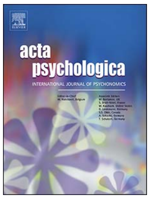
Acta Psychologica 183 (2018) 66–74

Contents lists available at ScienceDirect

Acta Psychologica

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



Size congruity influences visual search via the target template

Kenith V. Sobel[a,*], Amrita M. Puri[b]

[a] Department of Psychology and Counseling, University of Central Arkansas, United States
[b] Department of Biology, University of Central Arkansas, United States

ABSTRACT

In numerical comparison experiments, participants are presented with two digits that vary in numerical and physical size, and they select the numerically (or physically) larger (or smaller) of the two digits. Response times are typically faster when numerical and physical size are congruent than when they are incongruent, which is called the size congruity effect (SCE). Although numerical size is unlikely to be a guiding feature in visual search, recent studies have nevertheless observed the SCE in the visual search paradigm. To explain this puzzling fact, we hypothesized that the incongruity between a target's numerical and physical size affects visual search primarily when an attended item is compared to the target template in visual short-term memory. In three experiments, participants searched for a target whose numerical and physical size were distinct from non-target distractors. The SCE and shallow search slopes in Experiment 1 suggest that the target's physical size captured attention, and only then did incongruent numerical size interfere with the response. Instructing participants to attend to physical size in Experiment 2 abolished the SCE, suggesting that participants did not analyze the target's numerical size when they could be confident that physical size was a reliable target cue. Presenting each of two possible target digits in blocks as in Experiment 3 enabled participants to load the visual features of shape and physical size into their target template, and once again the SCE was abolished. The three experiments show that the SCE in visual search can be reduced or eliminated by restricting the target template based on specific physical features and thus discouraging participants from analyzing the target's numerical size.

1. Introduction

In traditional numerical comparison tasks (Moyer & Landauer, 1967), participants select one of two digits based on its numerical size. Besner and Coltheart (1979) extended this technique by varying the digits' physical size so numerical and physical size could be either congruent (e.g., 2 and 9) or incongruent (e.g., 2 and 9). In such a task, selecting the numerically (or physically) larger (or smaller) digit is generally faster when the numerical and physical size are congruent than when incongruent (Besner & Coltheart, 1979; Henik & Tzelgov, 1982). This result, called the size congruity effect (SCE), implies that a numeral's semantic (numerical size) and perceptual (physical size) characteristics interact mentally in a manner reminiscent of the classic Stroop (1935) effect.

2. Stroop and reverse Stroop effects in identification and localization

In one of Stroop's (1935) experiments, participants viewed either color words written in ink that was incongruent with the meaning of the word, or colored blocks. Naming the color of the ink was slower for incongruent color words than colored blocks, which has become known as the Stroop effect. Much less well known than the color naming experiment was one in which participants read the words aloud (MacLeod, 1991). In this experiment, word reading was no slower for color words written in an incongruent color ink than color words written in a neutral (i.e., black) ink. In a third experiment, Stroop showed that incongruent ink color *can* interfere with word reading, but only after several days practicing ink color naming, and this effect promptly vanished in a follow-up task. Although Stroop found interactions between a word's meaning and ink color in both color naming and word reading tasks, the first has become known as the Stroop effect and the second as the reverse Stroop effect.

The likely reason for this naming convention is that the Stroop effect is so much more robust than the reverse Stroop effect. Indeed, whereas MacLeod (1991) reviewed hundreds of articles replicating the Stroop effect, replications of the reverse Stroop effect are comparatively rare (Blais & Besner, 2006). This asymmetry between the Stroop and reverse Stroop effects has traditionally been explained as the result of automaticity (Besner, Stolz, & Boutilier, 1997; Blais, Harris, Guerrero, &

---

[*] Corresponding author at: Department of Psychology and Counseling, University of Central Arkansas, 201 Donaghey Ave., Mashburn Hall 260, Conway, AR 72035, United States.
E-mail address: ksobel@uca.edu (K.V. Sobel).

https://doi.org/10.1016/j.actpsy.2018.01.003
Received 13 September 2017; Received in revised form 7 January 2018; Accepted 7 January 2018

Bunge, 2012). That is, participants are presumed to have had much more experience reading words than naming colors, so incongruent word meaning interferes with ink color naming (Stroop) more than the other way around (reverse Stroop).

Whereas the automaticity account implies a special status for word reading, many studies replicating the reverse Stroop effect (e.g., Virzi & Egeth, 1985) have argued that Stroop interference results from the need to translate mental codes between the stimulus and response. Because identification of either the target's color (Stroop) or meaning (reverse Stroop) entails a verbal response, a visual stimulus (ink color) needs to be translated into a verbal code in Stroop tasks, but a verbal stimulus (word meaning) requires no such translation in reverse Stroop tasks. The translation account implies that tasks eliciting a visual response should invert the traditional asymmetry such that the Stroop effect should be smaller than the reverse Stroop effect. To support this claim, Durgin (2000; and a recent replication by Miller, Kubicki, Caffier, Kolski, & Naveteur, 2016) presented color words that appeared in a visual color, and instructed participants to localize one of four color patches that matched either the cue's color (Stroop) or meaning (reverse Stroop). The Stroop task required no translation between the cue's color and the matching color patch, but the reverse Stroop task *did* require the cue's meaning be translated into a visual code to match the corresponding color patch. Consistent with the translation account, the Stroop effect was smaller than the reverse Stroop effect.

Blais and Besner (2007) argued that a localization task such as the one used by Durgin (2000) should have been sufficient to elicit a reverse Stroop effect even without any need for translation. That is, localization tasks are more strongly associated with perceptual processing than semantic processing, so attending to the target's semantic feature (word meaning) in a localization task should elicit more interference than attending to its perceptual feature (color). In contrast, the traditional Stroop task is identification, which is more strongly associated with semantic processing. According to the strength-of-association account, this is why attending to the target's perceptual feature (color) in traditional Stroop tasks elicits more interference than attending to the target's semantic feature (word meaning) in reverse Stroop tasks. Blais and Besner (and a recent replication by Yamamoto, Incera, & McLennan, 2016) adapted Durgin's task by replacing the color patches with color words so no translation was required between the meaning of the cue and the meaning of the matching color word. Consistent with the strength-of-association account, they observed a reverse Stroop effect even though no translation was required.

