



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Goal-neglect links Stroop interference with working memory capacity

**Citation for published version:**

Morey, CC, Elliott, EM, Wiggers, J, Eaves, SD, Shelton, JT & Mall, JT 2012, 'Goal-neglect links Stroop interference with working memory capacity', *Acta Psychologica*, vol. 141, no. 2, pp. 250-260.  
<https://doi.org/10.1016/j.actpsy.2012.05.013>

**Digital Object Identifier (DOI):**

[10.1016/j.actpsy.2012.05.013](https://doi.org/10.1016/j.actpsy.2012.05.013)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Acta Psychologica

**Publisher Rights Statement:**

© Morey, C. C., Elliott, E. M., Wiggers, J., Eaves, S. D., Shelton, J. T., & Mall, J. T. (2012). Goal-neglect links Stroop interference with working memory capacity. *Acta Psychologica*, 141(2), 250-260.  
[10.1016/j.actpsy.2012.05.013](https://doi.org/10.1016/j.actpsy.2012.05.013)

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



PLEASE CITE:

Morey, C. C., Elliott, E. M., Wiggers, J., Eaves, S. D., Shelton, J. T., & Mall, J. T. (2012). Goal-neglect links Stroop interference with working memory capacity. *Acta Psychologica, 141*(2), 250–260.

Goal-neglect links Stroop interference with working memory capacity

Candice C. Morey

University of Groningen

Emily M. Elliott

Louisiana State University

Jody Wiggers

University of Groningen

Sharon D. Eaves & Jill T. Shelton

Louisiana State University

Jonathan T. Mall

University of Groningen

Author's Note

The research described in this manuscript was completed in partial fulfillment of the M.A. degree requirements of Jody Wiggers. Thanks to Gerdien Boven, Malte Bieler, Rieka Buggentien, Andrej Krakau, Irene Krap, Helene Meißner, Anne Muth, Sophia Oelmann, Nelly Rahmede, and Michael Wolff for assistance with data collection. Jill T. Shelton is now at the Department of Behavioral and Social Sciences, Lee University, Cleveland, TN and Sharon Eaves is now at the Department of Social Sciences, Shawnee State University, Portsmouth, OH. Correspondence should be addressed to Candice C. Morey, Experimental Psychology, Grote Kruisstraat 2/1, 9712TS Groningen, The Netherlands. Email: [c.c.morey@rug.nl](mailto:c.c.morey@rug.nl).

Word count: 9,530 (36 references)

Date of original submission: 15 July, 2011

Date of resubmission: 19 January, 2012

Date of 2<sup>nd</sup> resubmission: 16 May, 2012

## Abstract

Relationships between Stroop interference and working memory capacity may reflect individual differences in resolving conflict, susceptibility to goal neglect, or both of these factors. We compared relationships between working memory capacity and three Stroop tasks: a classic, printed color-word Stroop task, cross-modal Stroop, and a new version of cross-modal Stroop with a concurrent auditory monitoring component. Each of these tasks showed evidence of interference between the semantic meaning of the color word and the to-be-named color, suggesting these tasks each require resolution of interference. However, only Stroop interference in the print-based task with high proportions of congruent trials correlated significantly with working memory capacity. This evidence suggests that the relationships observed between Stroop interference and working memory capacity are primarily driven by individual differences in the propensity to actively maintain a goal.

Word count: 132

Keywords: working memory, Stroop, executive control, goal-neglect, selective attention

## Goal-neglect links Stroop interference with working memory capacity

Anyone who has experienced the sensation of arriving in one room and forgetting the reason for entering it understands that in the space of mere seconds, it is possible to forget what one is doing, or to forget the goal currently motivating behavior. This phenomenon is thought to be one factor underlying Stroop interference (Kane & Engle, 2003), in addition to the competition that arises from the conflict between perceiving some color and reading the name of another color: confronted with two operations that might be performed on the stimulus, participants forget to name the ink color and read the word instead. Thus, both resolving conflict arising from the irrelevant information and maintaining activation of goals contributes to successful selective attention performance. However, working memory capacity, which is believed by many to reflect ability to control attention (e.g., Cowan, Elliott, Saults, Morey, Mattox, Hismjatullina, & Conway, 2005; Kane, Bleckley, Conway, & Engle, 2001) does not always predict ability to ignore irrelevant information as expected. We aim to test the hypotheses about attention and goal neglect raised by Kane and Engle, and thereby clarify expectations about relationships between working memory capacity and selective attention using a variety of Stroop tasks to measure selective attention in contexts varying in their propensity to induce goal neglect and in the ease of goal recovery following an attentional lapse.

Kane and Engle (2003) compared groups of participants with extremely low or high working memory capacity on two versions of Stroop task, one in which the ink color and word were congruent on the majority of trials, and one in which the ink color and word were rarely or never congruent. They observed large differences in error rates between low- and high-capacity extreme groups as a function of the proportion of

congruent Stroop trials: when most trials included congruent ink colors and words, the low-capacity individuals committed significantly more errors than the high-capacity individuals (consistently with the expectations of Long and Prat (2002), that low-capacity individuals will not allocate much attention to maintaining the task goal), but when the ink color and word were usually incongruent, low- and high-capacity individuals did not differ in error rates. Kane and Engle suggested that the rarely-congruent context supported the ink-color naming goal best because reading the word would usually result in an error whereas in the frequently-congruent context, errors were rarely a consequence for reading the word because doing so usually produced the correct response. The group differences they uncovered suggested that this contextual support for goal maintenance was critical only for the low-capacity individuals; high-capacity individuals seemed to maintain the goal of naming the ink color even in the less supportive context, a finding which has been confirmed (Hutchison, 2011). This proposition was also supported by reductions in errors by low-capacity individuals when they completed the frequently-congruent block after the rarely-congruent block. Kane and Engle suggested that individual differences in selective attention, as measured by Stroop interference, depend on individual differences in maintaining the task goal, perhaps in addition to individual variability in resolving the response competition evoked by task-irrelevant information.

This distinction between maintaining a task goal and attending to some stimulus while excluding another may potentially explain the perplexing relationships observed between working memory capacity and the ability to ignore distracting information. Controlled attention views of working memory capacity predict that individuals with high working memory capacity will more effectively exclude task-irrelevant information from

mind than individuals with low working memory capacity. This has been demonstrated in dichotic listening, with low-capacity individuals noticing their name in the unattended channel more frequently than high-capacity individuals (Conway, Cowan, & Bunting, 2001). Visual change detection research also suggests that memory capacity and ability to filter task-irrelevant distractors are correlated (Vogel, McCullough, & Machizawa, 2005). However, evidence from two logically similar methods, the irrelevant sound paradigm and the cross-modal Stroop paradigm, is mixed or contradictory. In both of these paradigms, robust impairing effects of auditory distractors on performance are observed, but these effects do not consistently correlate with working memory capacity.

A critical difference between selective attention tasks that correlate with working memory capacity and selective attention tasks that do not could be the level of support for goal recovery offered by the task's context. In the irrelevant sound paradigm, serial memory performance is compared for visually-presented lists that are delivered in conditions of silence or during the auditory presentation of irrelevant words or tones. Elliott and Cowan (2005) examined relationships between memory span and the difference between list recall in silence or with irrelevant sounds, and only sometimes observed small correlations between working memory span and irrelevant sound effects (ISEs). Beaman (2004) examined relationships between operation span scores and the size of an ISE on memory for lists of digits or words, and found no difference between the size of the ISE for individuals with high versus low working memory capacity; in fact, the average ISE for high-capacity individuals was sometimes numerically larger than that for low-capacity individuals. In Beaman's series of studies, irrelevant speech only differentially impaired recall performance of low-capacity individuals when the irrelevant

words were semantically related to the memoranda; under this circumstance, low-capacity individuals were more likely to recall a related, spoken item as though it were one of the studied memoranda. Possibly, when the lists contained semantically related words, keeping track of which list should be attended required more attention. Elliott, Barrilleaux, and Cowan (2006) showed using regression that the presence of irrelevant sounds affected low-capacity participants more than high-capacity participants. Even so, this significant result held only for operation span and not for running span, which should also measure working memory capacity, and the proportions of variance that could be accounted for by this relationship were small. Thus, the costs of resolving conflict between irrelevant sounds and memoranda do not show a consistent pattern of relationships with working memory capacity, despite predictions that they should.

