each trial, and they were instructed to identify either the target's color (Stroop) or meaning (reverse Stroop). Experiments 1 and 2 manipulated the task, and thus the processing required for the task. Experiment 1 used an identification task, which is more strongly associated with verbal than visual processing, whereas Experiment 2 used a localization task, which is more strongly associated with visual than verbal processing. While manipulating the task, we controlled discriminability by using the same font size and colors in both experiments. Also, in Experiment 2 the cued feature and attended feature relied on the same codes (visual for the Stroop condition, verbal for the reverse Stroop condition), thereby eliminating the need for translation. Both experiments elicited a manual keypress response from participants. Thus, the crucial result supporting strength-of-processing would be an inversion of asymmetry between tasks: Stroop > reverse Stroop in identification (Experiment 1), but Stroop < reverse Stroop in localization (Experiment 2).

2. Experiment 1: Identification

2.1. Method

2.1.1. Participants

The University of Central Arkansas Institutional Review Board approved all experimental procedures, and we treated participants in accordance with the ethical guidelines stipulated by the American Psychological Association (2017). In light of recent studies that have revealed a numerical Stroop effect in visual search (Krause et al., 2017; Sobel et al., 2016), an effect with a similarly large $d = 1.25$ would require a minimum of 14 participants to achieve 80% power at an alpha of 0.05 (Bausell & Li, 2002). A total of 16 students (one male, 15 female) between 19 and 29 (M = 21.5) volunteered for Experiment 1 in exchange for course credit.

2.1.2. Apparatus

All experiments were conducted on a MacBook computer and CRT monitor with a screen resolution of 1024 × 768 pixels. Programs written in Xojo Basic presented stimulus arrays to the monitor and gathered responses from the keyboard.

2.1.3. Procedure

The experiment began by presenting instructions that participants could read at their own pace. After reading the instructions, participants viewed a series of displays that each contained one word from the following set: *Red*, *Green*, *Blue*, and *Yellow*. All words appeared in 64-point Helvetica font so from a viewing distance of 60 cm., each letter was about 1.3° tall × 1.0° wide. We used the same RGB settings as in Blais and Besner (2007): (255, 0, 0) for red, (0, 255, 0) for green, (0, 0, 255) for blue, and (255, 255, 0) for yellow. Participants were instructed to identify the target's color during one half of the experiment and its meaning during the other half. Block order was counterbalanced across participants.

For incongruent trials, the target word appeared in one of the three colors that was incongruent with its meaning. In each block of trials (i.e., identify the target's color in the Stroop block or its meaning in the reverse Stroop block), incongruent trials included four repetitions of every combination of four words and three incongruent colors for a total of 48 trials, and congruent trials included 12 repetitions of each of four words for a total of 48 trials. Besides the blocking of attended feature, all other variables were randomly interleaved so each block contained a total of 96 trials.

Each trial began with the presentation of a white fixation cross consisting of two orthogonal line segments 1° long in the middle of a black background. After 750 ms, the cross was replaced by a target word. Participants reported either the target's color or meaning (depending on the attended feature condition) by pressing one of four keys:

Table 2
Mean error rates (percent).

	Attended feature	
	Color (Stroop)	Word (reverse Stroop)
Experiment 1		
Congruent	2.47	5.34
Incongruent	6.77	6.64
Experiment 2		
Congruent	2.08	2.21
Incongruent	3.52	3.91
Experiment 3		
Congruent	4.42	4.03
Incongruent	3.91	5.47
Experiment 4		
Congruent	3.51	1.95
Incongruent	4.43	5.86
Experiment 5		
Congruent	2.86	2.60
Incongruent	3.08	4.30

d for red, *f* for green, *j* for blue, and *k* for yellow. Participants were instructed to place their hands in the "home" position on the keyboard and use one of four fingers to report each response. After a correct response, the target was replaced by the fixation cross to begin the next trial. After an error, the word "Incorrect" appeared for 750 ms before being replaced by the fixation cross. The time between the onset of the target word and the keypress was recorded for each trial. At the end of the first block, participants were invited to take a short break, and reminded that the attended feature would switch for the remainder of the experiment. The first six trials overall and the first six trials after the break were practice, so each participant carried out 204 (12 practice +192 experimental) trials.

2.2. Results

Mean error rates were submitted to a three-way Analysis of Variance (ANOVA) with congruence and attended feature as within-subjects factors, and block order as a between-subjects factor. None of the effects with block order as a factor were significant, all *ps* > .3, so the error rates in Table 2 are collapsed across both levels of block order. Error rates were higher in incongruent trials than congruent, $F(1, 14) = 8.57$, $p = .011$, $\eta_p^2 = 0.38$, but as can be seen in Fig. 1, responses were also slower in incongruent trials, so there is no evidence of

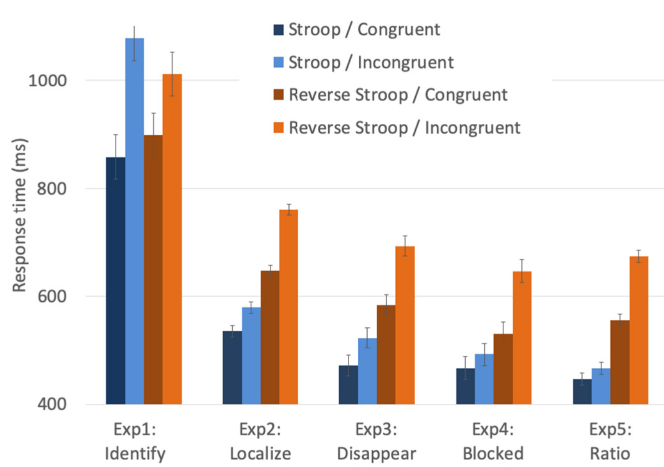

Fig. 1. Mean correct RTs in all experiments. Experiments 2 through 5 were all localization tasks. Error bars represent 95% confidence intervals (Loftus & Masson, 1994).

a speed/accuracy tradeoff. The main effect of attended feature and its interaction with congruence were not significant, both $ps > .1$.