Sobel, Puri, and Faulkenberry (2016) recently extended the size congruity paradigm to a visual search localization task (as Blais and Besner (2007) did for Stroop). This study included both a reverse Stroop task, in which participants localized the item with a unique numerical (semantic) size (Experiment 1), and a Stroop task in which they localized the item with a unique physical (perceptual) size (Experiment 2). In both experiments, every display contained one item that was both numerically and physically unique; the only difference between experiments was that participants were instructed to attend to numerical size in Experiment 1 and physical size in Experiment 2. In both experiments, RTs were faster for congruent targets than incongruent targets, but this SCE was significantly greater in Experiment 1 (reverse Stroop) than Experiment 2 (Stroop). Experiments 4 and 5 were also analogous to a reverse Stroop and Stroop task, respectively, but targets and distractors were three-digit numerals. Because salience of visual features increases with display density (Bravo & Nakayama, 1992; Sobel, Pickard, & Acklin, 2009; Todd & Kramer, 1994), packing more items into the same size display was intended to boost the salience of the target's physical size, thereby reducing the role of numerical size. As expected, the significant SCE in Experiment 4 (reverse Stroop) was not just reduced, but completely abolished, in Experiment 5 (Stroop).

3. The presence of the SCE in visual search is surprising

A larger SCE when participants attended to a target's numerical size rather than its physical size accords well with the strength-of-association prediction that reverse Stroop should be larger than Stroop effects for localization tasks, and yet the mere presence of the SCE in visual search is somewhat surprising. One obstacle to observing the SCE in visual search is that manipulating a search item's semantic associations typically also entails manipulating its shape (e.g., 9 is numerically larger than 2, but also has a different shape), so it is difficult to disentangle the effect of numerical size from the effect of shape (Krueger, 1984; Wolfe & Horowitz, 2004). Nevertheless, researchers have recently developed an assortment of techniques to control for an alphanumeric character's shape in visual search, enabling them to reveal the influence of the character's meaning on visual search (Godwin, Hout, & Menneer, 2014; Krause, Bekkering, Pratt, & Lindemann, 2017; Lupyan, 2008; Lupyan & Spivey, 2008; Schwarz & Eiselt, 2012; Sobel, Puri, & Hogan, 2015).

Whereas these studies tamed the confound between a target character's shape and meaning, a second obstacle to observing the SCE in visual search concerns the dubious status of numerical size as a guiding feature in visual search (Sobel, Puri, Faulkenberry, & Dague, 2017). A guiding feature is defined by its ability to limit the range of items through which search proceeds (Wolfe & Horowitz, 2004). If numerical size is not a guiding feature, how does it exert any influence on visual search? We believe that the target first captures attention due to its unique physical size (undoubtedly a guiding feature according to Wolfe & Horowitz), then, only after attention is directed to the physical size singleton, does its numerical size have the opportunity to interfere with the participant's decision to report that the attended item is the target. This echoes Risko, Maloney, and Fugelsang (2013), who argued that in traditional size congruity experiments with just two numbers to compare, one number captures attention, and only then does incongruent numerical size interfere with the participant's decision. However, Arend and Henik (2015) identified two methodological limitations that they claimed undermined the validity of Risko et al.'s conclusions: participants selected the numerically larger item, but were never asked to select the numerically smaller item, nor were they ever asked to attend to the items' physical size. To extend on Risko et al. while seeking to overcome the methodological limitations identified by Arend and Henik, in our experiments we included four conditions: participants searched for the numerically small, numerically large, physically small, and physically large item.

4. The role of the target template

Our attentional-capture-then-interference model of the SCE in visual search relies heavily on the role of the target template in visual short-term memory (VSTM). When participants search for a single item of interest from among several non-target distractors, they maintain a target template in VSTM for comparison with target candidates (Beck, Hollingworth, & Luck, 2011; Olivers, Peters, Houtkamp, & Roelfsema, 2011). The precision of the target template affects both attentional guidance and decision-making (Hout & Goldinger, 2015).

As a first step to probe the influence of the target template on the SCE in visual search, we noted that in our previous study when participants were instructed to search for the digit with a unique numerical size (Sobel et al., 2016, Experiment 1), the target's physical size varied randomly across trials. Because of this inter-trial interference, participants were prevented from developing a template that specified the target's physical size. What if each participant were exposed to exclusively congruent or incongruent targets?

It is well known that the Stroop effect is sensitive to the ratio of the frequency of congruent and incongruent trials (Blais et al., 2012; Jiménez & Méndez, 2013), but in our experiments we wanted to see if presenting exclusively congruent or incongruent targets would

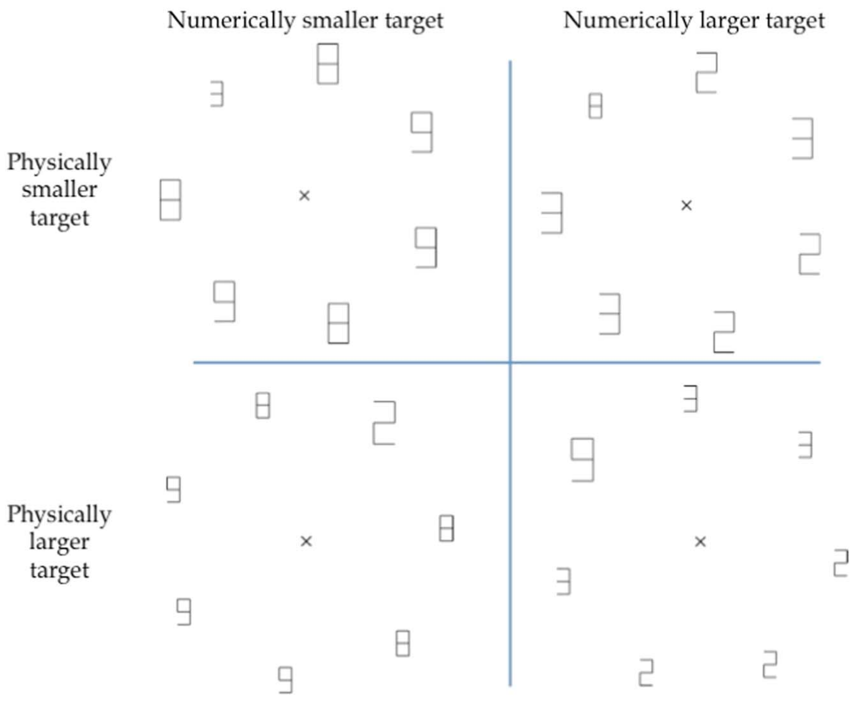


Fig. 1. Stimulus arrays containing seven items (one target and six distractors) in each of four conditions. The numerical and physical size of the target were congruent in the upper left and lower right displays, incongruent in the lower left and upper right displays.