Importantly, there is currently no evidence suggesting that individual differences in working memory affect Stroop interference as measured by cross-modal Stroop tasks. In cross-modal Stroop tasks, participants name the color of a square while distracting sounds are presented (e.g., Cowan & Barron, 1987; Elliott & Cowan, 2001; Elliott, Cowan, & Valle-Inclan, 1998). When the distracting sound is a color word whose meaning is incongruent with the color of the to-be-named square, significant slowing in color naming is observed compared with color-naming speed with no distracting sound or with a congruent color word. Although this slowing is not as large as that observed in printed-word Stroop tasks, cross-modal Stroop produces scale changes between incongruent and neutral response time distributions similar to those observed with print Stroop (Elliott, Morey, Morey, Eaves, & Shelton, in prep.). This is consistent with the idea that in both the print and cross-modal Stroop tasks, slowing occurs because of

conflict between the semantic meaning of the irrelevant word and the goal to name the color. However, no previous study measuring cross-modal Stroop interference and working memory capacity has uncovered any evidence that cross-modal Stroop interference and working memory capacity are negatively correlated (Elliott et al., 2006), a prediction that would follow logically from the supposition that working memory tasks measure one's ability to effectively cope with interference. Moreover, facilitation of congruent words on color naming is rarely observed during cross-modal Stroop, though such facilitation is consistently found in print-based Stroop tasks, even when the to-be-named color and the to-be-ignored word are spatially (MacLeod, 1998; Spieler, Balota, & Faust, 2000) or temporally (Roelofs, 2010) separated. Although facilitation may in part occur because of converging information (see Roelofs, 2010), facilitation may also reflect trials on which participants forgot that their goal was to name the color and instead quickly read the word (MacLeod & MacDonald, 2000). The absence of facilitation in cross-modal Stroop may suggest that goal recovery is less of a problem during cross-modal than print-based Stroop, but this task difference has not previously been examined.

The absence of any relationship between cross-modal Stroop interference and working memory capacity and the inconsistency of relationships between the size of ISEs and working memory capacity is surprising if one believes that working memory capacity is critically related to controlling attention, and that controlling attention is necessary for efficiently resolving conflicts between attended and irrelevant information. However, while performance in dichotic listening, cross-modal Stroop, and print-based Stroop tasks reflects some level of interference resolution, these tasks may differ in the level of contextual support typically provided for maintaining the appropriate task goal. Consider



the situation presented in the cross-modal Stroop paradigm, supposing that with some frequency, participants' attention will lapse and they will momentarily forget the goal to name colors. In this scenario, the participant recovers from an attentional lapse, focuses on the screen and sees a colored square. It is difficult to imagine what the participant would consider doing with this colored square besides naming the color; no competitive alternative, such as inadvertently reading the word in print-based Stroop, is apparent. Cross-modal Stroop is logically similar in this respect to spatially-separated Stroop paradigms. Evidence suggests that compared to traditional administrations of print-based Stroop, spatially-separated Stroop offers a context more supportive for reducing the impact of goal neglect on Stroop interference, reducing the incidence of very slow responses that might reflect attentional lapses rather than only interference resolution processes (Kane & Engle, 2003; Spieler, Balota, & Faust, 2000). Cross-modal Stroop is an even more drastic example, because there is no visually-presented word at all to be read. If the variation that is shared between print-based Stroop and working memory capacity reflects differences in the tendency to forget that the goal is to name the ink color in addition to individual differences in the ability to resolve conflict between the semantic meaning of the ink color and the intended word response, then the absence of any relationship between cross-modal Stroop and working memory capacity is not so surprising. One may attribute the absence of this relationship to the negligible consequences of goal neglect in cross-modal Stroop, where the context supports the quick recovery of the task goal and provides no plausible alternative goal that may provoke errors.

However, this explanation of the absence of a relationship between cross-modal Stroop and working memory capacity is not the only plausible one. Print-based and cross-modal Stroop tasks also differ in that attending to the printed word is obligatory in print-based Stroop, whereas attending to the aurally-presented information is discouraged in cross-modal Stroop. On each and every trial in print-based Stroop, participants must look at the word in order to extract and name the ink color, whereas in cross-modal Stroop, participants are instructed to ignore the sounds as much as possible and focus only on the colors onscreen. Also, although the slowing observed on incongruent trials in cross-modal Stroop suggests that interference occurs in the cross-modal as well as print-based versions, there is much less slowing in cross-modal than in print-based Stroop. It is therefore possible that in cross-modal Stroop, there is insufficient variance in the size of interference effects to detect correlations with working memory capacity.

We therefore set out to compare performance on three Stroop tasks that varied in the manner in which potentially interfering stimuli were attended and the ease with which the task goal could be recovered when lost. In Experiment 1, we compared performance on a classic, print-based Stroop task (hereafter called *classic Stroop*) with performance on a cross-modal Stroop task previously used by Elliott and colleagues (e.g., Elliott et al., 2006; hereafter *CM Stroop*). In these administrations, we included no congruent color trials, so that all participants would always need to actively maintain the goal to name colors in order to avoid making errors. Individual differences in active goal maintenance would therefore be minimized in this task context.

In Experiment 2, we repeated these tasks with a high proportion of congruent trials on a new sample of participants. With a high proportion of congruent trials,

individual variation in active goal maintenance may emerge because on the majority of trials, a correct response may occur regardless of whether the color-naming goal is active. This context should allow participants who are more prone to attentional lapses to neglect the explicit task goal. On the rare incongruent trials however, participants actively maintaining the color-naming goal should show fewer errors or less slowing.

We also conducted a new version of the cross-modal Stroop task, in which participants were instructed to monitor the sounds for a rarely-occurring target while naming the colors of the squares (*CM+ Stroop*). We designed the *CM+* Stroop task to include several important features that we think increase its comparability to classic Stroop. Although participants are instructed to ignore the meaning of the printed words in classic Stroop, they must attend to the printed words in order to name the ink color, whereas in *CM* Stroop, participants are instructed to ignore the sounds completely. In our *CM+* Stroop, participants must pay some attention to the spoken words, as they must pay some attention to the printed words in classic Stroop, thereby increasing the potential for semantic interference between the presented words and appropriate responses compared to *CM* Stroop tasks in which the auditory information is to be ignored. Merely increasing the size of the difference between latencies of color-naming in incongruent and neutral trials might allow correlations with working memory capacity to emerge in the *CM+* task. However, although *CM+* Stroop should provoke more interference than *CM* Stroop, we do not think that *CM+* Stroop presents a comparable risk for goal neglect as classic Stroop, because as in *CM* Stroop, it is not possible to inadvertently read a response. Error rates in *CM* Stroop are typically very low, suggesting that participants are not prone to

inadvertently repeating the aurally-presented words as responses to the color square stimuli.

In both experiments, we planned to compare correlations between Stroop interference effects and working memory capacity, as measured by automated Operation and Symmetry span tasks (Unsworth, Heitz, Schrock, & Engle, 2005). Although their experimental design exploited extreme group differences, we expected based on Kane and Engle's (2003) work that classic Stroop interference should negatively correlate with working memory capacity in a situation with low contextual support in which errors were discouraged. To the extent that CM+ Stroop better mimics the task demands of classic Stroop than other versions of CM Stroop, we expect CM+ Stroop to share variance with classic Stroop and working memory capacity, if the shared variance between classic Stroop and working memory capacity primarily reflects individual differences in resolving interference. However, if the shared variance between classic Stroop and working memory capacity mainly reflects individual differences in actively maintaining a task goal, then 1) correlations between working memory capacity and classic Stroop interference will only appear in a highly congruent task context of Experiment 2, and not in Experiment 1, which included no congruent trials, and 2) even with mostly congruent trials, the cross-modal Stroop tasks and working memory capacity will not relate to each other in the same manner as classic Stroop and working memory measures, because the ease of goal recovery in the cross-modal Stroop task makes actively maintaining the task goal trivial in cross-modal Stroop in any task context. Furthermore, CM+ Stroop contains a separate measure of goal neglect in the auditory monitoring component of the task, which we examine with respect to individual differences in working memory capacity.