For each participant in each of four conditions (two levels of congruence × two attended features), a trimming program removed all RTs that were more than three standard deviations away from the mean for that participant and condition or < 100 ms; a total of 2.0% of data points were removed. Mean correct RTs were submitted to a three-way ANOVA with congruence and attended feature as within-subjects factors, and block order as a between-subjects factor. None of the effects with block order as a factor were significant, all $ps > .2$, so the RTs in Fig. 1 are collapsed across both levels of block order.

Responses were faster for congruent than incongruent targets, $F(1, 14) = 70.35$, $p < .001$, $\eta_p^2 = 0.83$. The main effect of attended feature was not significant, $p = .84$, but the interaction between congruence and attended feature, $F(1, 14) = 5.17$, $p = .039$, $\eta_p^2 = 0.27$, shows that the Stroop effect was different than the reverse Stroop effect. Simple effects analysis confirmed that both the Stroop effect, $F(1, 15) = 43.25$, $p < .001$, $\eta_p^2 = 0.74$, and the reverse Stroop effect, $F(1, 15) = 18.03$, $p = .001$, $\eta_p^2 = 0.55$, were significant, and the Stroop effect size (Mean difference ± 95% confidence interval: 219.91 ± 41.07 ms, $\eta_p^2 = 0.74$) was larger than the reverse Stroop effect size (113.24 ± 41.07 ms, $\eta_p^2 = 0.55$).

2.3. Discussion

The Stroop effect was larger than the reverse Stroop effect for an identification task in which participants reported the target's color or meaning by pressing keys. These results are consistent with the discriminability (Dunbar & MacLeod, 1984; Melara & Mounts, 1993), translation (Sugg & McDonald, 1994; Virzi & Egeth, 1985), and strength-of-association (Blais & Besner, 2006, 2007) accounts. The fact that all three converge on the same outcome could be the reason that the classic Stroop asymmetry has historically been so robust.

While our results were consistent with these three accounts, they were inconsistent with the response modality account. There are a few possible reasons for this inconsistency. First, although the keypress responses entailed four different visuospatial locations, participants may have covertly assigned a verbal label to each response key so they could remember what each represented (Blais & Besner, 2006; Sugg & McDonald, 1994). Accordingly, identification of a target's meaning or color would be verbally mediated even though a manual keypress response was required. Second, if discriminability, translation, and strength of association additively combine to induce a larger Stroop effect than the reverse Stroop effect, their combined influence might overwhelm any effects associated with the visuospatial associations of each manual response. And third, while Grégoire et al. (2019) observed that recent studies revealing a reverse Stroop effect have elicited a manual response, the correlation between reverse Stroop effects and manual responses could be spurious. To distinguish between these possibilities will require experiments in which the response modality is manipulated while all other factors are held constant.

While the results from Experiment 1 are consistent with the discriminability, translation, strength-of-association accounts, the crucial test of the strength-of-association account was Experiment 2. The same font size and colors were used in Experiment 2 as in Experiment 1, and the need for translation was eliminated for both attended features in Experiment 2. At the same time, the task was manipulated between experiments, from identification in Experiment 1 to localization in Experiment 2.

3. Experiment 2: Localization

3.1. Method

3.1.1. Participants

A total of 16 students (nine male, seven female) between 19 and 24 (M = 20.1) volunteered for Experiment 2 in exchange for course credit. None had participated in Experiment 1.

3.1.2. Procedure

Participants were informed that a color cue would indicate the target's color in one half of the experiment, and a word cue would indicate the target's meaning in the other half; block order was counterbalanced across participants. At the beginning of each trial, a cue appeared in the middle of a black background. In one block the cue was a square 2.0° on a side in one of four colors (red, green, blue, or yellow), and in the other block the cue was one of four color words presented in white. After the cue was visible for 750 ms, four search items appeared and remained visible along with the cue until participants responded. All four words and colors appeared once in each search array. The centers of mass of the four search items were evenly distributed on an imaginary circle with a radius of 6.3°, centered on the cue. The target appeared in a randomly determined location in one of four quadrants. To ensure that the target's position was readily distinguishable from the vertical meridian, its center of mass was placed at least 30° of arc away from vertical (i.e., between 1 o'clock and 5 o'clock on the right side of the screen, and between 7 o'clock and 11 o'clock on the left side). After the search display appeared, participants reported the target's location by pressing the "/" key with their right hands if it appeared on the right side of the screen, or the "z" key with their left hands if it appeared on the left side. Errors were followed by a screen with the word "Incorrect" for 750 ms, then the cue for the next trial. Screenshots of visual displays appear in Fig. 2.

There are nine possible ways of assigning the four colors to four words such that the four words are all simultaneously incongruent with their color. We wanted to use three incongruent word-to-color mappings as in Experiment 1. To reduce the nine possible permutations to three, we envisioned two concentric circles, with the inner circle representing colors and the outer circle representing words. When red appears at the 12 o'clock position, green at 3 o'clock, blue at 6 o'clock,

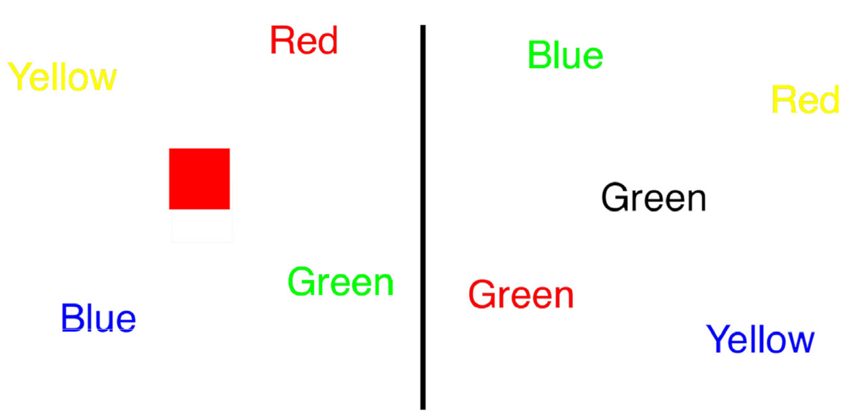

Fig. 2. Cues and search arrays for Experiment 2. In Experiments 3, 4, and 5, the cue was replaced with a fixation cross before the search arrays appeared. The left panel depicts a color cue (Stroop condition) with congruent search items, and the right panel depicts a word cue (reverse Stroop condition) with incongruent search items. All items were presented against a black background so word cues appeared in white, but the word cue in the right panel has been changed to black so it can be seen against a white background.

and yellow at 9 o'clock on both circles, the words and colors are all congruent. The inner circle can be rotated relative to the outer circle by 90°, 180°, or 270°. A counterclockwise rotation of the inner circle by 90° results in the word *Red* appearing in green, *Green* in blue, *Blue* in yellow, and *Yellow* in red. Rotations of 180° or 270° each yield a different set of incongruent mappings. Thus, each offset between the inner and outer circle represents one mapping of words to colors, with one congruent mapping (i.e., offset = 0°), and three incongruent mappings (i.e., offset = 90°, 180°, or 270°). These three incongruent word-color mappings were the ones we used for incongruent displays.