encourage participants to load the target template with the target's visual features. For that reason we looked to Stroop experiments in which congruent and incongruent trials were presented in blocks to gain some idea of what to expect. One problem with blocking congruent trials in Stroop tasks is that participants who are told to identify the target's visual color may actually rely on word reading rather than color naming (Algom & Fitousi, 2016). Because localization is more strongly associated with perceptual processing, the strength-of-association account predicts that participants in a localization task who are instructed to attend to the target's numerical size may instead rely on attending to the target's physical size. This is what we expected in congruent trials, but what about incongruent trials? If the target has a unique physical size in every trial, we expected that participants who view exclusively incongruent targets may also have an incentive to attend to physical size because the physical size singleton will always be the target.

In the experiments described below, we eliminated inter-trial interference of the target's physical size by manipulating physical size between subjects. As a result, the target's physical size was not just predictably unique, but also remained the same (either larger or smaller than distractors) throughout the experiment. This provided participants with an incentive to adopt a top-down strategy of loading a particular physical size into their target template, thereby making search more efficient (Kiss & Eimer, 2011). Furthermore, if participants explicitly noticed that the target always had a particular physical size, they could be expected to skip checking the physical size singleton's numerical size (because it always matched the target's numerical size), thereby abolishing, or at least severely curtailing, the SCE in visual search.

In Experiment 1, we manipulated numerical and physical size between subjects in a numerical comparison visual search task in order to test two hypotheses. First, that the target's unique physical size would capture attention, and only then would incongruent numerical size interfere with selecting the attended item. And second, that upon initially localizing the target, participants would be less likely to check whether the physical size singleton's numerical size matched the target's numerical size than when the target's physical size varied across trials.

5. Experiment 1: Search for a numerical size singleton

5.1. Method

5.1.1. Participants

We obtained permission to carry out all three experiments from the University of Central Arkansas Institutional Review Board, and treated participants in accordance with the ethical guidelines stipulated by the American Psychological Association (2017). In light of recent studies that have revealed a size congruity 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 .05 (Bausell & Li, 2002). A total of 56 undergraduate students (4 groups of 14) from the University of Central Arkansas volunteered for the experiment in exchange for class credit.

5.1.2. Apparatus

All experiments were conducted on a Macbook computer connected to a 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.

5.1.3. Stimuli

When selecting digits to use as targets and distractors, we wanted digits that would encourage participants to create a target template that was as simple as possible. Because a target template containing digits that are adjacent on the number line is simpler than when the target digits are separated (Sobel et al., 2015), we tried to find two pairs of adjacent digits that have similar shapes. Godwin et al. (2014) used multidimensional scaling to create a similarity map for the shapes of the ten digits in a Verdana font, but in their similarity map, none of the pairs of digits that are nearby in similarity space are adjacent on the number line. To boost the similarity of the ten digits, we rendered the digits out of line segments as on the faces of digital clocks (Sobel et al., 2015, and as can be seen in the screenshots in Fig. 1), then used a metric developed by Cohen (2009) to calculate the similarity between each pair of adjacent digits. According to Cohen's metric, the physical similarity (P) between any two digits in a “clockface” font = O/D,

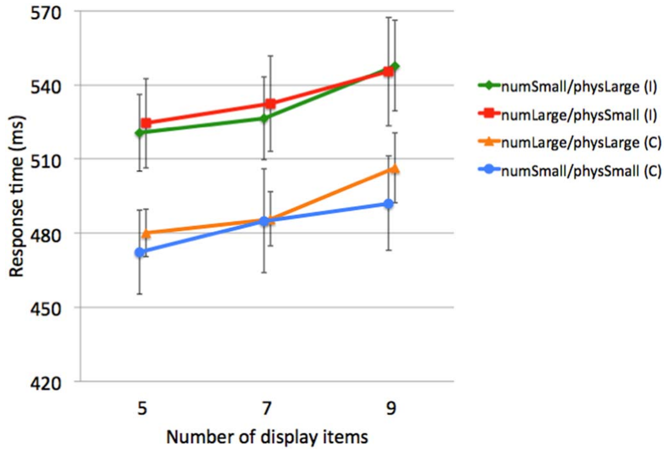


Fig. 2. Response time as a function of number of display items in Experiment 1. The letter in parentheses indicates whether numerical and physical size were congruent (C) or incongruent (I). Error bars represent standard error of the mean.

where O is the number of line segments that two digits share and D is the number of non-shared line segments. The only adjacent digit pairs with a P greater than one (i.e., with more line segments that overlap than line segments that differ) are 2 and 3 (P = 2), 5 and 6 (P = 5), and 8 and 9 (P = 6). Because we wanted digit pairs that were numerically small and numerically large (i.e., less than or greater than 5), we selected 2 and 3 as numerically small digits, and 8 and 9 as numerically large digits.

All four digits were used in all conditions. At a viewing distance of 56 cm, the component line segments for the physically large digits spanned 0.92° of visual angle so each digit was 0.92° wide × 1.8° tall, and line segments for the physically small digits spanned 0.61° so each digit was 0.61° wide × 1.2° tall. Each visual array contained one target digit and either four, six, or eight distractor digits. The search items (target plus distractors) were arranged on an imaginary circle with a radius of 5.9° and centered on a fixation cross consisting of two orthogonal line segments each 1.0° long. The fixation mark and digits were white against a black background. The target digit was positioned in one of four quadrant locations: upper right, lower right, lower left, or upper left. The participants' task in each trial was to indicate which side of the display contained the target. To ensure that the position of the target was readily distinguishable from the vertical meridian, targets were always placed at least 30° of arc away from vertical; i.e., in terms of a clock face, targets in the upper right quadrant were placed at a randomly determined location between 1 o'clock and 3 o'clock, in the lower right quadrant between 3 o'clock and 5 o'clock, in the lower left quadrant between 7 o'clock and 9 o'clock, and in the upper left quadrant between 9 o'clock and 11 o'clock.

