



# Providing goal reminders eliminates the relationship between working memory capacity and Stroop errors

Audrey V. B. Hood<sup>1</sup> · Keith A. Hutchison<sup>1</sup>

Accepted: 24 September 2020  
© The Psychonomic Society, Inc. 2020

## Abstract

Previous research has shown that list-wide effects in the Stroop task interact with working memory capacity (WMC). The predominant explanation for this relationship is goal maintenance. However, some researchers have challenged whether list-wide effects truly reflect goal-maintenance abilities. In the current study, we examined whether goal maintenance explains higher WMC individuals' better performance within mostly congruent (MC) Stroop lists by providing periodic goal reminders to some of the participants. Two hundred and twelve participants from Montana State University first completed the Automated Operation Span and were then assigned to either a true control, goal reminder, or nongoal reminder condition. During the Stroop task, the true control condition received rest breaks every 60 trials, whereas the goal reminder and nongoal reminder conditions stopped every 12 trials to vocalize either the task goal or a rehearsed statement, respectively. We regressed Stroop errors on reminder condition and WMC, comparing each group to the true control. For the Goal Reminder  $\times$  True Control comparison, there was an interaction, such that WMC negatively correlated with Stroop errors in the true control, but not in the goal reminder condition. In contrast, for the Nongoal Reminder  $\times$  True Control comparison, there was only an overall effect of WMC, with greater Stroop errors for those lower in WMC. These data provide evidence that goal reminders eliminate the relationship between WMC and Stroop interference.

**Keywords** Cognitive control · Stroop interference · Goal neglect · Working memory capacity

Cognitive control refers to the ability to orchestrate thought and action in accord with internal goals—particularly in situations that have the potential for distraction. A major component of cognitive control is goal maintenance, or continued access to information related to the current task regardless of interference stemming from habit (Engle, 2001). Individuals higher in working memory capacity (WMC) have greater cognitive control than individuals lower in WMC, as they are better at maintaining task goals in the face of distraction and overriding habitual responses to produce task-relevant responses (Conway, Cowan, & Bunting, 2001; Engle, 2002; Kane, Bleckley, Conway, & Engle, 2001). Lower WMC individuals, in contrast, are often subject to goal neglect—a momentary loss of the current task goal (Duncan, Emslie, Williams, Johnson, & Freer, 1996).

These WMC-related differences in ability to control attention particularly manifest when there is conflict. In these situations, one must successfully keep task goals active in order to respond appropriately when conflict arises (Engle & Kane, 2004). Higher WMC individuals' superior ability under these conditions has been evidenced in a multitude of tasks, including dichotic listening (Conway et al., 2001), the antisaccade task (Engle, 2001; Kane et al., 2001), the Simon task (A. Miller, Watson, & Strayer, 2012), the Sustained Attention to Response task (SART; McVay & Kane, 2009), and the AX-CPT (Redick, 2014). In each of these, the explanation for WMC-related differences in performance has been attributed to higher WMC individuals' ability to maintain task goals in an active state (leading to task-appropriate responses) and lower WMC individuals' inability to maintain such task goals (leading to automatic, habitual responses).

These WMC-related differences in ability to control attention are further demonstrated in studies using the Stroop task (Stroop, 1935). During this task, participants are presented with the name of a color presented in either congruent or incongruent font (e.g., GREEN in green font, or GREEN in blue font, respectively). The task is to name

---

✉ Keith A. Hutchison  
khutch@montana.edu

<sup>1</sup> Department of Psychology, Montana State University,  
P.O. Box 173440, Bozeman, MT 59717-3440, USA

the font color while ignoring the word itself. This presents a challenge, as individuals must overcome interference on incongruent trials by selectively orienting their attention toward the weaker color dimension, instead of relying on the more habitual response of word reading (E. K. Miller & Cohen, 2001). The Stroop effect represents the typical finding that participants respond slower and less accurately when naming the font color of incongruent words than when naming the font color of congruent words (for a review, see MacLeod, 1991).

Given the role of WMC in goal maintenance—particularly in the face of distraction—it is not surprising that WMC is a strong predictor of Stroop interference (Hutchison, 2011; Long & Prat, 2002; Unsworth & Spillers, 2010). Moreover, the proportion of congruent trials in a list moderates such WMC-related differences (Kane & Engle, 2003). As explained by Kane and Engle (2003), the response conflict triggered during each incongruent trial can serve as a reminder of the task goal to name the color, rather than the word. Thus, when the number of such incongruent trials is decreased, this requires more internal goal maintenance. Kane and Engle (2003) provided evidence consistent with this goal maintenance account by presenting participants with mostly congruent (MC) or mostly incongruent (MI) Stroop lists. Overall, they found that lower, compared with higher, WMC individuals committed more errors on incongruent trials when the list-wide proportion congruency was either 75% (Experiments 1–3) or 80% (Experiment 4). Because correct responses in such lists can usually be produced by either reading the word or naming the color, the infrequent reinforcement of the task goal presumably resulted in lower WMC individuals resorting to the habitual response of word reading.

In addition to the pattern they obtained in errors, Kane and Engle (2003) also found that individuals lower in WMC have greater response-time interference compared with individuals higher in WMC, although this effect was not as robust as that found in errors. Specifically, in Experiment 3, the WMC difference in Stroop reaction times was only significant when the 75% congruent task occurred subsequent to a 0% congruent task. However, when analyzing the overall data across experiments, they did find a significant effect in which individuals lower in WMC showed larger response-time interference compared with individuals higher in WMC. Because this effect was not particularly robust, they noted that larger samples may be required in order to detect this relationship.