## Experiment 1

**Method**

**Participants.** One hundred sixteen Louisiana State University psychology students participated in this experiment for course credit. All participants gave written consent indicating their willingness to take part in the study. The data of 17 participants were excluded for various reasons (high rates of false starts (i.e., triggering the microphone with non-speech sounds on 10% or more of the trials), program error, or non-completion of all of the tasks), resulting in a final sample of 99 (75 females, 18-28 years old,  $M=19.74$ ,  $SD=1.75$ ). All participants in this sample were native English speakers, who reported normal vision (some with contact lenses or glasses for correction) and normal hearing.

**Design and Materials.** All tasks were run in a within-participants design, using desktop computers and E-Prime software (Schneider, Eschmann, & Zuccolotto, 2002). Each participant performed the working memory screening first, and then was invited to return for a second session. The participants then completed the classic Stroop task and the CM Stroop task, with an unrelated filler task between the two Stroop versions. The order of the two Stroop tasks was counterbalanced ( $N=45$  with the classic task first, and  $N=54$  with the CM task first). Each Stroop task included three colors (red, green, and blue), and required spoken responses. Spoken response latency was recorded with a voice key and microphone, and the identity of the color-word response was recorded in real-time during the session by a trained experimenter.

**Working memory span tasks.** We chose two complex working memory span tasks, one with verbal and one with visual-spatial memoranda, with which to compute a

composite measure of working memory. Our participants completed automated versions of the Operation span and Symmetry span tasks (Unsworth, et al., 2005) in a separate experimental session from the Stroop tasks. In Operation span, participants judged whether a given equation was veridical or not and were given a consonant to remember. This sequence was repeated 3-7 times, and at the end of each sequence, participants attempted to recall the consonants from that trial in order. Similarly, in Symmetry span participants judged whether a visual block pattern was vertically symmetrical and then viewed a spatial location within a grid to remember. After this sequence was repeated 2-5 times, participants tried to recall the locations indicated in the grid in the correct serial order. All of the participants described above maintained at least 85% accuracy on the equation and symmetry judgments in these tasks.

***Classic Stroop task.*** Word objects appeared in the center of the screen, in capital letters in either red, green, or blue text. Three types of color-naming trials were presented: incongruent color words (e.g., RED in blue text), neutral stimuli (e.g., @@@@) and non-color words. Both versions of the Stroop task included these distractor conditions, to allow for comparisons across the two versions in the current study, as well as with other research employing these Stroop tasks (e.g., Elliott et al., in prep.). To select a relatively high-frequency set of non-color words to be comparable to the color word stimuli, the non-color words *short*, *long*, and *big* were chosen from the category of size words.

In order to have comparable numbers of trials contributing to the mean values in each distractor condition, some of the numerous incongruent trials were labeled randomly at run time as filler trials. These were excluded from analyses (as done by Kane and Engle, 2003) so that approximately equal numbers of trials were analyzed in the each

condition of each task. Doing this ensured that differences across conditions could not be attributed merely to a greater chance for the slowest responses to influence the mean when the mean is calculated from a smaller set of responses. We included two blocks of 135 trials, within which 111 incongruent trials were presented within each block but only 24 of these were analyzed. Participants were not aware of these filler trials, as they appeared just as any other trial during the session. The remaining 24 trials of the 135 within each block were divided evenly among the neutral and non-color conditions. Trial order was randomly determined.

***Cross-modal Stroop task (or CM Stroop).*** During the color-naming task, speech sounds (which the participants were instructed to ignore) were played through headphones. The onset of the sounds occurred simultaneously with the onset of a to-be-named colored square. With this timing, Elliott, Cowan, and Valle-Inclan (1998) observed the largest differences in color-naming speed between sound and silence conditions. Colored squares appeared in a color incongruent with the to-be-ignored sound (e.g., a blue square during presentation of “green”). Sounds were recorded and digitized; durations ranged from 400 to 700 ms, and the words were spoken in a male voice. The words in the color word condition were *red*, *blue*, and *green*, and the non-color word condition included *short*, *long*, and *big*, as in the classic Stroop task. The analog to the neutral trials in this version of the task was color-naming trials with a .wav file of generated silence. Participants completed the same number and balance of trials as in the classic Stroop task.

**Procedure.** A trained experimenter supervised each experimental session, beginning by explaining the task to the participant and supervising a short practice

session. Practice trials were always neutral or silent trials. Trials began with a 2000-ms reminder to name the color of the object as quickly and accurately as possible, followed by a 500-ms fixation cross and then the presentation of the to-be-named colored object. The experimenter entered the participant's response, and judged whether the activation of the voice key was triggered by the utterance of a color-word, and not by a cough, inhalation, or other noise. The experimenter also indicated whether s/he made an uncorrected input error when recording the response. As soon as the experimenter was finished, a new trial began. Breaks could be taken throughout if the experimenter delayed his or her responses. Each Stroop task lasted approximately 20 minutes.

## **Results**

For all results,  $p < .05$  was the criterion for declaring statistical significance. In reports of analyses of variance (ANOVAs) whenever the sphericity assumption was violated, the Greenhouse-Geiser correction was applied. Analyses of task order effects were conducted, but no effects of the order in which the Stroop tasks were completed were found. This factor was therefore collapsed in all subsequent analyses. Additionally, any trials coded by the experimenters as false starts were removed from the data, resulting in the removal of 4.71% and 5.13% of the data from CM Stroop and classic Stroop, respectively. We analyzed accuracy and response times of color naming within each Stroop task and correlations between the Stroop tasks and measures of working memory capacity.

**Accuracy analyses.** Proportions of correct responses were entered into a 2-way ANOVA with task (classic, CM) and distractor condition (incongruent color word, non-color word, or neutral/silence) as within-subjects factors. We observed a significant main



effect of task ( $F(1,98)=3.03$ ,  $MSE=.001$ ,  $\eta^2_p=.08$ ); the effect of distractor condition did not quite meet our criterion for statistical significance ( $F(1.98, 194.45)=3.03$ ,  $MSE=.001$ ,  $\eta^2_p=.03$ ,  $p=.05$ ). The main effect was qualified by a significant interaction between these factors ( $F(1.90, 186.56)=11.18$ ,  $MSE=.001$ ,  $\eta^2_p=0.10$ ).

We then conducted separate analyses for each Stroop task. Correct responses did not differ by distractor condition in the CM Stroop task ( $p=.18$ ); however, differences were observed in the classic task, ( $F(2,194.21)=11.22$ ,  $MSE=.001$ ,  $\eta^2_p=.10$ ). In the classic Stroop task, distractor condition significantly affected accuracy, with lower accuracy on incongruent trials ( $M=.97$ ,  $SEM=.002$ ) than on neutral trials ( $M=.98$ ,  $SEM=.003$ ) and the highest accuracy on non-color word trials ( $M=.99$ ,  $SEM=.003$ ).

**Response time analyses.** Before analyzing response times, all incorrect trials were removed from the data. A trimming procedure was used to eliminate outliers; all values below 180 ms were removed, and any response exceeding 3 standard deviations of the distractor condition mean were excluded, for each participant. This resulted in the removal of 1.17% and 3.16% of the data in CM Stroop and classic Stroop, respectively.

A repeated-measures ANOVA with task and distractor condition yielded a significant main effect of distractor condition ( $F(1.96,192.72)=19.14$ ,  $MSE=2655.67$ ,  $\eta^2_p=.16$ ) and a significant interaction ( $F(1.87,183.82)=18.39$ ,  $MSE=3235.62$ ,  $\eta^2_p=.16$ ). The main effect of task was not significant ( $p=.97$ ).