5.1.4. Procedure

Participants were randomly assigned to one of two numerical size conditions (numerically small target and numerically large target) and two physical size conditions (physically small target and physically large target), resulting in four between-subjects conditions. In the two numerically small conditions participants searched for a target that was either a 2 or a 3 among distractors that were 8 s and 9 s, and vice versa for participants in the numerically large conditions.

The experiment began by presenting a series of instructional windows that participants could read at their own pace then click a button labeled 'Next' to advance to the next window. The instructions directed participants to search for a number either < 5 or > 5 (depending on condition), but did not mention that the target digit would have a unique physical size in all displays. Each trial began with the onset of the stimulus array, which remained visible until participants responded by pressing either the 'z' key to indicate that the target appeared on the left side of the display or the '/' key to indicate that the target appeared on the right side of the display.

The time between the onset of the stimulus array and the keypress was recorded for each trial. If the correct response was given, the stimulus array disappeared, leaving only the fixation cross on the screen for 750 ms, followed by the stimulus array for the next trial. When participants made an error, a white screen with the word 'Incorrect' in the middle appeared for 750 ms, followed by the screen containing just the fixation mark for 750 ms until the stimulus array for the next trial appeared. Each participant completed 14 replications of every combination of target quadrant (4 levels), target digit (2 levels), and number of display items (3 levels), for a total of 336 experimental trials. After completing half of the trials participants were invited to take a short break. The first six trials overall and the first six trials after the break were considered practice so participants carried out a total of 348 (336 experimental + 12 practice) trials, lasting approximately 20 min. Results from error and practice trials were excluded from analysis.

5.2. Results

Error rates were submitted to a four-way Analysis of Variance (ANOVA) with numerical size and physical size as between-subjects factors, and number of display items and target digit as within-subjects factors. Accuracy was > 97% in all conditions. None of the main effects and none of the interactions were significant, all $ps > .05$. Perhaps the consistent null effects of error rates are attributable to our using localization tasks, for which error rates tend to be low and flat across varying numbers of display items compared to the more commonly used detection task in visual search (Dukewich & Klein, 2009).

Mean correct RTs (depicted in Fig. 2) were submitted to a four-way ANOVA with numerical size and physical size as between-subjects factors, and number of display items and target digit as within-subjects factors. As is common in visual search experiments, RTs increased with the number of display items, $F(2, 104) = 27.5$, $MSE = 15{,}847.43$, $p < .001$, $\eta_p^2 = .35$. Nevertheless, the mean slope of RT as a function of number of display items (5.89 ms per item) was shallower here than the conventional criterion for visual search (< 20 ms per item, Wolfe, 1998) indicating that the target popped out from the distractors. The main effects of numerical and physical size were not significant (both $Fs < 1$), but their interaction, $F(1, 52) = 4.58$, $MSE = 166{,}739.00$, $p = .037$, $\eta_p^2 = .086$, indicates that search was slower when numerical and physical size were incongruent than when they were congruent. None of the other effects were significant, all $ps > .05$.

If search in the incongruent conditions was slower because of a processing cost per search item, then the RT slopes from the incongruent conditions should be steeper than the congruent conditions. If, however, as we hypothesized, interference from incongruent physical size induced a fixed cost after the physical size singleton captured attention, RTs should be slower but no steeper for the incongruent conditions than the congruent conditions. To distinguish between these two possibilities, we submitted the RT slopes to a two-way ANOVA with numerical size and physical size as between-subjects factors. The two main effects and their interaction were not significant, all $Fs < 1$.

5.3. Discussion

The interaction between numerical and physical size revealed the presence of the SCE in visual search, replicating previous studies (Krause et al., 2017; Sobel et al., 2016), so it seems that manipulating numerical and physical size between subjects reduced, but did not abolish, the SCE. As we discovered in Sobel et al. (2016), the SCE in visual search is surprisingly tough to kill.

5.3.1. Alternative explanations for the observed SCE

Because manipulating numerical size entails a confounding manipulation of visual features (Wolfe & Horowitz, 2004), researchers who find an effect of numerical size on visual search need to discount alternative explanations based on visual features such as brightness and shape (Godwin et al., 2014; Schwarz & Eiselt, 2012; Sobel et al., 2015). The brightness of the digits used in Experiment 1 increased with the number of line segments each digit contained, as well as the line segments' lengths. Each of the numerically smaller digits (2 and 3) contained five line segments, so they were dimmer than the 8, which contained seven line segments, and the 9, which contained six line segments. Against a dark background, brighter items capture attention more than dimmer items (Braun, 1994; Nothdurft, 2006) so numerically large (more line segments) and physically large (longer line segments) targets had a brightness advantage. If this advantage affected search times, numerically and physically large targets should have elicited the fastest RTs and numerically and physically small targets the slowest. While an advantage for brightness can explain fast responses in the physically and numerically large target condition, it cannot explain fast responses in the physically and numerically small target condition. Another possible explanation for the SCE is that the target's physical size interacted with its shape.

In the similarity map developed by Godwin et al. (2014), 3 and 8 are near each other, but they are distant from both 2 and 9, which are distant from each other. However, their similarity map was based on digits presented in a standard (Verdana) font, whereas our "digital clock" digits built from line segments are probably more similar in shape to each other, which may explain the lack of any effect of target digit. By discounting explanations based on brightness and shape, we feel confident that the SCE indicates that physical size interacted with numerical size rather than any visual feature confounded with numerical size.

5.3.2. Support for our hypotheses

The results from Experiment 1 supported our first hypothesis that the physical size singleton would capture attention. First, the shallow slopes were characteristic of visual search for a singleton distinguished by a unique guiding feature. Second, although search was significantly slower in the incongruent conditions than the congruent conditions, RT slopes were not steeper. This shows that the slower RTs for the incongruent condition cannot be attributed to a processing cost per search item, as would be expected if incongruent numerical size interfered with the search process per se. Instead, the target's physical size captured attention, and only then did incongruent numerical size interfere with selecting the attended item as the target. This supports the account proposed by Risko et al. (2013). By asking participants to select both numerically large and small targets, we overcame the methodological limitations in Risko et al. identified by Arend and Henik (2015).