In each block, the incongruent trials included every combination of four target quadrants, four target words, and three offsets for a total of 48 trials, and congruent trials included three repetitions of every combination of four target quadrants and four target words for a total of 48 trials. Besides the blocking of attended feature, all other variables were randomly interleaved. At the end of the first block, participants were invited to take a short break and reminded that for the remainder of the experiment the cued target feature (word or color) would switch. As in Experiment 1, all participants carried out 12 practice and 192 experimental trials.

3.2. Results

Mean error rates were submitted to a three-way ANOVA with congruence and attended feature as within-subjects factors, and block order as a between-subjects factor. None of the main effects or interactions were significant, all *p*s > .1. The error rates in Table 2 are collapsed across both levels of block order.

The same RT trimming routine used in Experiment 1 removed a total of 2.1% of data points. Mean correct RTs were submitted to a three-way ANOVA with congruence and attended feature as within-subjects factors, and block order as a between-subjects factor. None of the effects with block order as a factor were significant, all *p*s > .5, so the RTs in Fig. 1 are collapsed across both levels of block order.

Responses were faster for congruent targets than incongruent, $F(1, 14) = 53.94$, $p < .001$, $\eta_p^2 = 0.79$, and were faster when participants attended to the target's color (Stroop) than its meaning (reverse Stroop), $F(1, 14) = 23.00$, $p < .001$, $\eta_p^2 = 0.62$. The interaction between congruence and attended feature, $F(1, 14) = 34.06$, $p < .001$, $\eta_p^2 = 0.71$, shows that the Stroop effect was different than the reverse Stroop effect. Simple effects analysis confirmed that both the Stroop effect, $F(1, 15) = 19.47$, $p = .001$, $\eta_p^2 = 0.56$, and the reverse Stroop effect, $F(1, 15) = 70.16$, $p < .001$, $\eta_p^2 = 0.82$, were significant, and that the Stroop effect size (43.72 ± 10.51 ms, $\eta_p^2 = 0.56$) was smaller than the reverse Stroop effect size (113.87 ± 10.51 ms, $\eta_p^2 = 0.82$), which is the inverse of Experiment 1, in which the Stroop effect was larger than the reverse Stroop effect.

The inversion of the asymmetry between Experiments 1 and 2 provides crucial support for our hypothesis that for a task requiring verbal processing, the Stroop effect should be larger than the reverse Stroop effect, but for a task requiring visual processing, the Stroop effect should be smaller than the reverse Stroop effect. In turn, the inverted two-way interactions between experiments imply that the three-way interaction between congruence, attended feature, and experiment should be significant. To examine this three-way interaction, mean correct RTs from both experiments were submitted to a three-way ANOVA with congruence and attended feature as within-subjects factors, and experiment as a between-subjects factor. The three-way interaction between congruence, attended feature, and experiment, $F(1, 30) = 14.06$, $p = .001$, $\eta_p^2 = 0.32$, confirms that the two-way interaction in Experiment 1 (i.e., Stroop > reverse Stroop) was significantly different than the two-way interaction in Experiment 2 (i.e., Stroop < reverse Stroop).

3.3. Discussion

Experiments 1 and 2 confirmed our hypothesis based on the strength-of-association account that the Stroop effect should be larger than the reverse Stroop effect for a task that relies on verbal processing (i.e., identification), but smaller than the reverse Stroop effect for a task that relies on visual processing (i.e., localization). At the same time, we controlled for discriminability by using the same font size and colors in both experiments, and we eliminated the need for translation in Experiment 2 by using the same features for the cues and targets: The cued features and target features were colors in the Stroop condition, words in the reverse Stroop condition. Our experiments represent the first instance in which the manipulation of task (and thus the processing required) led to an inversion of asymmetry between conditions (Stroop > reverse Stroop for verbal processing, but Stroop < reverse Stroop for visual processing) while controlling for discriminability and translation.

One outcome of Experiment 2 that was not predicted by the strength-of-association account is the faster overall RTs in the Stroop condition than the reverse Stroop condition. Although color is a guiding feature in visual search whereas word meaning is probably not (Wolfe & Horowitz, 2004, 2017), we must acknowledge that the color words we used had different shapes and sizes (e.g., with just three letters, *Red* is a smaller shape than *Yellow*, with six letters). Thus, when cued with a word in the reverse Stroop condition, participants may have relied on shape and size to help guide their search. Nevertheless, faster overall responses in the Stroop condition than the reverse Stroop condition suggest that color was more visually salient than shape and size. Whereas selection efficiency is driven primarily by salience, we have argued that congruity affects the decision stage only after an item has been selected (Risko et al., 2013; Sobel & Puri, 2018). Thus, if selection could be made more efficient in the reverse Stroop condition, the congruity effect should remain intact. We call this the separate stages hypothesis.

However, an alternative explanation for faster overall RTs and a smaller effect of interference in the Stroop condition than the reverse Stroop condition is that the congruity effect is scaled to selection efficiency: faster RTs lead to a smaller congruity effect. According to this scaling hypothesis, the Stroop effect was smaller than the reverse Stroop effect merely because responses were faster in the Stroop condition than the reverse Stroop condition. Consequently, if responses for the reverse Stroop condition could be made faster, the congruity effect would shrink.

Experiment 3 was intended to distinguish between the separate stages hypothesis and the scaling hypothesis. To do so, we wanted to speed responses while leaving the search displays as similar as possible to those in Experiment 2. The separate stages hypothesis predicts that the congruity effect should remain intact when overall RTs are faster, whereas the scaling hypothesis predicts that the congruity effect should shrink when overall RTs are faster.