As outlined by Kane and Engle (2003), these results can be explained via a two-process model of goal maintenance and conflict resolution. Stroop effects primarily manifest in errors when participants neglect the task goal, because they accidentally produce the incorrect word response (presumably) before the response conflict is detected. In contrast, Stroop effects primarily manifest in reaction times (RTs) when participants

attempt to resolve the competition between conflicting word and color responses on incongruent trials, suggesting that the goal was active prior to stimulus onset. As evidenced in their results, individuals higher in WMC outperform lower WMC individuals in both errors and RTs.

Similar findings have been found in other studies as well. For instance, Spieler, Balota, and Faust (1996) compared Stroop performance between healthy younger and older adults and individuals diagnosed with Alzheimer's disease. Although healthy older adults took longer to suppress the irrelevant dimension of the stimulus (i.e., the word) compared with younger adults, they were still able to do so, resulting in the production of a correct response. In contrast, individuals diagnosed with Alzheimer's disease were much more likely to produce errors on incongruent trials, suggesting that they were more likely to lose the goal of the task, leading to behavior driven by the irrelevant dimensions of the stimulus. Further, these error rates were shown to increase with the severity of the disease (see Balota et al., 2010; Hutchison, Balota, & Duchek, 2010, for further evidence of the utility of Stroop incongruent error rates for diagnosing Alzheimer's disease). In addition, Hutchison, Smith, and Ferris (2013) found that stereotype threat produced a significant increase in Stroop errors among those lower in WMC, but only under MC list conditions that require internal goal maintenance. Hutchison et al. (2013) explained that the distraction caused by stereotype threat-related thoughts resulted in a loss of the appropriate goal to suppress habitual (yet incorrect) response tendencies. To summarize, the findings from each of these studies demonstrate that if the goal is neglected, error rates on incongruent items increase. However, although likely not as robust as error effects, performance differences due to goal neglect may also manifest in larger Stroop RT effects, due to some fast responses on congruent trials (if participants quickly and mistakenly name the word) or slow responses on incongruent trials (if participants retrieve the goal in time to correct their output).

Although the goal-maintenance account of WMC differences in Stroop effects has since received support (Entel & Tzelgov, 2019; Hutchison, 2011; Hutchison et al., 2013; Morey et al., 2012), some researchers have challenged the assumption that list-based effects in Stroop performance actually reflect differences in goal maintenance. One alternative argument is based on the item-specific proportion congruency (ISPC) effect. The ISPC effect refers to the finding of reduced Stroop effects for specific words within a list that usually appear in an incongruent font color, relative to words that usually appear in a congruent font color. Jacoby, Lindsay, and Hessels (2003) first demonstrated this effect by manipulating congruency not across lists, but rather across items within an overall 50% proportion congruent (PC) list. Across all three of their experiments, MI items had a smaller Stroop effect than MC items.

Because all list-wide proportion congruency manipulations prior to 2003 had confounded list-wide with item-specific congruency (i.e., items in MC lists were MC themselves), the results from Jacoby et al. (2003) suggested that earlier list-wide congruency effects may have been due to automatic processes occurring at the item level, rather than a central top-down control mechanism (i.e., goal maintenance). If list-wide congruency effects are indeed due solely to item-level effects, then this would be problematic for explanations of WMC differences in MC lists. Instead, it is possible that even when goals are maintained, individuals lower in WMC may be more susceptible to the strong word reading response triggered by such MC items, as these individuals are typically less able to suppress prepotent responses (Kane et al., 2001; Kane & Engle, 2003).

However, more recent studies have demonstrated the list-wide proportion congruence effect does, in fact, reflect list-level control (Bugg & Chanani, 2011; Bugg, McDaniel, Scullin, & Braver, 2011; Hutchison, 2011). For instance, interference is reduced for items that appear in MI compared with MC lists, even when the congruency of the items themselves is equated across lists (Bugg & Chanani, 2011; Hutchison, 2011). For instance, Bugg et al. (2011) included neutral items (e.g., concrete English nouns) displayed in different font colors and found faster response times for neutral trials in MI lists than in MC lists. Because the neutral words did not have an item-specific proportion congruency bias, these results were attributed to list-level control. More recently, Cohen-Shikora, Diede, and Bugg (2018) examined whether individuals are sensitive to dynamic changes in experience (i.e., proportion congruency) as a list transitions from MC or MI (first 6 items) to 50% congruent (last 12 items). They found a rapid list-wide effect in the first six items (called “inducer” items) that persisted for the first half (six items) of the 50% list segment, but then disappeared. This gives evidence for a brief carryover of control settings, which then quickly adjusts to match the changing environmental conditions. In their second experiment, they found similar effects for separate transfer items that were always 50% congruent, providing evidence that the effects were driven by a global control mechanism, rather than triggered by an item-specific learning mechanism.<sup>1</sup>