The results also support our second hypothesis that eliminating inter-trial variations in physical size would discourage participants from checking the numerical size of the physical size singleton because the item with a particular physical size was always the target. The numerical and physical sizes in Experiment 1 were the same as in Experiment 1 in Sobel et al., so if the SCE were driven exclusively by interference *within* a display, the effect size of the interaction between numerical and physical size should be the same here as in Sobel et al. While we acknowledge that caution is warranted when comparing effect sizes between data sets, here the effect size of the SCE ($\eta_p^2 = .086$) was nearly an order of magnitude smaller than in Sobel et al. ($\eta_p^2 = .72$). Of course, here we manipulated numerical and physical size between subjects, so the resulting analysis is subject to between-subjects variance that would be disregarded in a within-subjects analysis, as in Sobel et al. As a result, it is impossible to isolate the effects of different treatments (i.e., exposing participants to just one level of physical size as we did here as opposed to both levels as in Sobel et al.) from different analyses, but we believe that at least some of the order of magnitude reduction in effect size can be attributed to the treatment in the present experiment. Furthermore, the preliminary support for the second hypothesis provided by Experiment 1 is supplemented by the results from Experiment 2.

Experiment 2 included the same stimuli as Experiment 1, but participants were instructed to attend to physical size, rather than numerical size. We had three reasons to expect that doing so would reduce the SCE. First, the presence of an SCE even when the target had a predictable physical size, as in Experiment 1, suggests that even though participants might have noticed that the physical size singleton was always the target, they seemed to retain some suspicion that the displays might switch, so they needed to check the physical size singleton's numerical size just to be certain it was the target. If participants had instead been reassured that the physical size singleton was always the target, they might have felt comfortable ignoring the physical size singleton's numerical size. Second, in Sobel et al. (2016), instructing participants to attend to physical size in Experiment 2 reduced the SCE. And third, the strength-of-association account (Blais & Besner, 2007) predicts that in localization tasks, Stroop effects (attend to the perceptual feature, as in Experiment 2) should be smaller than reverse Stroop effects (attend to the semantic feature, as in Experiment 1).

6. Experiment 2: Search for a physical size singleton

6.1. Method

6.1.1. Participants

A total of 56 undergraduate students (4 groups of 14) from the University of Central Arkansas volunteered for the experiment in exchange for class credit. None had participated in Experiment 1.

6.1.2. Procedure

The stimuli and conditions were the same as in Experiment 1. The only difference was that participants were instructed to search for the physically small item in two conditions and the physically large item in the other two conditions. The instructions did not mention that the target had a unique numerical size.

6.2. Results

Mean error rates were submitted to a four-way ANOVA with numerical size and physical size as between-subjects factors, and number of display items and target digit as within-subjects factors. Accuracy was > 98% in all conditions. None of the main effects and none of the interactions were significant, all *ps* > .05.

Mean correct RTs (depicted in Fig. 3) were submitted to a four-way ANOVA with numerical size and physical size as between-subjects factors, and number of display items and target digit as within-subjects factors. As in Experiment 1, RTs increased with the number of display

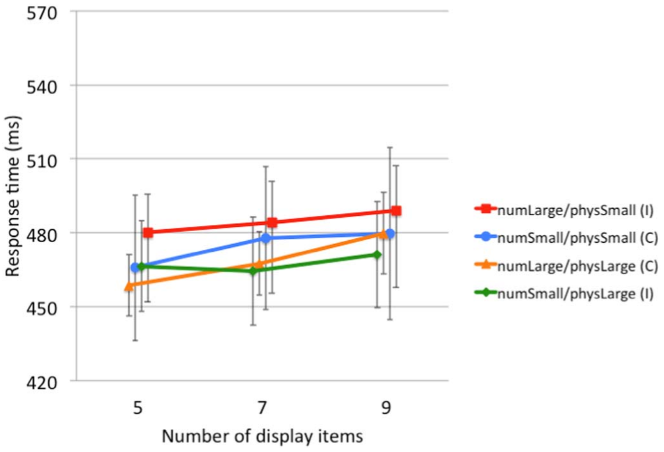


Fig. 3. Response time as a function of number of display items in Experiment 2. Error bars represent standard error of the mean.

items, $F(2, 104) = 16.3$, $MSE = 4116.75$, $p < .001$, $\eta_p^2 = .24$. The effect size of the number of display items was smaller than in Experiment 1 ($\eta_p^2 = .35$), and the mean slopes of RT as a function of number of display items (3.02 ms per item) were shallower than in Experiment 1 (5.89 ms per item), indicating that in both experiments, the target popped out from the distractors.

None of the main effects or two-way interactions were significant, all $ps > .05$, but some unexpected three-way interactions were significant. An interaction between numerical size, physical size, and target digit, $F(1, 52) = 11.50$, $MSE = 2060.59$, $p = .001$, $\eta_p^2 = .18$, indicates that responses were faster for target digits 3, 8, and 9 when they were physically large than when physically small, but for target digit 2, RTs were the same when it was physically large as when physically small. This effect appears to underlie two additional three-way interactions, one between numerical size, number of display items, and target digit, $F(2, 104) = 5.41$, $MSE = 1121.00$, $p = .006$, $\eta_p^2 = .094$, and another between numerical size, physical size, and number of display items, $F(2, 104) = 3.78$, $MSE = 948.86$, $p = .026$, $\eta_p^2 = .068$. As can be seen in Fig. 3, RTs increase with the number of display items for both numerically large conditions, but for the numerically small conditions, RTs were flatter for the physically large condition than the physically small condition, because RTs increased with number of display items for numerically large targets (8 and 9), and also for the numerically small target when it was 3, but remained flat when it was 2.

As mentioned above, all of these three-way interactions were unexpected, and as they seem to be driven by a single anomaly (i.e., RTs did not vary with physical size for target digit 2 but did for all the other target digits), they are not particularly interesting in and of themselves. What may be interesting, however, is the possibility that the target's shape influenced search behavior in Experiment 2. This *does* fit with our expectation that instructing participants to attend to physical size, a perceptual attribute, discouraged them from checking the target's numerical size. That is, these interactions with target digit as a factor suggest that participants processed the target digit as a shape rather than as a symbol associated with some numerical size.