Because the cue remained visible after the search items appeared in Experiment 2, the cue may have held the participants' attention, thereby preventing participants from directing their attention to the search items. To promote participants' ability to disengage their attention from the cue's location in Experiment 3, the cue was only briefly visible, then disappeared before the search items appeared. By removing the cue from the display, we hoped to speed responses so there would be no RT difference between the reverse Stroop condition in Experiment 3 and the Stroop condition in Experiment 2.

4. Experiment 3: Disappearing cues

4.1. Method

4.1.1. Participants

A total of 17 students (five male, 12 female) between 19 and 25

(M = 21.1) volunteered for Experiment 3 in exchange for course credit. None had participated in either of the previous experiments. Data from one participant were excluded from analysis because his error rate (40%) for incongruent displays in the reverse Stroop condition suggests that he misunderstood the instructions. Apparently, he mistakenly believed that the word cue in the reverse Stroop condition indicated the target's color so he needed to translate the meaning of the word cue into a visual code. Our speculation that he translated the cue's meaning in the reverse Stroop condition but not the Stroop condition is supported by the fact that his RTs were much longer than other participants in the reverse Stroop condition but not the Stroop condition. His high error rate suggests that by chance, in only about half of the displays, the actual target was on the same side of the display as the item he searched for.

4.1.2. Procedure

As in Experiment 2, each trial began with the presentation of either a color or word cue, but the cue only remained visible for 750 ms, after which it was replaced by a fixation cross consisting of two orthogonal line segments each 1° long. After the fixation cross was visible for 750 ms, the four search items appeared, and remained visible along with the fixation cross until participants made a response.

4.2. Results

Mean error rates were submitted to a three-way ANOVA with congruence and attended feature as within-subjects factors, and block order as a between-subjects factor. None of the effects were significant, all $ps > .1$, so the error rates in Table 2 are collapsed across both levels of block order.

The RT trimming routine removed a total of 2.0% of data points. Mean correct RTs were submitted to a three-way ANOVA with congruence and attended feature as within-subjects factors, and block order as a between-subjects factor. None of the effects with block order as a factor were significant, all $ps > .07$, so the RTs in Fig. 1 are collapsed across both levels of block order.

The results replicated those from Experiment 2. Responses were faster for congruent targets than incongruent, $F(1, 14) = 51.33$, $p < .001$, $\eta_p^2 = 0.79$, and were faster when participants attended to the target's color (Stroop) than its meaning (reverse Stroop), $F(1, 14) = 21.82$, $p < .001$, $\eta_p^2 = 0.61$. The interaction between congruence and attended feature, $F(1, 14) = 7.12$, $p = .018$, $\eta_p^2 = 0.34$, shows that the Stroop effect was different than the reverse Stroop effect. Simple effects analysis confirmed that both the Stroop effect, $F(1, 15) = 10.89$, $p = .005$, $\eta_p^2 = 0.42$, and the reverse Stroop effect, $F(1, 15) = 42.69$, $p < .001$, $\eta_p^2 = 0.74$, were significant, and that the Stroop effect size (51.55 ± 18.93 ms, $\eta_p^2 = 0.42$) was smaller than the reverse Stroop effect size (109.27 ± 18.93 ms, $\eta_p^2 = 0.74$), as in Experiment 2.

Removing the cue before presenting the search array made responses faster, and the reverse Stroop effect (109 ms) was about the same as it had been in Experiment 2 (114 ms). Nevertheless, as can be seen in Fig. 1, RTs in the reverse Stroop condition in Experiment 3 still appear to be longer than for the Stroop condition in Experiment 2. To compare them, we submitted the mean RTs from the Stroop condition in Experiment 2 and the reverse Stroop condition in Experiment 3 to a two-way ANOVA with congruence as a within-subjects factor and experiment as a between-subjects factor. The interaction between congruence and experiment, $F(1, 30) = 11.37$, $p = .002$, $\eta_p^2 = 0.27$, confirms that the reverse Stroop effect in Experiment 3 was larger than the Stroop effect in Experiment 2. However, the main effect of experiment, $F(1, 30) = 3.77$, $p = .062$, $\eta_p^2 = 0.11$, shows that responses were marginally slower for the reverse Stroop condition in Experiment 3 than the Stroop condition in Experiment 2. Because responses were still (marginally) faster for the Stroop condition in Experiment 2 than the reverse Stroop condition in Experiment 3, scaling remains a possible explanation for a smaller Stroop effect in Experiment 2 than the reverse Stroop effect in Experiment 3.

4.3. Discussion

We intended to hasten responses in Experiment 3 so responses in the reverse Stroop condition would be as fast as the Stroop condition in Experiment 2. By doing so, we hoped to determine whether smaller Stroop effects than the reverse in Experiment 2 could be explained by a scaling effect. Unfortunately, the RTs for the reverse Stroop condition in Experiment 3 were still marginally slower than for the Stroop condition in Experiment 2. To make responses even faster for the reverse Stroop condition in Experiment 4, we eliminated the need for participants to process the meaning of the cue word by presenting them in blocks. Thus, for example, participants only had to search for *Red* throughout one block of trials, *Green* in another block, et cetera. Presenting the word cues in blocks should make the cues more predictable within each block, thereby speeding responses for the reverse Stroop condition in Experiment 4.

5. Experiment 4: Blocked word order in reverse Stroop

5.1. Method

5.1.1. Participants

A total of 17 students (five male, 12 female) between 19 and 24 (M = 21.6) volunteered for Experiment 4 in exchange for course credit. None had participated in any of the previous experiments. The data from one participant were excluded from analysis because her error rate (46%) for incongruent displays in the reverse Stroop condition suggests that she misunderstood the instructions in the same way as the participant whose data were removed from the previous experiment.

5.1.2. Procedure

The stimuli and procedure were the same as in the previous experiment, except that in the reverse Stroop condition the cue words were presented in blocks rather than randomly ordered. For example, some participants viewed 24 consecutive displays for which the word cue was *Red*, followed by 24 with *Green* as the cue, then *Blue*, then *Yellow*; block order of word cues was counterbalanced across participants. The presentation order of congruity remained random so each block of 24 identical word cues contained 12 congruent and 12 incongruent displays in random order.