In addition to these studies, Hutchison (2011) examined the interactivity of list-wide and ISPC effects with WMC. In this study, MC and MI words were embedded within lists consisting of filler items that were either 100% congruent or 100% incongruent, resulting in MC or MI lists. This allowed him to factorially test the contributions of list-wide and item-specific effects, and their relation to WMC. Hutchison replicated Kane and Engle’s (2003) finding that WMC-related

differences only occurred for MC lists, even when ISPC was held constant, providing additional support for the goal maintenance account of list-wide congruency effects. Interestingly, however, ISPC effects also significantly interacted with list-wide PC and with WMC, such that ISPC effects were greater when the list-wide proportion was MC and among those lower in WMC. Hutchison interpreted this pattern to suggest that MC items attract attention toward word reading, and the impact of this bottom-up salience effect is greater when top-down control is relaxed, as it would be for lower WMC individuals or those receiving a MC list. In contrast, MI items implicitly trigger control, which can compensate for relaxed or impaired goal maintenance. Overall, these results suggest both goal maintenance and implicit learning play a role in WMC-related differences in Stroop performance. As a result, testing a goal maintenance account of WMC differences in Stroop performance requires a more direct manipulation of external goal support than relying on proportion congruency manipulations across lists or items.

As an alternative approach to item or list PC, some studies have used precues signaling the probability of congruency for an upcoming Stroop trial or list. Such precuing allows one to manipulate expectations of upcoming conflict (Bugg, Diede, Cohen-Shikora, & Selmeczy 2015; Correa, Rao, & Nobre, 2009; Gratton, Coles, & Dorchin, 1992; Logan & Zbrodoff, 1982; see Bugg & Smallwood, 2014, for a review). For instance, Bugg et al. (2015, Experiment 5) used precues to vary expectations for each upcoming list of 20 trials, while holding experience (i.e., proportion congruency, which was 50%) constant. Participants showed a larger Stroop effect when cued that an upcoming list was MC than when cued it would be MI. However, similar to the Cohen-Shikora et al. (2018) study presented above, this cueing effect was only significant for the first half of the list. In addition, Hutchison, Bugg, Lim, and Olsen (2016) used trial-by-trial congruency precues and found that ISPC effects were absent when participants were told the next trial would likely be incongruent. This finding bolstered Hutchison’s (2011) findings in demonstrating that increasing top-down control reduces the tendency for distractor words to attract or deflect attention.

Although each of these studies were able to manipulate expectations of upcoming conflict using precues, a criticism is that such cues are unnatural for the Stroop task. Specifically, typical instructions for the Stroop task instruct participants to respond to the color and ignore the word. However, using cues such as “matching” or “easy” directs participants’ attention to word reading. In such cases, the experimenter has drawn participants’ attention to what is normally considered the “irrelevant” dimension in Stroop studies. This changes the nature of the task from one requiring sustained selective attention away from an irrelevant dimension to one in which participants can switch their attention between stimulus dimensions on a trial-by-trial basis.

<sup>1</sup> There is also an alternative temporal learning model that attempts to explain effects typically attributed to listwide PC (Schmidt, 2013a, 2014). We will delay discussion of this model until the [discussion](#) section.

## Current study

The purpose of the current study is to examine whether goal maintenance indeed explains higher WMC individuals' superior performance within MC Stroop lists. According to the goal maintenance account, when the context promotes goal neglect by failing to reinforce the task goal (i.e., during MC lists), lower WMC individuals commit more errors than higher WMC individuals (Kane & Engle, 2003). The current study will further test the role of goal maintenance in MC Stroop lists by keeping the task context (i.e., proportion congruency) constant, while providing periodic goal reminders to some of the participants. If goal maintenance indeed mediates the relation of WMC to error performance in MC lists, providing reminders should eliminate WMC-related differences in Stroop errors. Following previous studies, we hypothesized that there would be WMC-related differences in performance in a control group matched to previous studies (Hutchison, 2011; Kane & Engle, 2003). However, we hypothesized that there would not be WMC-related differences in the goal reminder condition.

In addition, we also included a yoked control group (nongoal reminder) to examine whether simply providing a break every 12 trials is enough to eliminate WMC-related differences in performance. Meier and Kane (2013) found that WMC-related differences in Stroop performance appear when there are many consecutive, uninterrupted congruent trials. Under such conditions, those lower in WMC are more likely to fall into the habitual response of word reading. Because of this, it is possible that giving a break every 12 trials (and thus preventing many consecutive, uninterrupted, congruent trials) could also eliminate WMC differences.

## Method

### Participants and design

In accord with Simmons, Nelson, and Simonsohn (2011), we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. Two-hundred and eleven undergraduates from Montana State University participated in the study for partial course credit. Although we did not ask participants to report their age or gender, the population of Introductory Psychology students at MSU features primarily freshman between 18–20 years old, of whom approximately 55–60% are female. Based on data from Hutchison (2011) examining the relationship between AOSPAN performance and Stroop effects in MC lists, we predicted a medium-sized effect (.30). Using G-power, the *a priori* power analysis indicated a sample size of 67 to have 80% power for detecting a medium-sized effect when employing the traditional .05 criterion of statistical

significance. To stay consistent with Hutchison (2011), we decided to run 70 participants per group.

We removed data from one participant for only completing 114 trials (52%) of the Stroop task. This resulted in usable data from 210 participants. Each participant was tested individually in a laboratory session lasting approximately thirty minutes.