As with the accuracy data, separate analyses were performed for each task to explore the interaction. The effect of distractor condition on response times was significant for both the classic Stroop task, ( $F(1.81,17.79)=6.65$ ,  $MSE=5252.27$ ,  $\eta^2_p=.06$ ) and the CM Stroop task, ( $F(1.64,160.77)=69.74$ ,  $MSE=1374.01$ ,  $\eta^2_p=.42$ ). Follow-up

comparisons were Bonferroni-corrected. In the classic Stroop task, color-naming performance was slowest in the incongruent color word condition ( $M=569$  ms,  $SEM=17.97$ ), slower than in either the neutral ( $M=540$  ms,  $SEM=16.71$ ) or the non-color word conditions ( $M=536$  ms,  $SEM=18.32$ ), which did not differ from each other. These findings suggest that the non-color words chosen for this task served as an additional form of a neutral stimulus. In the CM Stroop task, all three distractor conditions differed significantly, with the fastest responses occurring in the silent condition ( $M=521$  ms,  $SEM=7.97$ ), followed by the color word condition ( $M=548$  ms,  $SEM=7.71$ ), and the non-color word condition ( $M=577$  ms,  $SEM=9.13$ ).<sup>1</sup>

**Correlations with working memory capacity.** We calculated a composite working memory measure from the Operation Span ( $M=60.17$ ,  $SD=12.71$ , range: 6-75, where 75 is the maximum possible score) and Symmetry Span ( $M=30.48$ ,  $SD=7.22$ , range: 8-42, where 42 was the maximum possible score) data. We used the leniently scored total values calculated by Unsworth et al.'s (2005) automated span programs (as recommended by Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005). This measure includes all correct responses within a trial toward the task score, regardless of whether the entire

---

<sup>1</sup> Subsequent analyses suggest that the non-color words were auditory oddballs in this task context (see Schröger & Wolff, 1998). The task set created by the frequent presentation of incongruent color words may have led to the expectation that all auditory distractor stimuli would be colors; thus the presentation of the non-color word stimuli may have caused participants to hesitate even longer on these trials. To investigate this hypothesis, we compared mean response times of the distractor conditions by block of trials. The analyses supported this hypothesis; response times for the non-color word trials were significantly slower in Block 1 ( $M=589$  ms,  $SD=105$ ) compared to Block 2 ( $M=566$  ms,  $SD=87$ ;  $t_{(98)}=2.99$ , Bonferroni-corrected). The other distractor conditions did not show this pattern to the same extent (Bonferroni-corrected  $t$ s each nonsignificant), although there was a trend toward speeding across blocks in the other conditions as well. The results of the cross-modal Stroop task in Experiment 2, which showed no slowing for congruent color words compared to silence, demonstrate that it is not the case that the mere presentation of any sound slows color-naming responses.

list was correctly recalled. We normalized these scores for each task and added them together into a composite value.<sup>2</sup> We tested statistical significance using one-tailed tests, because we believed a priori (based on Kane and Engle, 2003) that any relationship between working memory capacity and Stroop interference should be negative, such that as working memory capacity increases, Stroop interference decreases. To assess Stroop interference, difference scores were calculated for both versions of the Stroop task by subtracting performance in the neutral/silence condition from the incongruent color word condition (classic Stroop effect  $M=30$  ms,  $SD=105$ ; CM Stroop effect  $M=27$  ms,  $SD=43$ ). The correlations between the difference scores and the working memory composite score were not significant for the CM Stroop task ( $r=-.03$ ,  $p>.38$ ), and interestingly, were significant but positive for the classic Stroop task ( $r=.18$ ). However, because the sign of this correlation was opposite of our expectations, we also computed a two-tailed test of the relationship, which was not statistically significant ( $p>.07$ ).

The absence of predicted negative correlations between Stroop and working memory capacity cannot be attributed to low reliability in our Stroop tasks. We calculated split-half reliability by assigning alternating trials to groups and calculating correlations between these halves of each data set. These values, given on the diagonal of Table 1, demonstrate high internal consistency.

## Discussion

Kane and Engle (2003; see also Hutchison, 2011) argued that in a task context with few congruent trials, participants did not vary much in their tendency to neglect the task goal because most trials reinforced the goal, ensuring that it would remain active.

---

<sup>2</sup> We observed a statistically significant correlation between Operation and Symmetry span in our sample,  $r=.493$ .

Across a large sample, we did not find evidence suggesting that larger Stroop interference effects occurred in individuals with lower working memory capacity; if anything, our data tended to suggest the opposite pattern, namely that individuals with higher working memory capacity spent even more time responding to incongruent color trials than lower capacity individuals. Compared with Stroop interference typically observed when congruent color trials are incorporated into the task setting, the mean size of Stroop interference in our sample was quite small, close to the size of the interference effect observed in CM Stroop.

With Experiment 2, we examined the relationships between classic and CM Stroop tasks and working memory capacity when most of the trials in a session were congruent color trials, increasing the probability that the color-naming goal in classic Stroop would be neglected. We included a new version of cross-modal Stroop, the CM+ Stroop task, in which participants were instructed to monitor the sounds presented during the session for occasional targets. In addition to increasing the size of the interference effect we are likely to observe in a cross-modal task, CM+ Stroop provides an additional measure of goal neglect (i.e., the number of valid responses given on the auditory monitoring task), which will enable us to further untangle the patterns observed in the relationships between working memory capacity and selective attention tasks.

## Experiment 2

### **Method**

**Participants.** One hundred and twenty University of Groningen psychology students participated in this experiment in partial fulfillment of course requirements. All participants gave written consent indicating their willingness to take part in the study.

The data of eight participants were excluded from analysis due to colorblindness, according to a brief Ishihara test (Ishihara, 1966), and the data of eight more participants were excluded for various reasons (six exclusions were due to high rates of false starts, i.e., triggering the microphone with non-speech sounds on 10% or more of the trials, one participant indicated non-fluency with all three languages in which we offered the tasks, and one reported a diagnosed hearing problem), resulting in a final sample of 104 (84 females, 18-29 years old,  $M=20.59$ ,  $SD=1.66$ ). All participants in this sample reported normal vision (some with contact lenses or glasses for correction) and normal hearing.

**Materials.** The participants were seated in a small, private room in front of a computer screen. Each participant performed the classic Stroop task and at least one of the CM Stroop and CM+ Stroop and color-naming tasks. In each task version, 60% of the trials were congruent, meaning that whatever the source of potential distracting color information, its semantic content matched the correct color-naming response, while 20% of the trials included no potential distracting content and 20% included color information incongruent with the correct color-naming response. As in Experiment 1, a random subset of the more numerous congruent trials were marked as filler at runtime, and were not analyzed. We offered the Stroop tasks in English, German, and Dutch to our participants, who came from a sample of international students studying psychology in either English or Dutch. Participants were encouraged to take part in the language they understood most easily. For the majority of our sample (89%), the language they chose to participate in was their native language. In all tasks, participants were instructed to name the color of the object appearing onscreen aloud as quickly as possible while avoiding errors.

Participants completed the same automated working memory span tasks (Unsworth et al., 2005) used in Experiment 1.

***Classic Stroop task.*** The Classic Stroop task was similar to that in Experiment 1, except that congruently colored words (e.g., RED in red text) were included in the design as well as incongruently colored words and neutral trials. Noncolor words were not included. On each trial, a .wav file of generated silence (created using the free software Audacity (Oetzmann & Mazzoni, 2011)) was played with the same timing within a trial as a sound would occur in the two cross-modal task versions. All participants completed two 90-trial blocks, for a total of 180 Classic Stroop trials.

***Cross-modal Stroop task (CM Stroop).*** The CM Stroop task was similar to that in Experiment 1, except that congruent trials (e.g., a red square during presentation of a recorded spoken word “red”) were included, along with incongruent and neutral (e.g., silent) trials. Noncolor words were not included. The possible words that could be played were the color words *red*, *blue*, and *green* (in the language chosen by the participant), and were played with approximately equal frequency. A subset ( $N=65$ ) of our sample completed two 90-trial blocks, for a total of 180 trials.