5.2. Results

Mean error rates were submitted to a four-way ANOVA with congruence and attended feature as within-subjects factors, and two kinds of block order (attended feature and word cue) as between-subjects factors. Neither of the main effects of block order were significant, both $ps > .08$, so the error rates in Table 2 are collapsed across all levels of block order. Most of the interactions with either of the block orders as factors were not significant, $ps > .09$, but the interaction between congruence and (attended feature) block order was significant, $F(1, 8) = 6.91$, $p = .030$, $\eta_p^2 = 0.46$. This interaction indicates that mean error rates for participants who attended to the target word in the first block were about the same for congruent (3.13%) and incongruent (3.52%) targets, but for participants who attended to the target's color in the first block, mean error rates were much lower for congruent (2.39%) than incongruent (6.77%) targets. Perhaps the blocked presentation of target words for participants who attended to target word in the first half of the experiment encouraged them to ignore the meanings of search items so incongruent word meanings influenced them less than participants for whom the first half of the experiment was not blocked. Mean error rates were higher for incongruent targets than congruent targets, $F(1, 8) = 9.85$, $p = .014$, $\eta_p^2 = 0.55$, but RTs were also slower for incongruent targets so there is no evidence of a speed/accuracy tradeoff. The main effect of attended feature and its interaction with congruence were not significant, both $ps > .06$.

The RT trimming routine removed a total of 2.2% of data points. Mean correct RTs were submitted to a four-way ANOVA with congruence and attended feature as within-subjects factors, and two kinds of block order (attended feature and word cue) as between-subjects factors. None of the effects with either block order as factors were significant, all $ps > .3$, so the RTs in Fig. 1 are collapsed across all levels of block order.

The results replicated Experiments 2 and 3. Responses were faster for congruent targets than incongruent, $F(1, 8) = 27.26$, $p = .001$, $\eta_p^2 = 0.77$, and faster when participants attended to the target's color (Stroop) than its meaning (reverse Stroop), $F(1, 8) = 12.70$, $p = .007$, $\eta_p^2 = 0.61$. The interaction between congruence and attended feature, $F(1, 8) = 10.70$, $p = .011$, $\eta_p^2 = 0.57$, shows that the Stroop effect was different than the reverse Stroop effect. Simple effects analysis confirmed that both the Stroop effect, $F(1, 15) = 9.90$, $p = .007$, $\eta_p^2 = 0.56$, and the reverse Stroop effect, $F(1, 15) = 26.00$, $p < .001$, $\eta_p^2 = 0.63$, were significant, and that the Stroop effect size (25.99 ± 21.19 ms, $\eta_p^2 = 0.56$) was smaller than the reverse Stroop effect size (115.32 ± 21.19 ms, $\eta_p^2 = 0.63$), as in Experiments 2 and 3.

Presenting cue words in blocks made responses for the reverse Stroop condition faster than in Experiment 3, and yet the reverse Stroop effect (115 ms) was no smaller than it had been in Experiment 2 (114 ms). To verify that the reverse Stroop effect in Experiment 4 was larger than the Stroop effect in Experiment 2 even though responses were no slower, we submitted the mean RTs from the Stroop condition in Experiment 2 and the reverse Stroop condition in Experiment 4 to a two-way ANOVA with congruence as a within-subjects factor and experiment as a between-subjects factor. The interaction between congruence and experiment, $F(1, 30) = 8.41$, $p = .007$, $\eta_p^2 = 0.22$, confirms that the reverse Stroop effect in Experiment 4 was larger than the Stroop effect in Experiment 2, even though RTs were no different between experiments: $p = .55$ for the main effect of experiment.

5.3. Discussion

Presenting cue words in blocks rather than in random order elicited faster responses for the reverse Stroop condition with the same displays as in Experiment 3. Although responses for the reverse Stroop condition in Experiment 4 were no slower than the Stroop condition in Experiment 2, the reverse Stroop effect in Experiment 4 remained larger than the Stroop effect in Experiment 2. These results subvert the scaling hypothesis, which lends support to the separate stages hypothesis that interference affects the decision stage only after an item has been selected.

However, another alternative explanation for slower responses and larger interference effects in the reverse Stroop condition than the Stroop condition is that participants who were instructed and cued to attend to semantic color in the reverse Stroop condition may have covertly translated the cue word's meaning into its corresponding color. By covertly translating the cue's meaning in the reverse Stroop condition, they could use color to efficiently locate the target. And the translation step in the reverse Stroop condition would introduce a delay compared to the Stroop condition. Indeed, the two participants who were excluded from analyses (one in Experiment 3 and another in Experiment 4) seemed to do just that: search for the color named by the cue in the reverse Stroop condition even though they had been instructed to search for the *word* that matched the cue. Furthermore, by presenting as many congruent targets as incongruent targets, we may have inadvertently provided participants with an incentive to covertly translate the cue's meaning into its corresponding color.

To understand why, consider a participant who is cued with the word *Red* in the reverse Stroop condition. A participant who translates the meaning of the cue into a color could efficiently locate the target in the half of trials in which the target's color and meaning are congruent. For incongruent trials, after initially selecting the wrong item (i.e., the item whose *color* matches the cue's meaning), the participant would then need to fall back on a strategy to search for the *word* with the same meaning as the cue. Although translating the meaning of the cue word into the corresponding color would represent a cost in the reverse Stroop condition, efficient search in one half of the trials might make the effort worthwhile. According to the covert translation account, the slower responses in the reverse Stroop condition could indicate the time required to translate the cued word meaning into a color, and the larger reverse Stroop effects than Stroop effects would be attributable to the need to engage in a two-tiered search strategy for incongruent targets in the reverse Stroop conditions, rather than interference from the irrelevant target feature, as in the traditional Stroop paradigm.

To address the possibility that presenting the same number of congruent as incongruent targets may have provided an incentive to covertly translate the cued word meaning in the reverse Stroop conditions for Experiments 2 through 4, in Experiment 5 we wanted to eliminate this incentive by boosting the number of incongruent displays relative to congruent displays. Because there are three ways for a given word to have an incongruent color (i.e., *Red* presented in green, blue, or yellow) but only one way for the word to have a congruent color (i.e., *Red* in red), the most natural way to boost the ratio of incongruent to congruent displays would be to present three times as many incongruent displays as congruent displays. By using a three-to-one ratio, covert translation of the cue word in the reverse Stroop condition would efficiently locate the target in just one quarter of the trials. In Experiment 5, the cue words in the reverse Stroop condition were presented in random order as in Experiment 3, but there were three times as many incongruent displays as congruent displays.