### Measures and apparatus

We used E-studio E-prime software from Psychology Software Tools (Version 2.0.8.90) to program and present the Stroop stimuli. Stimuli were presented using a Dell Optiplex 9020, with an Intel Core processing unit with 8.00 GB of RAM and were displayed on a 16-inch Dell monitor with 1024 x 768 screen resolution.

**Automated operation span (AOSPAN; Unsworth, Heitz, Schrock, & Engle, 2005)** Participants first completed the Automated Operation Span (AOSPAN; Unsworth, et al., 2005). During this task, participants were asked to solve simple math problems (e.g.,  $2 + 3 / 2 = ?$ ) while remembering letters in between each math problem. After participants made a “true” or “false” decision via a mouse click on a math problem, a letter would appear for 250 ms for the participant to memorize. After each set of trials, a recall screen was presented listing 12 possible letters and the participant was instructed to click the mouse next to the letters in the correct order that they were presented. The task was composed of three blocks, with each containing five sets of between three to seven trials, for a total of 75 letters and 75 math problems. The AOSPAN was scored by summing the total number of letters recalled in the correct serial position, as recommended by Conway et al. (2005).<sup>2</sup>

**Stroop (Stroop, 1935)** Participants completed a version of the Stroop task based on reminder condition. All three versions contained 75% congruent trials. Stimuli consisted of one of four words (RED, GREEN, BLUE, YELLOW) presented upon a black background 42 times in the congruent color (e.g., the word “GREEN” in green font) and four times in each of the three incongruent colors (e.g., the word “GREEN” in blue, red, or yellow font). Stimuli were presented in the center of the screen in 18-point Courier New font for 3,000 ms or until a response. Participants were instructed to name the color of the written word while ignoring the word itself.

Both accuracy and speed were emphasized, and participants responded by speaking into a microphone. The

<sup>2</sup> We did not use any exclusion criteria for AOSPAN performance. The positive relationship between processing accuracy and storage/recall suggests that using a processing score cutoff would remove more lower span than higher span individuals (see Unsworth, Redick, Heitz, Broadway, & Engle 2009). Of note, our results remain consistent when an 85% criterion is used (see Appendix Tables 3 and 4).



experimenter, seated next to the participant, then coded the participant's response on an attached keyboard in which the keys were labeled with colored stickers. Microphone errors were coded as scratch trials and not analyzed. Following the coding response, a 1,000 ms blank intertrial interval preceded the next stimulus.

## Procedures

This study received permission from the Institutional Review Board at Montana State University. Upon receiving informed consent, participants completed the AOSPAN and were then randomly assigned to either the true control, nongoal reminder, or goal reminder condition. Participants then completed the Stroop task based on condition. During the Stroop task, the true control group received rest breaks every 60 trials, whereas the goal reminder and nongoal reminder groups stopped every 12 trials to vocalize either the task goal (e.g., “name the color not the word”) or a task-irrelevant statement (e.g., “this fulfills my psych 100 requirement”), respectively.<sup>3</sup> The rehearsed statements appeared on the computer screen, and the participants were instructed to read them aloud. Overall, the task contained a total of 216 trials and was preceded with 16 practice trials with the same stimuli and proportion congruency. Stimuli were presented randomly for each participant.

## Results

### WMC

We conducted a one-way analysis of variance (ANOVA) examining the effect of condition on WMC in order to ensure there were no preexisting differences in WMC between our three conditions. The effect of condition was nonsignificant,  $F(2, 205) = .092$ ,  $\eta_p^2 = .001$ ,  $p = .912$ , indicating participants had similar AOSPAN scores in the goal reminder ( $M = 37.54$ ,  $SE = 2.12$ ), nongoal reminder ( $M = 37.10$ ,  $SE = 2.12$ ), and true control ( $M = 38.35$ ,  $SE = 2.06$ ) conditions.

### Stroop errors

The relation between WMC, condition, and Stroop errors are shown in Fig. 1. We conducted three multiple regression analyses. First, we examined the effect of giving participants a goal reminder by comparing the goal reminder condition