6.3. Discussion

We hypothesized that if participants were reassured that the physical size singleton was always the target, they would feel comfortable responding to it without bothering to check its numerical size. This hypothesis was supported by the results from Experiment 2, in which participants were instructed to find the target with a specific physical size, thereby abolishing the SCE. Furthermore, the three-way interactions with target digit as a factor were unexpected, but they suggest that participants viewed the search items as abstract shapes rather than as symbols associated with some numerical size. The presence of the SCE when participants were instructed to attend to numerical size, as in Experiment 1, coupled with the lack of the SCE when participants were instructed to attend to physical size, as in Experiment 2, replicates the same pattern from Experiments 4 and 5 in Sobel et al. (2016). Furthermore, this pattern is consistent with a reverse Stroop effect (attend to semantic feature) that is larger than a Stroop effect (attend to perceptual feature) in localization (Blais & Besner, 2007).

Although the combined results from Experiments 1 and 2 are consistent with previous studies, the null SCE in Experiment 2 itself appears, at least initially, to be a failure to replicate the SCE in Krause et al. (2017). In that study, participants were instructed to attend to physical size, and searched for a physical size singleton that had either a congruent or incongruent numerical size. Although there were several differences in the apparatus and stimuli between their experiment and ours, we do not believe that these implementation-level differences can explain our apparent failure to replicate their SCE. After all, in Sobel et al. (2016, Experiment 2) we used the same apparatus, stimuli, and instructions (attend to physical size) as we did here, and yet we obtained a robust SCE in that experiment. We believe that the relevant difference is that in Krause et al. (as well as Sobel et al., 2016), congruent trials were randomly interleaved with incongruent trials. Extending the results from our Experiments 1 and 2 to Krause et al. implies that 1) if they had included a condition in which they instructed participants to attend to numerical size, the resulting SCE would be larger than the one they observed in their experiment, in which they instructed participants to attend to physical size, and 2) if they had included conditions in which they manipulated congruence between subjects, the SCE should be smaller in both instruction conditions than in their experiment, in which they manipulated congruence within subjects.

The lack of the SCE in Experiment 2 suggests that the target's numerical size did not intrude on the participants' decision to respond to the physical size singleton as the target. Is it possible to prevent semantic processing of the physical size singleton even when participants are instructed to attend to numerical size, as in Experiment 1? To answer this, consider the results from Experiment 1 in light of Wolfe's (2012) claim that search for multiple targets is a kind of hybrid search, because it entails both a search through the visual field as well as through the target template. In visual searches for multiple targets, once a visual item is attended, the participant must then search through the target template to determine if the attended item matches any of the target items. Because numbers are commonly presumed to be arranged along a mental number line (Feigenson, Dehaene, & Spelke, 2004; Pinhas, Pothos, & Tzelgov, 2013), visual search for digits is faster and more efficient when the targets are adjacent to each other on the number line than when they are separated (Sobel et al., 2015). Thus, by instructing participants to search for multiple target digits, we believe participants in Experiment 1 encoded the target digits' number line positions in their target templates. If participants had instead searched for a single target digit, a template containing the target's shape might be simpler than a template containing the target's number line position. To test this hypothesis, in Experiment 3, participants were instructed to attend to numerical size as in Experiment 1, but the presentation of the target digits (i.e., 2 and 3 in the small numerical size conditions, 8 and 9 in the large numerical size conditions) was blocked, so just one of the two possible target digits appeared in each half of the experiment.

7. Experiment 3: Search for a numerical size singleton with blocked trial order

7.1. Method

7.1.1. Participants

A total of 56 undergraduate students (4 groups of 14) from the University of Central Arkansas volunteered for the experiment in exchange for class credit. None had participated in Experiments 1 or 2.

7.1.2. Procedure

The stimuli, conditions, and instructions were the same as in Experiment 1. The only difference was that the presentation of each of the two possible target digits in each numerical size condition was blocked rather than randomly interleaved. Thus, throughout the first half of the experiment, half of the participants in the numerically small target condition were presented with displays in which the target was always the digit 2, then in the second half of the experiment the target was always the digit 3. Block order was counterbalanced across participants. The same was true for target digits 8 and 9 in the numerically large conditions.

7.2. Results

Mean error rates were submitted to a five-way ANOVA with numerical size, physical size, and block order as between-subjects factors, and number of display items and target digit as within-subjects factors.

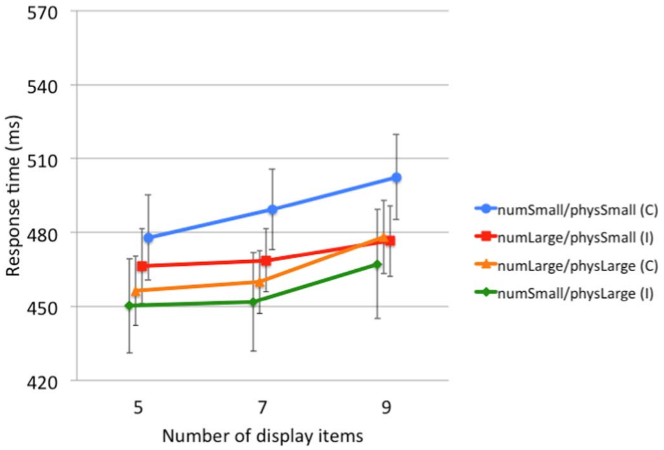


Fig. 4. Response time as a function of number of display items in Experiment 3. Error bars represent standard error of the mean.

Accuracy was > 98% in all conditions. The interaction between block order and target digit, $F(1, 48) = 4.63$, $MSE = 7.44$, $p = .036$, $\eta_p^2 = .088$, is evidence of a practice effect; error rates were higher for the target digit viewed in the first block than for the target digit viewed in the second block. None of the other effects were significant, all $ps > .05$.

Mean correct RTs (depicted in Fig. 4) were submitted to a five-way mixed ANOVA with numerical size, physical size, and block order as between-subjects factors, and number of display items and target digit as within-subjects factors. As in the previous experiments, RTs increased with the number of display items, $F(2, 104) = 30.6$, $p < .001$, $\eta_p^2 = .37$. The effect size of the number of display items was about the same as in Experiment 1 ($\eta_p^2 = .35$), and the mean slopes of RT as a function of the number of display items (4.76 ms per item) were also about the same as in Experiment 1 (5.89 ms per item). None of the other effects were significant, all $ps > .05$.