In summary, we hypothesized that the reverse Stroop effects we observed in Experiments 2, 3, and 4 represent interference from the incongruent color and not the effect of covert translation. Thus, eliminating the incentive to covertly translate the cue's meaning in the reverse Stroop condition in Experiment 5 should have no effect on the size of the reverse Stroop effect. On the other hand, the covert translation hypothesis asserts that the reverse Stroop effects in Experiments 2, 3, and 4 represent the need to engage a two-tiered search strategy after initially selecting the wrong item among incongruent displays. This alternative account predicts that eliminating the incentive to covertly translate in the reverse Stroop condition in Experiment 5 should reduce the size of the reverse Stroop effect. Furthermore, a reduction in both the Stroop and reverse Stroop effects would be consistent with Melara and Algom's (2003) observation that across experiments, those with a higher proportion of incongruent trials (resulting in lower "contingency" between colors and words) showed smaller congruity effects.

6. Experiment 5: Three-to-one ratio of incongruent to congruent

6.1. Method

6.1.1. Participants

A total of 18 students (three male, 15 female) between 18 and 63 (M = 24.2) volunteered for Experiment 5 in exchange for course credit. None had participated in any of the previous experiments. The data from two participants were excluded from analysis. One excluded participant's mean RT was greater than the mean of the other participants' RTs plus three standard deviations. The other excluded participant's error rate of 50% or more in all conditions suggests she generated random responses.

6.1.2. Procedure

The stimuli and procedure were the same as in Experiment 3, except that in Experiment 5 there were three times as many incongruent trials as congruent trials. In each of two blocks we presented three repetitions of four target quadrants, four target words, and four offsets (of which only 0° produced congruent displays, and the other three offsets, 90°, 180°, and 270°, produced incongruent displays), randomly interleaved within the block. Thus, participants carried out a total of 396 (12 practice + 384 experimental) trials.

6.2. Results

Mean error rates were submitted to a three-way ANOVA with congruence and attended feature as within-subjects factors, and block order as a between-subjects factor. None of the effects with block order as a factor were significant, all $ps > .4$, so the error rates in Table 1 are collapsed across both levels of block order. Mean error rates were higher for incongruent targets than congruent, $F(1, 14) = 5.42$, $p = .035$, $\eta_p^2 = 0.28$, and when participants attended to the cue's meaning (reverse Stroop) than when they attended to the cue's color (Stroop), $F(1, 14) = 5.95$, $p = .029$, $\eta_p^2 = 0.30$. However, responses were also slower for incongruent targets and the reverse Stroop condition than for congruent targets and the Stroop condition, respectively, which shows that there was no speed-accuracy tradeoff.

The RT trimming routine removed a total of 2.0% of data points. Mean correct RTs were submitted to a four-way ANOVA with congruence and attended feature as within-subjects factors, and block order as a between-subjects factor. None of the effects with block order as a factor were significant, all $ps > .1$, so the RTs in Fig. 1 are collapsed across both levels of block order.

The results replicated the previous localization experiments. Responses were faster for congruent targets than incongruent, $F(1, 14) = 34.51$, $p < .001$, $\eta_p^2 = 0.71$, and were faster when participants attended to the target's color than its meaning, $F(1, 14) = 89.80$, $p < .001$, $\eta_p^2 = 0.87$. The interaction between congruence and attended feature, $F(1, 14) = 58.03$, $p < .001$, $\eta_p^2 = 0.81$, shows that the Stroop effect was different than the reverse Stroop effect. Simple effects analysis confirmed that the Stroop effect, $F(1, 15) = 6.41$, $p = .023$, $\eta_p^2 = 0.30$, and the reverse Stroop effect, $F(1, 15) = 47.21$, $p < .001$, $\eta_p^2 = 0.76$, were significant, and that the Stroop effect size (20.25 ± 11.24 ms, $\eta_p^2 = 0.30$) was smaller than the reverse Stroop effect size (118.35 ± 11.24 ms, $\eta_p^2 = 0.76$), as in the previous localization experiments. Because Experiment 5 was intended to find if presenting a majority of incongruent displays would reduce the size of the reverse Stroop effect, we submitted the results from Experiments 3 and 5 (in these experiments, the cue words in the reverse Stroop conditions were randomly interleaved but in Experiment 4 the cue words were presented in blocks) to a three-way ANOVA with congruence and attended feature as within-subjects factors and experiment as a between-subjects factor. None of the effects with experiment as a factor were significant, all $ps > .1$.

6.3. Discussion

Because presenting a majority of displays containing an incongruent target in Experiment 5 should have eliminated the incentive to covertly translate the cued word's meaning in the reverse Stroop condition, the covert translation hypothesis predicted that the reverse Stroop effect should be smaller in Experiment 5 than in previous experiments. The results from Experiment 5 did not support this hypothesis: the reverse Stroop effect was no smaller in Experiment 5 (118 ms) than in Experiment 3 (109 ms). Apparently, when instructed to search on the basis of a cue's meaning, participants will comply, even if searching on the basis of color would be more efficient. This is consistent with results from the numerical Stroop paradigm in visual search. Searching on the basis of physical size is more efficient than searching on the basis of numerical size, and yet participants search on the basis of numerical size when instructed to do so (Sobel et al., 2016), even when participants are primed to use physical size to guide their search (Wilson & Sobel, 2018).

7. General discussion

The strength-of-association account asserts that the interference between a word's color and meaning depends on the type of processing associated with the task demands. Accordingly, identification, as required in the traditional Stroop task, is more strongly associated with verbal than visual processing, so incongruent word meaning should interfere with identification of the targets' color (Stroop) more than incongruent color interferes with word reading (reverse Stroop). On the other hand, localization is more strongly associated with visual than verbal processing, so incongruent color should interfere with localization of the target word (reverse Stroop) more than incongruent word meaning interferes with localization of the target color (Stroop). Furthermore, these effects should occur even in the absence of any need for translation between verbal and visual codes.