***Cross-modal Stroop task, plus auditory monitoring (or CM+ Stroop).*** Stimuli in this task were the same as stimuli in the CM Stroop task, except that participants were instructed to monitor the sounds they heard for a target sound. Occasionally (on 6 trials interspersed randomly in each of two 96-trial blocks, for a total of 12 trials out of 192), a target sound was played, and participants were to respond by pressing a labeled button on a Psychology Software Tools response box in addition to naming the color of the square. Although a color-naming response was also recorded on these 12 trials, they were

omitted from all analyses of color-naming speed and accuracy. Participants were allowed to respond with their button press before or after naming the color square, up to a limit of 2000 ms after the onset of the sound. A subset ( $N=85$ ) of our sample completed a version of the CM+ Stroop task. Over two planned experiments, we manipulated the nature of the target sound. In our first sample ( $N=46$ ), participants listened for the specific word *gray*. In our second sample, some participants ( $N=21$ ) were instructed to listen for a tone (any of a set including 250 Hz, 475 Hz, or 550 Hz) and others were instructed to listen for a word that was not a color word ( $N=18$ ). Non-color words were selected so that they shared phonological characteristics with the color-word sounds. In English, these were *glue*, *reef*, and *bread*; in Dutch, *gauw*, *roem*, and *brood*, where the color words were *blauw*, *groen*, and *rood*; and in German, *rau*, *grund*, and *boot*, where the color words were *blau*, *grün*, and *rot*. These versions differed in the extent to which semantic processing of the sounds was necessary, but we found no significant differences in the size of Stroop interference (i.e., the difference between response times on incongruent and neutral trials) between these versions of the CM+ task, suggesting that the instruction to pay attention to the sounds at all was sufficient to increase processing of the meaning of the acoustically-presented words. We therefore collapsed across these versions in our reported analyses.

**Procedure.** The procedure was the same as in Experiment 1, except that some participants completed all three Stroop tasks. Each task was presented to the participant in a separate block, with its own short practice session of 6 trials. Practice trials were always neutral or silent trials. All participants completed at least two Stroop tasks, with order determined randomly. Each task was given in two sub-blocks of approximately

equal duration, and participants were given an opportunity for a break after each sub-block of each task.

## Results

As in Experiment 1,  $p < .05$  was the criterion for declaring statistical significance. In reports of analyses of variance (ANOVAs) whenever the sphericity assumption was violated, the Greenhouse-Geiser correction was applied. Trials on which the experimenter indicated that a false start or uncorrected experimenter error occurred as well as trials marked as filler were also excluded. Because this is the first report of a cross-modal Stroop task with an auditory monitoring component, we begin our analyses by comparing the three Stroop tasks with each other, for the participants for whom we have data for all three tasks.

**Accuracy analyses.** Proportions of correct responses were entered into an ANOVA with task (classic, CM, and CM+) and distractor condition (neutral, congruent or incongruent semantic information) as within-subjects factors on the subset of our sample that completed all three Stroop tasks ( $N=46$ ). We observed significant main effects of task ( $F(1.45, 65.05)=17.08, MSE=.001, \eta^2_p=.28$ ) and distractor condition ( $F(1.14, 51.10)=38.32, MSE=.001, \eta^2_p=.46$ ), a significant interaction between these factors ( $F(1.44, 64.71)=24.48, MSE=.002, \eta^2_p=0.35$ ). We analyzed the effect of distractor condition separately for each task in order to understand this interaction. In the classic Stroop task, distractor condition significantly affected accuracy, with lower accuracy on incongruent trials ( $M=.940, SEM=.01$ ) than on congruent ( $M=.999, SEM=.001$ ) or neutral trials ( $M=.999, SEM=.001$ ), which did not significantly differ ( $p \approx 1$ ). The same pattern was observed for CM+ Stroop: significantly lower accuracy was observed during



incongruent trials ( $M=.988$ ,  $SEM=.003$ ) than during congruent ( $M=.996$ ,  $SEM=.002$ ) or neutral ( $M=.999$ ,  $SEM=.001$ ) trials, which did not significantly differ ( $p>.48$ ). However, for CM Stroop, there was no significant effect of distractor condition ( $p>.08$ ;  $M$ s from 0.991-0.997). A significant interaction ( $F(1.01, 49.55)=25.59$ ,  $MSE=.001$ ,  $\eta^2_p=.36$ ) in a 2-way ANOVA of task and distractor condition including only the classic and CM+ Stroop tasks indicated that the effect of distractor condition on accuracy was larger in classic than in CM+ Stroop.

These analyses suggest that CM+ Stroop is intermediate between CM and classic Stroop in its proclivity for provoking errors on incongruent trials. However, in all tasks, the error rate was extremely low (means  $<2\%$  in each distractor condition of each task), far lower than the mean error rates observed by Kane and Engle (2003), which ranged from 1-18% depending on experimental condition and working memory span group. We therefore focused on response times when analyzing relationships between Stroop interference and working memory span.

**Response time analysis.** Error trials were excluded from our response time analysis. We trimmed responses faster than 180 ms from our data set, and excluded responses slower than 3 standard deviations from the mean separately for each task. This trimming procedure resulted in the removal of  $<3\%$  of otherwise valid trials.

An ANOVA with task (classic, CM, or CM+ Stroop) and distractor condition (congruent, incongruent or neutral) yielded significant main effects of both task ( $F(1.64,73.86)=126.47$ ,  $MSE=4759.70$ ,  $\eta^2_p=.74$ ) and distractor condition ( $F(2,90)=379.14$ ,  $MSE=1541.59$ ,  $\eta^2_p=.89$ ) and a significant interaction ( $F(2.85,128.44)=77.63$ ,  $MSE=952.66$ ,  $\eta^2_p=.63$ ). These values are depicted in Figure 1. In the classic Stroop task,

incongruent trials were considerably slower than neutral trials, and congruent trials were slightly faster than neutral trials. In the CM Stroop task, we observed more modest slowing on incongruent trials but no speeding on congruent trials. The pattern of the CM+ Stroop tasks was intermediate between these, with more slowing on incongruent trials than in CM Stroop, but less than in classic Stroop, and unlike classic Stroop, no evidence of facilitation effects. The lack of significant facilitation in both CM and CM+ Stroop is consistent with the assumption that the context of the cross-modal tasks supports goal recovery, allowing no way to respond quickly but accurately using an incorrect goal (such as reading the word; see MacLeod & MacDonald, 2000).

To understand the interaction between task and distractor condition, we ran separate ANOVAs for each task, which confirmed that incongruent trials were significantly slower than congruent and neutral trials in each task. Only in classic Stroop were congruent trials significantly faster than neutral trials. The interaction in the 2-way ANOVA could be due to differences in the amount of slowing on incongruent trials, or to the absence of differences between congruent and neutral trials in the CM and CM+ Stroop tasks. We tested hypotheses about the source of the task by distractor condition interaction by running separate ANOVAs comparing the CM+ Stroop task in turn with the CM Stroop task and with the classic Stroop task, including only incongruent and neutral trials. In the ANOVA including CM Stroop and CM+ Stroop, we observed a significant interaction between task type and distractor condition ( $F(1,45)=28.34$ ,  $MSE=386$ ,  $\eta^2_p=.39$ ), supporting the hypothesis that more slowing should occur in CM+ than in CM Stroop. In the ANOVA including classic Stroop and CM+ Stroop, the interaction between task type and distractor condition was also statistically significant

( $F(1,45)=44.85$ ,  $MSE=1093$ ,  $\eta^2_p=.50$ ), consistent with the assumption that more slowing occurred in classic than in CM+ Stroop. We also analyzed difference scores as dependent variables, where Stroop interference effects equaled the difference between incongruent and neutral RTs. By task, each Stroop interference score significantly differed, with a mean interference effect of 63.60 ( $SEM=4.98$ ) ms in CM Stroop, 94.44 ( $SEM=6.92$  ms) in CM+ Stroop, and 159.74 ms ( $SEM=7.36$ ) in classic Stroop.