7.3. Discussion

As expected, repeatedly presenting the same target digit throughout an entire block of trials encouraged participants to load the target template with the target's shape rather than its numerical value or position on the number line, thereby abolishing the interaction between numerical and physical size. The unexpected three-way interactions in Experiment 2 suggested that participants who were instructed to attend to physical size may have processed the target digits as abstract shapes rather than as symbols associated with some numerical size. If so, then effects with target digit as a factor could have been expected in this experiment, because blocking the presentation of the target digit encouraged participants to load the target digit's shape into their target template. The lack of any such effects in Experiment 3 suggests that the three-way interactions in Experiment 2 were merely anomalous.

8. General discussion

Early attempts to discover how an alphanumeric character's meaning can influence visual search yielded mixed results. For example, Jonides and Gleitman (1972) found that search for an ‘O’ target was affected by whether participants thought the target was the letter (oh) or the number (zero), but Duncan (1983) failed to replicate this effect. Krueger (1984) argued that the role of a character's meaning in visual search could be explained more parsimoniously in terms of its shape, which led to a two decade dry spell that ended only when Lupyan (2008; Lupyan & Spivey, 2008) met Krueger's challenge by carefully controlling characters' shapes. This opened the door for other researchers to develop their own techniques to show how a character's meaning influences visual search (Godwin et al., 2014; Schwarz & Eiselt, 2012; Sobel et al., 2015). After Risko et al. (2013) showed that visual attention plays a key role in the SCE in traditional numerical comparison tasks, it was only a matter of time before the numerical comparison task was adapted to a visual search paradigm (Krause et al., 2017; Sobel et al., 2016). Notwithstanding the results from these studies which revealed an SCE in visual search, the presence of the SCE in visual search is surprising, in light of the claim that numerical size is not a guiding feature in visual search (Sobel et al., 2017; Wolfe & Horowitz, 2004).

To explain the presence of the SCE in visual search given the dubious status of numerical size as a guiding feature, we developed a hypothesis that extended on Risko et al. (2013): when searching for a target that is distinct from distractors due to its unique numerical and physical size, the target's physical size first captures attention, and only then does incongruent numerical size interfere with deciding whether the attended item is the target. Here we sought to explore the role of the target template in determining whether congruity between a target's numerical and physical size affects visual search.

In a previous study that instructed participants to attend to the target's numerical size in one experiment and its physical size in another (Sobel et al., 2016), physical size varied randomly between trials. The resulting inter-trial interference precluded participants from creating a target template in which the target's physical size could be precisely defined. In Experiment 1 here, the target's numerical and physical size were manipulated between subjects so each participant was exposed to just one numerical size and one physical size condition. Search slopes were shallow and no different between congruent and incongruent conditions, indicating that the target's physical size captured attention, and the interference from incongruent numerical size was a fixed cost rather than a processing cost per search item. The SCE in Experiment 1 replicated Experiment 1 in Sobel et al. (2016), and as expected, the effect size was smaller than when numerical and physical size were manipulated within subjects as in Sobel et al. Although some of the reduction in effect size could be attributed to variability between subjects, it nevertheless lent preliminary support to our hypothesis that a target with a predictable physical size should reduce participants' incentive to check the attended item's numerical size to make certain it is the target. Supplementary support for this hypothesis was provided by the results from Experiment 2.

In Experiment 2, instructing participants to attend to the target's physical size eliminated the SCE. A larger SCE when participants attended to numerical size rather than physical size replicates Sobel et al., and is broadly consistent with reverse Stroop effects (attend to semantic meaning) that are larger than Stroop effects (attend to perceptual color) in localization tasks (Blais & Besner, 2007). We hypothesized that the SCE was abolished in Experiment 2 because participants didn't submit the physical size singleton to semantic processing, and wondered if semantic processing could be prevented even when participants were instructed to attend to numerical size, as in Experiment 1.

To find out, participants in Experiment 3 were instructed to attend to the target's numerical size as in Experiment 1, but the presentation order of the two target digits (i.e., 2 and 3 in the small numerical size conditions, 8 and 9 in the large numerical size conditions) was blocked rather than randomly interleaved across trials. Because the target's physical size and shape were predictable, participants could load the target template with the two purely visual features of size and shape. The lack of an SCE in Experiment 3 suggests that visual search for multiple digits requires participants to load the target template with the targets' number line position, but search for a single digit allows participants to load the target template with the target's shape.

Here we must acknowledge a possible alternative reason that the target digits may have been divorced from their meaning in Experiment 3. A common experience when people verbally repeat any meaningful word several times is that the word eventually seems to lose its meaning, an effect called semantic satiation (see Balota & Black, 1997, for a brief review). Experiment 3 was intended to show that visual search for a single digit throughout the block encourages participant to load the target template with the target's shape independent of its meaning, but the repeated presentation of one target digit could also have disconnected it from its meaning due to semantic satiation. Although semantic satiation is a possible explanation for why digits were divorced from their semantic associations, we don't believe that semantic satiation as an alternative hypothesis is particularly devastating to the conclusions we draw from Experiment 3. First of all, semantic satiation is primarily an auditory phenomenon whereas our experiments were visual. And second, even if semantic satiation does extend into the visual realm, it is not mutually exclusive with our claim that repeated exposure to the same digit encouraged participants to load the target digit's shape rather than its meaning into the target template. The current findings do not allow us to distinguish between the two explanations, but future studies could be designed to address this issue.

9. Conclusions

Along with Wolfe and Horowitz (2004), we have long doubted that semantic associations could influence visual search. For that reason, we never expected to find the SCE in visual search, and were shocked when we did (Experiment 1 in Sobel et al., 2016). In a series of four follow-up experiments, we sought to abolish the SCE in visual search by emphasizing the target's physical size within each search display. The SCE in visual search proved to be surprisingly robust, as we failed to abolish the SCE in two out of the four follow-up experiments. Here we extended on those experiments by looking at whether variability across, rather than within, trials influences the precision of the target template. In Experiment 1, we eliminated the inter-trial variability from Sobel et al. (2016) by manipulating the four target size conditions between subjects. This reduced, but did not abolish the SCE in visual search. In Experiments 2 and 3, we exposed participants to conditions that we expected to discourage them from submitting the physical size singleton to semantic processing. Would the SCE be larger if we exposed participants to conditions intended to *encourage* semantic processing? We look forward to investigating this question in the future.