In Experiment 1, the Stroop effect was larger than the reverse Stroop effect, which is consistent with the discriminability, translation, and strength-of-association accounts. If these factors additively combine to induce a larger Stroop effect than the reverse, that could be the reason the classic Stroop asymmetry has historically been so robust. To explain the claim by Grégoire et al. (2019) that recent studies finding a reverse Stroop effect have all elicited manual responses, we speculated that manual responses entail more visuospatial engagement than vocal responses. Nevertheless, the Stroop asymmetry we observed in Experiment 1, in which we elicited manual responses, is inconsistent with the response modality account. We argued that it is unclear how to explain this inconsistency until further experiments are carried out in which the response modality is manipulated while other factors are held constant.

Experiment 2 used the same font size and colors as in Experiment 1, so the discriminability account predicts that the relative sizes of the Stroop and reverse Stroop effects should have remained the same across both experiments. We eliminated the need for translation in Experiment 2 by using cues that relied on the same kind of code as the attended feature; color patch cues indicated that participants should attend to the target's color in the Stroop condition, and word cues indicated that participants should attend to the target's meaning in the reverse Stroop condition. By manipulating the task requirements so identification in Experiment 1 required verbal processing but localization in Experiment 2 required visual processing, we tested the strength-of-association account's prediction of a larger Stroop effect than the reverse in Experiment 1, but smaller Stroop effect than the reverse in Experiment 2. The results from Experiments 1 and 2, in which the Stroop effect was larger than the reverse for identification in Experiment 1, but smaller than the reverse for localization in Experiment 2, were consistent with the predictions of the strength-of-association account.

This is not to say that the results from Experiment 2 were necessarily inconsistent with the translation account. As we have speculated, discriminability, translation, and strength of association may have additively contributed to the Stroop asymmetry we observed in Experiment 1. If so, then eliminating the need for translation in Experiment 2 simply means that translation made no contribution to the asymmetry in that experiment. However, the inconsistency between the results from Experiment 2 and the predictions of the discriminability account does seem to require some explanation.

One possible explanation for the inconsistency between the Experiment 2 results and the discriminability account is that, as with strength of association, the role of discriminability switches with task demands. Specifically, a task that relies on verbal processing may enhance the discriminability of semantic colors whereas a task that relies on visual processing may enhance the discriminability of colors. Thus, even if two tasks use the same font size and colors (as we did in Experiments 1 and 2), the words could be more discriminable than the colors for identification but less discriminable than the colors for localization. This hypothesis could be tested by a series of experiments in the spirit of Dunbar and MacLeod (1984) and Melara and Mounts (1993). That is, they argued that the Stroop effect is typically larger than the reverse Stroop effect because the typical words are more discriminable than the typical colors, but reducing the discriminability of words inverts the asymmetry. By extension, we consistently found larger reverse Stroop effects in visual search experiments, which might indicate that for a task requiring visual processing, the colors we used were more discriminable than the words. Thus, reducing the discriminability of the colors could invert the asymmetry we observed, as in Dunbar and MacLeod (1984) and Melara and Mounts (1993). Such a result would suggest that even though we tried to control discriminability by using the same font size and colors in Experiments 1 and 2, manipulating the task demands switched the advantage for words in identification to an advantage for colors in localization. In turn, this would reconcile the discriminability account with our Experiment 2 results.

And while we intended to control the discriminability of the target's color and meaning between the identification (Experiment 1) and localization tasks (Experiments 2–5), we must acknowledge that targets were presented at fixation in Experiment 1 but were presented peripherally in the remaining experiments. Nevertheless, while visual acuity falls off with eccentricity, it is not clear whether visual eccentricity would differentially affect the discriminability of the targets' color or meaning. Also, participants may have examined the search items by shifting their gaze from one item to another, in which case they would examine each item with their fovea. Further experiments using eye tracking equipment will be needed to reveal whether participants shift their gaze to foveate search items.

Whereas the inversion of asymmetry we observed between Experiments 1 and 2 (larger Stroop than the reverse for identification versus smaller Stroop than the reverse for localization) was predicted by strength of association, the slower responses for the reverse Stroop condition for localization in Experiment 2 was not. We argued that this effect is attributable to the target colors being more visually salient than the target word, which led to more efficient search when participants were cued with a color in the Stroop condition than when they were cued with a word in the reverse Stroop condition. Furthermore, we argued that while salience influenced the selection stage, incongruity between the target's color and meaning affected the decision stage only after the target had been selected for further processing. By affecting different processing stages, the effect of salience (faster responses in the Stroop condition than the reverse Stroop condition) were independent of the effect of incongruity (smaller Stroop effect than reverse Stroop effect).

Experiments 3, 4, and 5 investigated alternative explanations for why responses were faster in the Stroop condition than the reverse Stroop condition, and the Stroop effect was smaller than the reverse Stroop effect, in Experiment 2. Perhaps the congruity effect was scaled on selection efficiency; according to this scaling hypothesis, the Stroop effect was smaller than the reverse in Experiment 2 simply because responses were faster in the Stroop condition. In Experiment 3, responses were marginally slower for the reverse Stroop condition than for the Stroop condition in Experiment 2, but in Experiment 4 responses were no slower for the reverse Stroop condition than for the Stroop condition in Experiment 2. Nevertheless, the reverse Stroop effect in Experiment 4 was larger than the Stroop effect in Experiment 2, which shows that the larger reverse Stroop effect could not be the result of scaling.

A second alternative is that participants implicitly recognize that attending to a color can more efficiently guide attention than attending to a word. Further, if half of the displays are congruent, translating the cue's meaning into the corresponding visual color in the reverse Stroop condition would enable a participant to initially select the target in half of the trials; a secondary search stage would only be required for incongruent displays. Our previous work with numerical Stroop in visual search (Sobel et al., 2016; Sobel & Puri, 2018; Wilson & Sobel, 2018) made us skeptical of the covert translation hypothesis, but nevertheless in Experiment 5 we directly tested this alternative. Presenting a majority of trials in which displays were incongruent should have eliminated the incentive to covertly translate the cue's meaning in the reverse Stroop condition, and yet the reverse Stroop effect was no smaller than it had been in previous experiments. This suggests that, as with numerical Stroop in visual search, participants who are instructed to attend to a cue word do not covertly translate the cue's meaning into a visual code to take advantage of a more efficient selection stage.