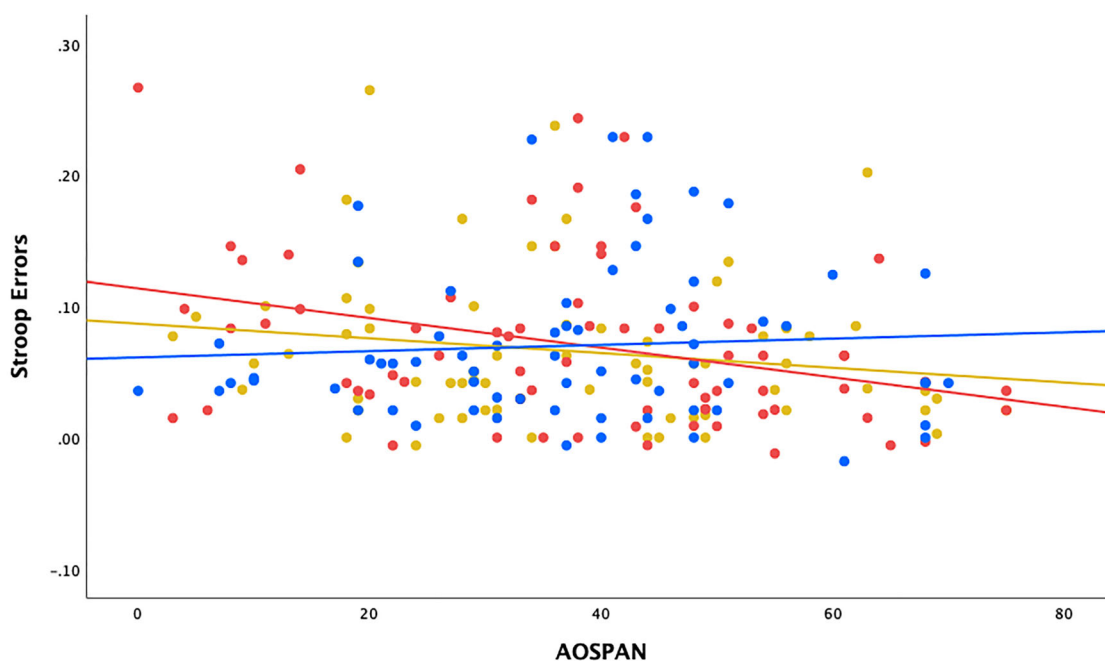
(coded as +1) to the other two conditions (coded as -1, collapsing across our two control conditions). In this analysis, we regressed Stroop errors on reminder condition and WMC (mean-centered AOSPAN score) in Step 1 to examine the main effects and entered the WMC  $\times$  Reminder Condition interaction in Step 2. Next, we examined the two reminder conditions separately, comparing each condition with the true control. Specifically, we compared the goal reminder condition with the true control and the nongoal reminder condition to the true control. This allowed us to obtain the specific comparisons of whether receiving periodic goal reminders or stopping to rehearse a nongoal reminding statement benefits performance relative to the standard condition of simply receiving a rest break every 60 trials. Table 1 shows the results of these analyses.

In addition to the frequentist analyses, we also conducted Bayesian analyses in JASP Version 0.11.1 (JASP Team, 2019) to explore the evidence for (or against) a main effect of WMC, a main effect of reminder condition, and an interaction of WMC  $\times$  Reminder Condition. In each case, we compared the model containing each of these effects to the null model that is missing the component in question. The resulting Bayes factor ( $BF_{10}$ ) shows the ratio of how much better the model with the component predicts the data over the null model missing that component. For instance, a  $BF_{10}$  of 3 indicates the model containing the effect is three times more likely than the null model missing that component and a  $BF_{10}$  of 0.33 means the null model is three times more likely than the model containing the effect. According to the classification scheme from Lee and Wagenmakers (2013; adjusted from Jeffreys, 1961), a  $BF_{10}$  of 10–30 = strong evidence, 3–10 = moderate evidence, 1–3 = anecdotal (weak) evidence, and 1 = no evidence. (Note that values <1 equal evidence for the null, such that 0.33 and 0.10 equal moderate and strong evidence for the null hypothesis, respectively.)<sup>4</sup>

For the comparison of goal reminder versus the combined controls, the main effect of reminder condition was not significant ( $\beta = .013$ ,  $t = 0.19$ ,  $p = .850$ ,  $BF_{10} = 0.22$ ). However, there was an overall main effect of WMC, with larger Stroop effects for those lower in WMC ( $\beta = -.156$ ,  $t = -2.26$ ,  $p = .025$ ,  $BF_{10} = 2.33$ ). There was also a significant WMC  $\times$  Reminder Condition interaction ( $\beta = .361$ ,  $t = 2.11$ ,  $p = .036$ ,  $BF_{10} = 1.99$ ), such that the relation between WMC and Stroop errors was significant in the combined control ( $r = -.256$ ,  $p = .002$ ), but not the goal reminder condition ( $r = .066$ ,  $p = .59$ ).

<sup>3</sup> For the true control group, we considered 60 trials a standard rest break, as 60 trials is close to the average of past studies in our lab with trials per rest break of 40, 100, 72, and 52 in Hutchison (2007), Hutchison et al. (2013), and Hutchison et al. (2016, Experiments 1 & 2), respectively. Although we are not aware of any suggested guidelines for rest breaks, 60 trials between breaks is also consistent with other studies (Cothran & Larson, 2008; Entel & Tvelgov, 2019; Naber, Vedder, Brown, & Nieuwehuis, 2016).

<sup>4</sup> Although these labels help facilitate scientific communication, they are an approximation of standards of evidence, and one should avoid the threshold mentality that plagues null significance hypothesis testing. In paraphrasing Rosnow and Rosenthal (1989), Wagenmakers et al. (2018) stated, “Surely, God loves the Bayes factor of 2.5 nearly as much as he loves the Bayes factor of 3.5” (p. 67).



Note: Blue = Goal Reminder, Gold = Nongoal Reminder, Red = True Control.

Fig. 1 Stroop errors as a function of WMC

For the goal reminder versus true control comparison, neither the main effect of reminder condition ( $\beta = -.007, t = -.078, p = .938, BF_{10} = 0.26$ ), nor the main effect of WMC ( $\beta = -.151, t = -1.79, p = .076, BF_{10} = 1.11$ ) were significant. However, there was a WMC  $\times$  Reminder Condition interaction ( $\beta = .464, t = 2.30, p = .023, BF_{10} = 3.12$ ), such that WMC correlated with Stroop errors in the true control ( $r = -.325, p = .006$ ), but not in the goal reminder condition ( $r = .066, p = .59$ ). In contrast, for the nongoal reminder versus true control comparison, there was only a main effect of WMC ( $\beta = -.255, t = -3.08, p = .003, BF_{10} = 16.77$ ), such that Stroop effects were larger for those lower in WMC. Neither the main effect of Reminder Condition

( $\beta = -.045, t = -0.54, p = .588, BF_{10} = 0.27$ ) nor the WMC  $\times$  Reminder Condition interaction ( $\beta = .196, t = 1.01, p = .315, BF_{10} = 0.46$ ) were significant.