References

Algom, D., & Fitousi, D. (2016). Half a century of research on Garner interference and the separability–integrality distinction. *Psychological Bulletin, 142*(12), 1352–1383. http://dx.doi.org/10.1037/bul0000072.

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

Arend, I., & Henik, A. (2015). Choosing the larger versus choosing the smaller: Asymmetries in the size congruity effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(6), 1821–1830. http://dx.doi.org/10.1037/xlm0000135.

Balota, D. A., & Black, S. (1997). Semantic satiation in healthy young and older adults. *Memory & Cognition, 25*(2), 190–202. http://dx.doi.org/10.3758/BF03201112.

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.

Beck, V. M., Hollingworth, A., & Luck, S. J. (2011). Simultaneous control of attention by multiple working memory representations. *Psychological Science, 23*, 887–898. http://dx.doi.org/10.1177/0956797612439068.

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia, 17*(5), 467–472. http://dx.doi.org/10.1016/0028-3932(79)90053-8.

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

Blais, C., & Besner, D. (2006). Reverse Stroop effects with untranslated responses. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1345–1353. http://dx.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. http://dx.doi.org/10.3758/BF03194090.

Blais, C., Harris, M. B., Guerrero, J. V., & Bunge, S. A. (2012). Rethinking the role of automaticity in cognitive control. *The Quarterly Journal of Experimental Psychology, 65*(2), 268–276. http://dx.doi.org/10.1080/17470211003775234.

Braun, J. (1994). Visual search among items of different salience: Removal of visual attention mimics a lesion in extrastriate area V4. *The Journal of Neuroscience, 14*, 554–567 (doi: 10.1.1.655.8147).

Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics, 51*, 465–472. http://dx.doi.org/10.3758/BF03211642.

Cohen, D. J. (2009). Integers do not automatically activate their quantity representation. *Psychonomic Bulletin & Review, 16*, 332–336. http://dx.doi.org/10.3758/PBR.16.2.332.

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. http://dx.doi.org/10.1037/a0012780.

Duncan, J. (1983). Category effects in visual search: A failure to replicate the “oh-zero” phenomenon. *Perception & Psychophysics, 34*(3), 221–232. http://dx.doi.org/10.3758/BF03202949.

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

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307–314. http://dx.doi.org/10.1016/j.tics.2004.05.002.

Godwin, H. J., Hout, M. C., & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review, 21*, 689–695. http://dx.doi.org/10.3758/s13423-013-0547-4.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition, 10*(4), 389–395. http://dx.doi.org/10.3758/BF03202431.

Hout, M. C., & Goldinger, S. D. (2015). Target templates: The precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics, 77*(1), 128–149. http://dx.doi.org/10.3758/s13414-014-0764-6.

Jiménez, L., & Méndez, A. (2013). It is not what you expect: Dissociating conflict adaptation from expectancies in a Stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 271–284. http://dx.doi.org/10.1037/a0027734.

Jonides, J., & Gleitman, H. (1972). A conceptual category effect in visual search: O as letter or as digit. *Perception & Psychophysics, 12*(6), 457–460. http://dx.doi.org/10.3758/BF03210934.

Kiss, M., & Eimer, M. (2011). Attentional capture by size singletons is determined by top-down search goals. *Psychophysiology, 48*, 784–787. http://dx.doi.org/10.1111/j.1469-8986.2010.01145.x.

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

Krueger, L. E. (1984). The category effect in visual search depends on physical rather than conceptual differences. *Perception & Psychophysics, 35*, 558–564. http://dx.doi.org/10.3758/BF03205953.

Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition, 108*, 566–577. http://dx.doi.org/10.1016/j.cognition.2008.03.009.

Lupyan, G., & Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology, 18*, R410–R412. http://dx.doi.org/10.1016/j.cub.2008.02.073.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163–203. http://dx.doi.org/10.1037/0033-2909.109.2.163.

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. http://dx.doi.org/10.1080/10447318.2016.1150642.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215*(5109), 1519–1520. http://dx.doi.org/10.1038/2151519a0.

Nothdurft, H. C. (2006). Salience-controlled visual search: Are the brightest and the least bright targets found by different processes? *Visual Cognition, 13*, 700–732. http://dx.doi.org/10.1080/13506280544000237.

Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Science, 15*, 327–334. http://dx.doi.org/10.1016/j.tics.2011.05.004.

Pinhas, M., Pothos, E. M., & Tzelgov, J. (2013). Zooming in and out from the mental number line: Evidence for a number range effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 39*, 972–976. http://dx.doi.org/10.1037/a0029527.

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. http://dx.doi.org/10.3758/s13414-013-0477-2.

Schwarz, W., & Eiselt, A. K. (2012). Numerical distance effects in visual search. *Attention, Perception, & Psychophysics, 74*, 1098–1103. http://dx.doi.org/10.3758/s1314-012-0342-8.

Sobel, K. V., Pickard, M. D., & Acklin, W. T. (2009). Using feature preview to investigate the roles of top-down and bottom-up processing in conjunction search. *Acta Psychologica, 132*, 22–30. http://dx.doi.org/10.1016/j.actpsy.2009.06.003.

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. http://dx.doi.org/10.3758/s13414-016-1098-3.

Sobel, K. V., Puri, A. M., Faulkenberry, T. J., & Dague, T. D. (2017). Visual search for conjunctions of physical and numerical size shows that they are processed independently. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 444–453. http://dx.doi.org/10.1037/xhp0000323.

Sobel, K. V., Puri, A. M., & Hogan, J. (2015). Target grouping in visual search for multiple digits. *Attention, Perception, & Psychophysics, 77*, 67–77. http://dx.doi.org/10.3758/s13414-014-0761-9.

Todd, S., & Kramer, A. F. (1994). Attentional misguidance in visual search. *Perception &*

Psychophysics, 56, 198–210. http://dx.doi.org/10.3758/BF03213898.

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

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

Wolfe, J. M. (2012). Saved by a log: How do humans perform hybrid visual and memory search? *Psychological Science, 23*, 698–703. http://dx.doi.org/10.1177/0956797612443968.

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. http://dx.doi.org/10.1038/nrn1411.

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