By discounting the scaling hypothesis and the covert translation hypothesis, Experiments 3, 4, and 5 corroborate our initial conclusions from Experiments 1 and 2. Specifically, the Stroop effect is larger than the reverse Stroop effect for identification because the task is more strongly associated with verbal than visual processing, whereas the Stroop effect is smaller than the reverse Stroop effect for localization because the task is more strongly associated with visual than verbal processing.

7.1. Conclusions

In contrast to the classic Stroop asymmetry in which Stroop effects are typically larger than reverse Stroop effects, we observed reverse Stroop effects that were larger than Stroop effects in four localization experiments. The current study shows that in localization tasks, the reverse Stroop effect is much more robust than previously thought. By obtaining these results across various manipulations of our displays, it seems that the classic Stroop asymmetry could be a historical accident. For decades, the traditional Stroop paradigm has required identification, but if localization tasks had been as widely used, perhaps the classic Stroop asymmetry would never have had the opportunity to become a classic.

CRediT authorship contribution statement

Kenith V. Sobel: Conceptualization, Methodology, Software, Formal analysis, Resources, Writing - original draft, Writing - review & editing.**Amrita M. Puri:**Methodology, Formal analysis, Writing - review & editing.**A. Kane York:**Investigation, Project administration.

References

American Psychological Association (2017). Ethical principles of psychologists and code of conduct. Retrieved from http://www.apa.org/ethics/code.

Bausell, R. B., & Li, Y.-F. (2002). *Power analysis for experimental research: A practical guide for the biological, medical, and social sciences.* Cambridge University Press.

Blais, C., & Besner, D. (2006). Reverse Stroop effects with untranslated responses. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1345–1353. https://doi.org/10.1037/0096-1523.32.6.1345.

Blais, C., & Besner, D. (2007). A reverse Stroop effect without translation or reading difficulty. *Psychonomic Bulletin & Review, 14*(3), 466–469. https://doi.org/10.3758/BF03194090.

Dadon, G., & Henik, A. (2017). Adjustment of control in the numerical Stroop task. *Memory and Cognition, 45*, 891–902. https://doi.org/10.3758/s13421-017-0703-6.

Dukewich, K. R., & Klein, R. M. (2009). Finding the target in search tasks using detection, localization, and identification responses. *Canadian Journal Of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale, 63*(1), 1–7. https://doi.org/10.1037/a0012780.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 622–639. https://doi.org/10.1037/0096-1523.10.5.622.

Durgin, F. H. (2000). The reverse Stroop effect. *Psychonomic Bulletin & Review, 7*(1), 121–125. https://doi.org/10.3758/BF03210730.

Grégoire, L., Poulin-Charronnat, B., & Perruchet, P. (2019). Stroop interference depends also on the level of automaticity of the to-be-interfered process. *Acta Psychologica, 197*, 143–152. https://doi.org/10.1016/j.actpsy.2019.05.013.

Kinoshita, S., Mills, L., & Norris, D. (2018). The semantic Stroop effect is controlled by endogenous attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(11), 1730–1742. https://doi.org/10.1037/xlm0000552.

Krause, F., Bekkering, H., Pratt, J., & Lindemann, O. (2017). Interaction between numbers and size during visual search. *Psychological Research, 81*(3), 664–677. https://doi.org/10.1007/s00426-016-0771-4.

Loftus, G. R., & Masson, M. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*(4), 476–490. https://doi.org/10.3758/

BF03210951.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110*(3), 422–471. https://doi-org.ucark.idm.oclc.org/10.1037/0033-295X.110.3.422.

Melara, R. D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition, 21*(5), 627–645. https://doi.org/10.3758/BF03197195.

Miller, H. C., Kubicki, S., Caffier, D., Kolski, C., & Naveteur, J. (2016). The Stroop and reverse Stroop effects as measured by an interactive tabletop. *International Journal Of Human-Computer Interaction, 32*(5), 363–372. https://doi.org/10.1080/10447318.2016.1150642.

Risko, E. F., Maloney, E. A., & Fugelsang, J. A. (2013). Paying attention to attention: Evidence for an attentional contribution to the size congruity effect. *Attention, Perception, & Psychophysics, 75*, 1137–1147. https://doi.org/10.3758/s13414-013-0477-2.

Sobel, K. V., & Puri, A. M. (2018). Size congruity influences visual search via the target template. *Acta Psychologica, 183*, 66–74. https://doi.org/10.1016/j.actpsy.2018.01.003.

Sobel, K. V., Puri, A. M., & Faulkenberry, T. J. (2016). Bottom-up and top-down attentional contributions to the size congruity effect. *Attention, Perception, & Psychophysics, 78*, 1324–1336. https://doi.org/10.3758/s13414-016-1098-3.

Song, Y., & Hakoda, Y. (2015). An fMRI study of the functional mechanisms of Stroop/reverse-Stroop effects. *Behavioural Brain Research, 290*, 187–196. https://doi.org/10.1016/j.bbr.2015.04.047.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662. https://doi.org/10.1037/h0054651.

Sugg, M. J., & McDonald, J. E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 20*(3), 647–675. https://doi.org/10.1037/0096-1523.20.3.647.

Uleman, J. S., & Reeves, J. (1971). A reversal of the Stroop interference effect, through scanning. *Perception & Psychophysics, 9*(3–A), 293–295. https://doi.org/10.3758/BF03212651.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition, 13*(4), 304–319. https://doi.org/10.3758/BF03202499.

Wilson, K., & Sobel, K. V. (2018). The size congruity effect in visual search: Further evidence for late interaction. *Psi Chi Journal of Psychological Research, 23*, 311–319. https://doi.org/10.24839/2325-7342.JN23.4.311.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.). *Attention*. East Sussex, U. K: Psychology Press Ltd.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*, 1–7. https://doi.org/10.1038/nrn1411.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*(3), 58. https://doi.org/10.1038/s41562-017-0058.

Yamamoto, N., Incera, S., & McLennan, C. T. (2016). A reverse Stroop task with mouse tracking. *Frontiers in Psychology, 7*, 670. https://doi.org/10.3389/fpsyg.2016.00670.