Figure 2 displays Stroop interference and facilitation effects as differences between the distributions in question at equally-distributed points. These delta plots show the differences between incongruent and neutral trials (left panel), or Stroop interference, for each task, and also differences between congruent and neutral trials, or Stroop facilitation (right panel). These plots show that while the sizes of the effects of Stroop interference clearly differ between tasks, the relationships between these trial types are similar throughout their distributions for each task. Considering interference effects, the size of the difference between incongruent and neutral response times increases as responses slow, producing positively-sloped lines. This is a pattern typically observed in Stroop tasks, which differ from other superficially similar tasks such as the Simon task (Pratte, Rouder, Morey, & Feng, 2010). Stroop facilitation shows another pattern: constant speeding throughout the distributions. Though in the CM and CM+ Stroop tasks mean facilitation effects are not statistically significant, the relationship between congruent and neutral response time distributions remains constant as response times increase, as facilitation effects do in classic Stroop.

**Correlational analyses: Working memory capacity, interference resolution, and attentional lapses.** Although it is clear that CM and CM+ Stroop produce smaller

interference effects than classic Stroop, our response time analysis, consistently with previous research on cross-modal Stroop (e.g., Elliott et al., 1998; Elliott et al., in prep.), does not suggest that the cross-modal and classic tasks measure fundamentally different conflict resolution processes. This probable inter-task similarity, at least where the nature if not the size of the interference evoked is concerned, is important to establish before interpreting the relationships between Stroop interference and working memory span described in the next section. Our response time evidence is consistent with the assumption that in each of these tasks, the semantic meanings of distracting words provoke interference with color naming. We therefore tested whether variance in Stroop interference effects was shared with variation in working memory capacity.

We calculated a composite working memory measure from the Operation Span ( $M=41.03$ ,  $SD=16.90$ , range: 6-75) and Symmetry Span ( $M=20.24$ ,  $SD=8.88$ , range: 4-37) data in the same manner as in Experiment 1.<sup>3</sup> Pearson's correlations are given in Table 2. Split-half reliabilities calculated for the Stroop tasks (these are given along the diagonal) indicated good internal consistency. Internal consistency is also evinced by the significant correlation between CM and CM+ Stroop. We tested statistical significance of correlations using one-tailed tests, because we believed a priori (based on Kane and Engle, 2003) that any relationship between working memory and Stroop interference should be negative, such that as working memory capacity increases, Stroop interference decreases.

We found a significant, negative correlation between interference in the classic Stroop task and our working memory composite scores,  $r(102)=-.252$ . This value

---

<sup>3</sup> We observed a statistically significant correlation between Operation and Symmetry span in our sample,  $r=.366$ .

significantly differed from the  $r$ -value we observed in the same test in Experiment 1, according to a  $z$ -test performed on Fisher  $r$ -to- $z$  transformed values ( $z=-3.07$ ). Neither the interference scores from CM Stroop nor the interference scores from CM+ Stroop correlated significantly with working memory capacity. For the CM Stroop task, a null result was expected (see Elliott et al., 2006). For the CM+ Stroop task, this null correlation was possibly due to insufficient statistical power; we tested only 85 participants in the CM+ task compared with the 104 in the classic Stroop task. However, assuming that the observed value of the correlation in our data between working memory and CM+ Stroop interference is close to the true value, we would require at least 832 participants to achieve sufficient power (i.e., 0.80; calculations carried out with G\*Power; Faul, Erdfelder, Lang, & Buchner, 2007) to detect a statistically significant correlation of  $r=-.086$ , suggesting that any relationship between working memory and the interference measured by CM+ Stroop is likely weaker than the relationship we observed between working memory and classic Stroop. We also tested whether the correlation between classic Stroop and working memory scores remained statistically significant if variance shared with CM+ Stroop interference is removed. Partialling out shared variance with CM+ Stroop interference, the correlation between classic Stroop interference and working memory remained significant ( $r(82)=-.199$ ). However, partialling out shared variance with classic Stroop did not make the correlation between CM+ Stroop interference and working memory statistically significant ( $r(82)=-.047$ ). The significant correlation between classic Stroop facilitation effects and working memory span given in Table 2 did not remain statistically significant after accounting for shared variance with classic Stroop interference effects, therefore we do not attempt to interpret it.

While CM+ Stroop interference did not significantly correlate with working memory capacity, a measure of goal neglect taken from the CM+ Stroop task did. We found a significant relationship between the number of valid responses ( $M=11.13$ ,  $SD=1.33$ , range: 5-12) collected in the auditory monitoring portion of the CM+ Stroop task and working memory capacity ( $r=.31$ ). Participants with higher working memory spans were more likely to notice and respond to the target sounds within the 2000 ms period from the onset of the sound during which these responses were collected. Because the monitoring task itself was simple, we consider omissions in this task to be attributable to goal neglect. Target trials were rare, occurring only 12 times in a 96-trial block, which should have created an unsupportive context for actively maintaining this goal.

## **Discussion**

Our results yielded several important findings. First, we confirmed that the relationship between working memory span and Stroop interference predicted by Kane and Engle (2003) is present in a large sample of individuals with varying working memory capacities; thus their results were not dependent on an extreme groups design. That this assumption should hold is essential for making inferences about the meaning of the relationship between working memory capacity and Stroop interference. The correlation between working memory capacity and print-based Stroop interference depends on the task's context, manipulated here as the proportion of congruent trials in a session, confirming previous expectations (Hutchison, 2011).

In contrast, we observed no correlation between CM Stroop interference and working memory capacity, despite observing robust Stroop interference in both cross-modal Stroop tasks. We think that these patterns of correlation constitute reasonable

evidence of a selective relationship between classic Stroop with a high proportion of congruent trials and working memory capacity that does not exist, or at least not at the same strength, between cross-modal Stroop tasks and working memory capacity, or between classic Stroop with no congruent trials and working memory capacity. This pattern of evidence can be explained by supposing that relationships between working memory and selective attention are driven primarily by individual differences in tendencies to actively maintain the task goals, rather than individual differences in resolving conflicts between task-relevant and task-irrelevant information. This conclusion is bolstered by the significant relationship we observed between auditory monitoring responses and working memory capacity; because the target sounds occurred so rarely during the CM+ Stroop task, actively maintaining this goal was necessary for success but poorly supported by the task's context.

### General Discussion

In different ways, the results of both Experiment 1 and Experiment 2 confirm that relationships between Stroop interference and working memory capacity depend on task context. First, we confirm that classic Stroop interference only correlates with working memory capacity under circumstances where task goal maintenance also varies, such as when the frequent occurrence of congruent trials decreases the likelihood of actively maintaining the task goal (see also Hutchison, 2011). In both contexts, goal maintenance is crucial for performance on incongruent trials, but when most trials are congruent, participants prone to attentional lapses may be more liable to neglect the task goal than in a context in which most trials are incongruent. When goal neglect is likely, the Stroop interference effect reflects both resolution of conflict and recovery of the task goal. In

cross-modal versions of the Stroop task, the ease of recovering the task goal likely precludes the need to actively maintain it, regardless of the proportion of congruent trials. Even though increasing the proportion of congruent trials increased the size of the interference effect in CM Stroop, especially when the sounds must be attended (as in CM+ Stroop), significant correlations with working memory capacity did not emerge. We believe that this is because these cross-modal tasks offer a supportive context for goal recovery, whereas in print-based Stroop, recovery of the color-naming goal is challenging because of the competing possibility of reading the word.

The absence of negative correlations between CM Stroop and working memory capacity and between classic Stroop with no congruent trials and working memory capacity is unexpected assuming the hypothesis that working capacity reflects the attentional control needed to resolve interference between task-relevant and task-irrelevant stimuli. However, these findings present no problem for the related assumption that working memory capacity indexes a broader level of attention, affecting the propensity of an individual to actively maintain the current task goal, or perhaps to experience attentional lapses that make recovering the task goal necessary (see also McVay & Kane, 2009; Mecklinger, Weber, Gunter, & Engle, 2003; Unsworth, Redick, Lakey, & Young, 2010). Because cross-modal Stroop tasks likely differ in the contextual support available for recovering the task goal in the event of an attentional lapse, one might not predict as strong a relationship between these tasks and working memory capacity as one would predict between classic Stroop and working memory capacity. These findings suggest a need for further research capable of distinguishing between



broad and fine attentional control operations, in order to better distinguish between these constructs.