**First-half analyses**

As mentioned previously, Bugg et al. (2015) showed evidence for cognitive control that only lasted for the first half of the trials following the control signal. Because of this, we wanted to examine the effect of getting a break plus goal reminder, relative to the true control, and the effect of receiving a break with no goal reminder, relative to the true control.

Table 1 Results of regression analyses (Stroop errors)

			B	SE	Beta	t	F	Adj. R <sup>2</sup>
Goal Reminder vs. Combined Control	Step 1	AOSPAN	-.054	.024	-.156	-.226*	2.57*	.015
		Reminder Condition	.084	.443	.013	.189		
	Step 2	Reminder Condition $\times$ AOSPAN	.055	.026	.361	.211*	3.23*	.031
True Control vs. Goal Reminder	Step 1	AOSPAN	-.053	.030	-.151	-.179	1.60	.009
		Reminder Condition	-.041	.521	-.007	-.078		
	Step 2	Reminder Condition $\times$ AOSPAN	.068	.030	.464	2.30*	2.86*	.039
True Control vs. Nongoal Reminder	Step 1	AOSPAN	-.087	.028	-.255	-3.08*	4.84*	.053
		Reminder Condition	-.237	.502	-.045	-.543		
	Step 2	Reminder Condition $\times$ AOSPAN	.028	.028	.196	1.01	3.57	.053

Note. AOSPAN = Automated Operation Span. \* $p < .05$

Specifically, we examined whether effects of goal reminder and/or break would be stronger for the first half of the trials after the break (relative to the same yoked six trials in the true control condition) than for the second half of the trials after the break.

For the goal reminder versus true control comparison, the results show the benefit of goal reminding obtained above was primarily due to the first half of the trials after the reminder. Specifically, the interaction of WMC and Condition was significant in the first half of the trials ( $\beta = .468, t = 2.30, p = .023, BF_{10} = 3.16$ ), but not in the second half of the trials ( $\beta = .348, t = 1.72, p = .088, BF_{10} = 1.13$ ). In contrast, the main effect of WMC was not significant in the first half of the trials ( $\beta = -.061, t = -0.71, p = .478, BF_{10} = 0.34$ ), but was significant in the second half of the trials ( $\beta = -.178, t = -2.11, p = .037, BF_{10} = 1.93$ ). Finally, there was no main effect of condition in either the first ( $\beta = -.032, t = -0.38, p = .706, BF_{10} = 0.29$ ) or the second half of the trials ( $\beta = .020, t = .232, p = .817, BF_{10} = 0.26$ ).

For the nongoal reminder versus true control comparison, the overall pattern of only a main effect of WMC was consistent for both the first ( $\beta = -.203, t = -2.42, p = .017, BF_{10} = 3.54$ ) and second ( $\beta = -.207, t = -2.46, p = .015, BF_{10} = 3.87$ ) half of the trials. There was no effect of reminder condition in either the first ( $\beta = -.051, t = -.612, p = .541, BF_{10} = 0.29$ ) or second ( $\beta = -.005, t = -.056, p = .955, BF_{10} = 0.25$ ) half of the trials. Finally, there was no WMC  $\times$  Reminder Condition interaction in either the first half ( $\beta = .120, t = 0.61, p = .545, BF_{10} = 0.36$ ) or second half ( $\beta = .270, t = 1.38, p = .171, BF_{10} = 0.70$ ) of the trials.

## Reaction times

The reaction time (RT) analysis was conducted on accurate responses only. We first removed any trials with a RT less than 50 (which removed 0.8% of RTs). We then used Van Selst and Jolicoeur's (1994) nonrecursive method, which removed an additional 2.9% of RTs. The results from the reaction time data are presented in Table 2. To anticipate, none of the effects were significant.

As with errors, we first examined the effect of giving participants a goal reminder by comparing the goal reminder condition (coded as +1) to the other two conditions (coded as -1, collapsing across our two control conditions). In this analysis, we regressed Stroop RT on reminder condition and WMC (mean-centered AOSPAN score) in Step 1 to examine the main effects and entered the WMC  $\times$  Reminder Condition interaction in Step 2. These results are shown at the top of Table 2. There was no main effect of WMC [ $\beta = -.065, t = -.933, p = .352, BF_{10} = 0.33$ ] or reminder condition ( $\beta = .098, t = 1.41, p = .160, BF_{10} = 0.57$ ). The interaction was also nonsignificant ( $\beta = .225, t = 1.30, p = .195, BF_{10} = 0.59$ ).

For the goal reminder versus true control comparison, there was no main effect of WMC ( $\beta = -.059, t = -.687, p = .493, BF_{10} = 0.33$ ) or reminder condition ( $\beta = .065, t = .760, p = .449, BF_{10} = 0.35$ ). The interaction was also nonsignificant ( $\beta = .280, t = 1.36, p = .175, BF_{10} = 0.73$ ). For the nongoal reminder versus true control comparison, there was no main effect of WMC ( $\beta = -.134, t = -1.58, p = .117, BF_{10} = 0.81$ ) or reminder condition ( $\beta = -.083, t = -.974, p = .332, BF_{10} = 0.40$ ). The interaction was also nonsignificant ( $\beta = .106, t = .532, p = .596, BF_{10} = 0.36$ ).

We then examined RTs during only the first half of the trials following the goal reminder. These effects were consistent with the overall data. For the goal reminder versus true control comparison, there was no main effect of WMC ( $\beta = -.082, t = -.961, p = .338, BF_{10} = 0.41$ ) or reminder condition ( $\beta = .075, t = .883, p = .379, BF_{10} = 0.38$ ). The interaction was also nonsignificant ( $\beta = .320, t = 1.56, p = .121, BF_{10} = 0.93$ ). For the nongoal reminder versus true control comparison, there was no main effect of WMC ( $\beta = -.149, t = -1.76, p = .081, BF_{10} = 1.05$ ) or reminder condition ( $\beta = -.073, t = -.858, p = .392, BF_{10} = 0.36$ ). The interaction was also nonsignificant ( $\beta = .187, t = .943, p = .347, BF_{10} = 0.47$ ).

## Composite measure

Although these RT effects were nonsignificant, they were in the same direction and had the same qualitative pattern as the error effects, with Stroop effects numerically decreasing with WMC in the true control and nongoal reminder conditions, yet numerically increasing with WMC in the goal reminder condition. Thus, as discussed in the Introduction, it is likely that effects of goal maintenance also affect Stroop RTs, but that the effect is not as robust as in errors. To test this, we combined Stroop RT and error effects into a  $z$ -score composite measure (i.e., average standardized Stroop effects across RTs and errors). To summarize, this composite analysis replicated all our previous effects obtained in the error analyses. Specifically, for the comparison of goal reminder versus the combined controls, there was again a significant WMC  $\times$  Condition interaction ( $\beta = .369, t = 2.16, p = .032, BF_{10} = 2.18$ ), such that the relation between WMC and Stroop errors was significant in the combined control ( $r = -.242, p = .004$ ), but not the goal reminder condition ( $r = .086, p = .49$ ). This interaction also replicated for the goal-reminder versus true control condition ( $\beta = .469, t = 2.32, p = .022, BF_{10} = 3.25$ ), with WMC correlating with Stroop errors in the true control ( $r = -.312, p = .008$ ), but not in the goal reminder condition ( $r = .086, p = .49$ ). In contrast, and as before, for the nongoal reminder versus true control comparison, there was only a main effect of WMC ( $\beta = -.244, t = -2.94, p = .004, BF_{10} = 11.74$ ), such that Stroop effects were larger for those lower in WMC. The WMC  $\times$  Condition interaction ( $\beta = .189, t = 0.97, p = .332, BF_{10} = 0.44$ ) was again not significant.

**Table 2** Results of regression analyses (Stroop reaction times)

			<i>B</i>	<i>SE</i>	Beta	<i>t</i>	<i>F</i>	Adj. <i>R</i> <sup>2</sup>
Goal Reminder vs. Combined Control	Step 1	AOSPAN	-.235	.252	-.065	-.933	1.44	.004
		Reminder Condition	6.58	4.66	.098	1.41		
	Step 2	Reminder Condition × AOSPAN	.359	.276	.225	1.30	1.52	.008
True Control vs. Goal Reminder	Step 1	AOSPAN	-.218	.317	-.059	-.687	.535	-.007
		Reminder Condition	4.19	5.52	.065	.760		
	Step 2	Reminder Condition × AOSPAN	.435	.319	.280	1.36	.978	.000
True Control vs. Nongoal Reminder	Step 1	AOSPAN	-.454	.288	-.134	-1.58	1.68	.010
		Reminder Condition	-5.00	5.13	-.083	-.974		
	Step 2	Reminder Condition × AOSPAN	.154	.289	.106	.532	1.21	.004

Note. AOSPAN = Automated Operation Span

## Discussion

In the current study, we investigated the role of goal maintenance in MC Stroop lists. We found that providing goal reminders eliminates the relation between WMC and Stroop errors. Specifically, our control condition replicated previous findings, demonstrating a negative correlation between WMC and Stroop errors. In contrast, our goal reminder condition eliminated WMC-related differences in Stroop errors. When separately comparing the goal reminder and nongoal reminder conditions to the true control, only the goal reminder condition significantly differed from the true control. Specifically, the goal reminder comparison showed a significant WMC × Reminder Condition interaction, whereas the nongoal reminder only showed only a main effect of WMC and no interaction.

The current results add support to the goal-maintenance explanation of WMC-related differences in task performance (Engle, 2001). According to this account, higher WMC individuals have superior performance in conflict tasks such as the Stroop task due to their ability to internally maintain the goal of the task. In contrast, lower WMC individuals lack this internal goal maintenance. However, previous research has shown that once the task-context has been manipulated to include external support, WMC-related differences in performance no longer exist. Specifically, when the percentage of incongruent trials increases (as in MI lists), little demand is placed on working memory, as the frequent incongruent trials serve as external reminders of the task goal (Kane & Engle, 2003). It is in this environment where we no longer see WMC-related differences, as the external support of the list boosts their performance.