This interpretation of the differences between these two Stroop tasks is strengthened by the results we observed with a version of the CM Stroop task in which participants were also instructed to monitor the sounds for a particular target (i.e., CM+ Stroop). In the CM+ Stroop task, we observed increased Stroop interference compared to CM Stroop, but again no significant correlation with working memory capacity. Compared with classic Stroop, we think CM+ Stroop, like CM Stroop, offers greater contextual support for recovering a task goal after an attentional lapse, because there is no obvious alternative response to make to the appearance of a color square (e.g., reading it). The low error rates typically observed in cross-modal Stroop tasks (as well as in our experiments) diminish the possibility that participants are likely to inadvertently repeat a spoken distractor as their response. Although CM+ Stroop interference did not significantly correlate with working memory capacity, the number of valid responses to its auditory monitoring component did. Participants with lower working memory capacity made fewer valid responses to the auditory targets, which is consistent with the assumption that they sometimes forgot this secondary task goal, and consistent with the assumption that working memory capacity is crucial for maintaining attention at this broader level of a task. This result corroborates previous findings in which working memory capacities were found to be related to event-based prospective memory (Brewer, Knight, Marsh, & Unsworth, 2010), and also seems consistent with claims that *g*, which is strongly related to working memory measures (Conway, Kane, & Engle, 2003), reflects

broad control of attention as measured by tasks designed to induce goal neglect (Duncan, Emslie, Williams, Johnson, & Freer, 1996).

Given the strength of Kane and Engle's (2003) results, we expected to observe stronger relationships between Stroop interference with frequent congruent trials and working memory capacity, and also to observe somewhat more variability in error rates, despite our instructions to attend to accuracy. We think that in order to observe such high error rates, it would be necessary for experimenters to emphasize speed more than we did; our experimenters emphasized accuracy because we wanted to reduce errors in classic Stroop, so that error rates would not drastically differ between classic and CM Stroop tasks. Although we did manage to observe a statistically significant sample-wide correlation in the highly congruent task context, it was of modest strength. This likely reflects that multiple factors, for instance interference resolution and goal neglect, underlie Stroop interference effects. Even if we assume that goal neglect is primarily the component that shares variance with working memory capacity, the amount of shared variance that can be expected between these tasks is limited by the multi-faceted nature of the Stroop interference effect. Also, the cognitive control necessary for preventing goal neglect may also vary considerably with the passage of time, possibly diminishing effect sizes and making individual differences even more difficult to detect. Meier and Kane (2010) found that differences in the size of Stroop interference effects between groups of participants with low, medium, and high working memory capacity grew smaller over the course of an experimental session. Taken together, these factors undermine the utility of using Stroop alone as a measure of attentional lapsing, although stronger correlations could possibly be found in Stroop tasks if administrators take additional measures to

increase the chances of goal neglect. Collecting multiple measures of goal neglect, as we did by also collecting auditory monitoring responses, is another way to ensure that some measure of attentional lapsing is available. However, auditory monitoring or prospective memory tasks, while conceptually quite direct measures of goal maintenance, are also limited in that only a few responses can be collected during an experimental session, meaning that large samples will likely be required to detect these effects.

Other research, particularly research regarding inhibition or attentional filtering in visual short-term memory (McNab & Klingberg, 2008; Vogel et al., 2005), seems to strongly support the hypothesis that irrelevant items are more likely to be encoded by individuals with lower working memory capacity because these individuals allow these items to be encoded into a limited-capacity system even though they are task-irrelevant. Although these findings would seem at first glance to support the notion that the attentional control related to working memory capacity functions to keep irrelevant information out of mind, we think that these findings may require a more complex explanation. For instance, there is good reason to suppose that attentional filtering depends strongly on the perceptual load posed by the to-be-encoded and to-be-ignored items (Lavie, Hirst, de Focker, & Viding, 2004); thus experimental outcomes (and perhaps also theoretical predictions) should vary with changes in load. Also, maintenance of task goal or context (cf. Braver, Gray, & Burgess, 2007) would seem to be a prerequisite for successful attentional filtering; one cannot intentionally exclude some information from mind without remembering from moment to moment which information is relevant and which is irrelevant. Perhaps for many published examples of relationships between inhibition or attentional filtering and working memory capacity, it

is variation in the ability to maintain and access these global goals, rather than variation in operating on some goal, that drives the observed relationships. Our results suggest that relationships between working memory capacity and attentional filtering are only likely to emerge in task contexts that where goal recovery is non-trivial and retrieving the incorrect goal has consequences.

Consider the evidence of Conway et al. (2001), in which individuals with low working memory capacity were more likely to hear their name in the unattended channel of a dichotic listening task. Suppose that attention sometimes lapses during a dichotic listening task; in this case, the participant would not selectively attend to either ear until recalling which ear was to be attended, and might then hear their name in the to-be-ignored stream. Beaman (2004)'s observation that working memory capacity predicted the size of ISEs most strongly when the irrelevant sounds were semantically related to the memoranda might also be explained in terms of contextual effects on goal-recovery. Whether to-be-ignored words were related or unrelated to the to-be-remembered words, attention might lapse during word presentation, but perhaps before the goal (i.e., recalling the *visually-presented* words) was recovered, participants encoded some aurally presented to-be-ignored words, which were then more likely to be recalled when they related to the intended memoranda than when they did not. In many tasks used to investigate the effects of irrelevant stimuli on memory, both fine and broad attentional control might be supposed necessary, but often these are not distinguishable from each other. It is impossible to declare for certain which of these levels of attentional control, the fine or the broad, is more responsible for the variation in individual differences in attentional selection previously reported in the literature. Our evidence does not eliminate

the possibility that attentional control is needed in many circumstances to suppress irrelevant stimuli, but it does confirm that broader attentional control needed to prevent disengagement is a key factor upon which individuals differ.

In conclusion, our results provide clear evidence that working memory capacity and Stroop interference are related in a sample-wide analysis for circumstances in which active goal maintenance is critical for task performance. When active goal maintenance is less important (as we believe it to be in our cross-modal Stroop tasks) or unlikely to vary much between individuals (as in classic Stroop without congruent trials), relationships between working memory and Stroop interference are not detectable. This suggests a possible qualification to proposed relationships between attention and working memory capacity, namely that working memory capacity may be primarily related to actively maintaining the task's goal, which is necessary for inhibiting the encoding of irrelevant information.

## References

- Beaman, C. P. (2004). The irrelevant sound phenomenon revisited: What role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1106-1118.
- Braver, T. S., Gray, J.R., & Burgess, G.C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. *Variation in working memory* (A.R.A. Conway, C. Jarrold, M.J. Kane, A. Miyake, & J. Towse (Eds.), pp. 76–106). Oxford, U.K.: Oxford University Press.
- Brewer, G. A., Knight, J. B., Marsh, R. L., & Unsworth, N. (2010). Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory & Cognition*, *38*, 304-311.
- Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*, 331-335.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.
- Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*, 547–552.
- Cowan, N., & Barron, A. (1987). Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Perception & Psychophysics*, *41*, 393-401.

- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*, 42-100.
- Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, *30*, 257-303.
- Elliott, E. M., & Cowan, N. (2001). Habituation to auditory distractors in a cross-modal, color-word interference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 654-667.
- Elliott, E. M., & Cowan, N. (2005). Coherence of the irrelevant-sound effect: Individual profiles of short-term memory and susceptibility to task-irrelevant materials. *Memory & Cognition*, *33*, 664-675.
- Elliott, E. M., Barrilleaux, K. M., & Cowan, N. (2006). Individual differences in the ability to avoid distracting sounds. *European Journal of Cognitive Psychology*, *18*, 90-108.
- Elliott, E. M., Cowan, N., & Valle-Inclan, F. (1998). The nature of cross-modal color-word interference effects. *Perception & Psychophysics*, *60*, 761-767.
- Elliott, E.M., Morey, C.C., Morey, R.D., Eaves, S.D., & Shelton, J.A. (invited resubmission in preparation). Assessing interference effects across modalities: Stroop tasks reveal more similarities than differences.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

- Hutchison, K. A. (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 851-860.
- Ishihara, S. (1966). *Tests for colourblindness*. Tokyo, Japan: Kanehara Shuppan.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47-70.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*, 169-183.
- Lavie, N., Hirst, A., de Fockert, J., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology-General*, *133*, 339-354.
- Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition*, *30*, 294-301.
- MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition*, *26*, 201-211.
- MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*, 383-391.
- McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*, 103-107.



- McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 196-204.
- Mecklinger, A., Weber, K., Gunter, T. C., & Engle, R. W. (2003). Dissociable brain mechanisms for inhibitory control: Effects of interference content and working memory capacity. *Cognitive Brain Research*, *18*, 26-38.
- Meier, M.E., & Kane, M.J. (2010). Response set, repetitions, and runs moderate working memory capacity's effect on Stroop interference. Poster presented at the 51<sup>st</sup> annual meeting of the Psychonomic Society, St. Louis, MO.
- Oetzmann, A., & Mazzoni, D. (2011, June 21). *Audacity, v1.2*. Retrieved from <http://audacity.sourceforge.net/manual-1.2/>.
- Pratte, M. S., Rouder, J. N., Morey, R. d., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophysics*, *72*, 2013-2025.
- Roelofs, A. (2010). Attention and facilitation: Converging information versus inadvertent reading in Stroop task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 411-422.
- Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.
- Schröger, E., & Wolff, C. (1998). Behavioral and electrophysiological effects of task-irrelevant sound change: A new distraction paradigm. *Cognitive Brain Research*, *7*, 71-87.

- Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 506–526.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505.
- Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, *38*, 111-122.
- Vogel, E. K., McCullough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500-503.

Table 1. Pearson's correlations between Stroop response time effects and working memory span, Experiment 1.

	1.	2.	3.
1. WM Score			
2. Classic Effect	.18** <sup>a</sup>	.93	
3. CM Effect	-.03	-.08	.96

Note. Values indicated with an \* are statistically significant according to a one-tailed test at  $p < .05$ . Stroop effects refer to interference calculated as the incongruent – neutral.

Values on the diagonal for the Stroop tasks are split-half reliabilities, with alternating trials assigned to each half. Reliability for working memory tasks was assessed separately for each task using Chronbach's alpha. Adequate values were observed in both tasks (Operation span  $\alpha = 0.88$ , Symmetry span  $\alpha = 0.77$ ).

<sup>a</sup> Two-tailed  $p < .07$ ; the sign of the value was opposite to expectations.

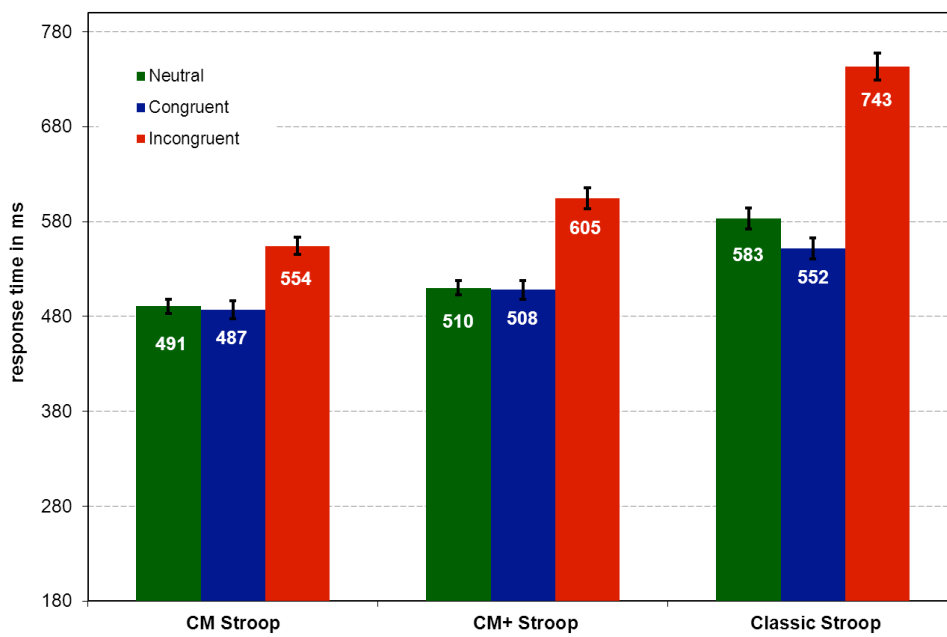
Table 2: Pearson's correlations between Stroop response time effects and working memory span, Experiment 2

	1.	2.	3.	4.	5.
1. WM Score					
2. Classic Effect ( $N=104$ )	-.25*	.94			
3. Classic Facilitation	-.22*	.30*	.94		
4. CM Effect ( $N=65$ )	.08	.08	.03	.95	
5. CM+ Effect ( $N=85$ )	-.09	.20*	-.01	.57* <sup>b</sup>	.94

Note. Values indicated with an \* are statistically significant according to a one-tailed test at  $p < .05$ . Stroop effects refer to interference calculated as the incongruent – neutral, whereas facilitation refers to congruent – neutral. Though some participants completed all 3 Stroop tasks as well as the WM measures ( $N=46$ ) most completed classic Stroop, the WM measures, plus either CM or CM+. Values on the diagonal are split-half reliabilities in the Stroop tasks, with alternating trials assigned to each half. Reliability for working memory tasks was assessed separately for each task using Chronbach's alpha. Adequate values were observed in both tasks (Operation span  $\alpha=0.82$ , Symmetry span  $\alpha=0.66$ ).

<sup>b</sup> For this value only,  $N=46$ .

Figure 1



*Figure 1.* Trimmed mean response times (with standard errors of the mean) for each task by distractor condition, for the 46 participants who took part in all 3 tasks in Experiment 2.

Figure 2

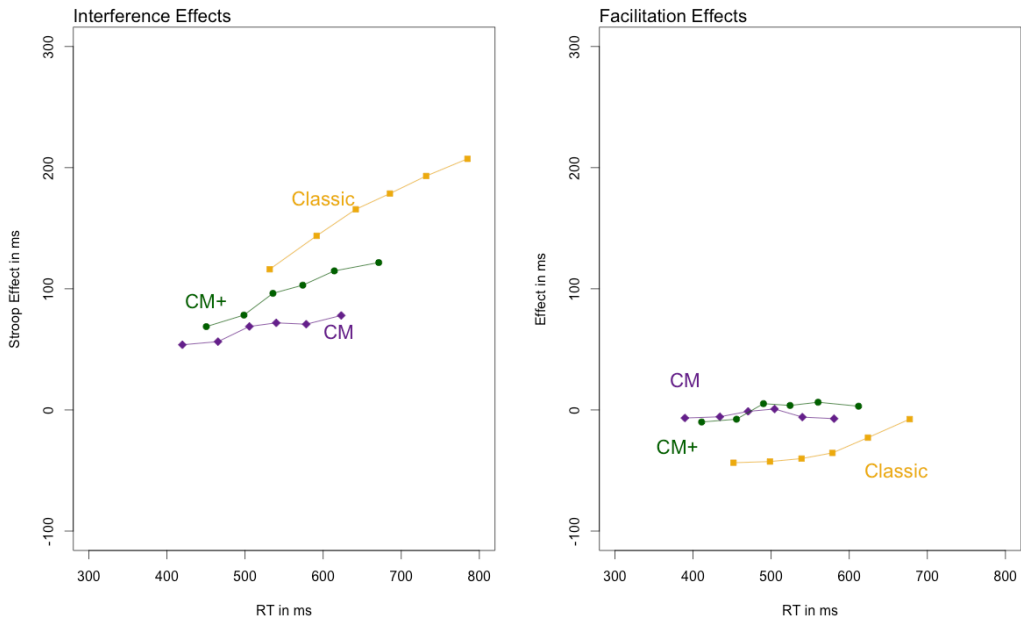


Figure 2. Delta plots, by task type, comparing incongruent and neutral distributions (i.e., Stroop interference, left) and congruent and neutral distributions (i.e., Stroop facilitation, right), Experiment 2.