As discussed previously, a few studies have also attempted to provide external support in the form of precueing (Bugg et al., 2015; Correa et al., 2009; Gratton et al., 1992; Hutchison et al., 2016; Logan & Zbrodoff, 1982; see Bugg & Smallwood, 2014, for a review).

Although precueing has been shown to successfully influence expectations of upcoming conflict, the cues used in previous studies to signal a likely congruent trial (e.g., “easy,” “matching”) draw attention to the supposed “irrelevant” dimension, changing the nature of the task. In contrast to using precues, the reminders used in the current study allowed for a cleaner method of providing external goal support, by simply reminding participants of the task instructions every 12 trials.

Despite receiving a great deal of early support, some researchers have challenged the assumption that list-based effects in Stroop performance actually reflect differences in sustained goal maintenance across trials. In addition to the frequent ISPC confound discussed in the Introduction, there are other accounts of larger Stroop effects in MC lists. One possibility is that list-wide PC effects could be due not to sustained internal goal maintenance across trials, but instead to short-lived conflict adaptation in which Stroop effects diminish following a previous incongruent trial (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Gratton et al., 1992). Such sequence effects (i.e., larger Stroop effects following previous congruent than incongruent trials) are typically explained as due to stimulus-triggered goal retrieval and subsequent top-down control that is signaled by the conflict experienced during an incongruent trial (but see Schmidt, 2013b; Verguts & Notebaert, 2009, for other explanations of supposed conflict adaptation effects). Importantly, however, the current paradigm eliminates this explanation, as all three conditions used the same PC list. Therefore, although sequential effects may have contributed to performance, they do not explain the difference in performance across goal reminder groups.

Another alternative account of list-wide effects is that they reflect temporal learning of response times (Schmidt 2013a, 2014). According to this account, people in MC lists develop a rhythm of responding more quickly, which causes greater problems when they encounter an



incongruent item (but see Cohen-Shikora, Suh, & Bugg, 2019; Spinelli, Perry, & Lupker, 2019, for evidence against the temporal learning account of list-wide PC effects). There is at least some support for this account in our data. Specifically, the effect of WMC on Stroop errors in the nongoal reminder condition is in-between that of the goal reminder and true control conditions, while not significantly differing from either. Thus, breaking up response rhythms by participants every 12 trials to vocalize a rehearsed statement may partially protect participants from errors due to such temporal learning effects. It is unclear, however, the extent to which this differs from a goal-maintenance account. In fact, falling into a rhythmic pattern of quick responding is often a signal of mind wandering (McVay & Kane, 2009). For instance, McVay and Kane (2009) found that RTs on the four trials preceding off-task reports were significantly faster than those preceding on-task thoughts. Furthermore, individuals lower in WMC are generally more likely to mind wander (Kane et al., 2007). Such an account, in which a pattern of quick habitual responding occurs during mind wandering episodes, is consistent with Meier and Kane's (2013) finding that individuals lower in WMC fall into the habitual response of word reading when there are many consecutive, uninterrupted congruent trials. Nonetheless, it is important to note that only the goal reminder condition showed a qualitatively different pattern of WMC effects on Stroop errors, and only this condition significantly reduced the influence of WMC, relative to the true control condition. Perhaps future studies could replicate the current results with goal reminders that appear less often, allowing for more consecutive, uninterrupted congruent trials, to help rule out that simply providing a rest-break is enough to reduce or eliminate WMC-related differences in performance.

In regard to our RT results, although nonsignificant, they were in the same direction and had the same qualitative pattern as the error effects. Importantly, when we combined RT and errors into one composite measure, all of our effects replicated. This supports previous findings that RT effects are likely not as robust to problems with goal maintenance as are error effects (Hutchison et al., 2010; Hutchison et al., 2013; Spieler et al., 1996). Recently, Draheim, Mashburn, Martin, and Engle (2019) described psychometric issues for why Stroop RT results may not be as sensitive to individual differences in attentional control. Specifically, when examining a difference score measure, the reliability of the measure decreases as the correlation between the two component scores increases, which reduces any possible outside correlation with another variable. Although this is an issue whenever using

difference scores, it was a particular problem for our RT measure, in which the correlation between RTs in the congruent and incongruent condition was  $+0.764$ ,  $p < .001$ . In contrast, participants essentially made zero errors in the congruent condition [congruent condition = 0.5% errors, range 0%–3%, with 46% of the participants having zero errors], resulting in a much smaller correlation in errors ( $r = .129$ ,  $p = .062$ ). In fact, because there were almost no errors in the congruent condition, all of our results replicate if we only use incongruent errors alone as our outcome variable (as opposed to using the incongruent – congruent difference score).

When examining our first half versus second half analyses, we found the benefit of goal reminding was primarily due to the first half of the trials after the goal reminder. Specifically, the WMC  $\times$  Reminder Condition interaction was significant in the first half of the trials, but not in the second half. In contrast, the main effect of WMC was significant in the second half of the trials, but not the first half. This finding suggests a signature of the temporal dynamics of goal reminders in initially overcoming the goal neglect experienced by individuals lower in WMC, but ultimately failing to have a sustained effect. Perhaps future studies could more precisely explore the time course of goal reminder benefits on performance.

## Conclusion

Individual differences in WMC are seen in many tasks, but are particularly evident on conflict tasks such as the Stroop. These WMC differences are thought to occur due to differences in the ability to guide the focus of attention in a goal-directed manner. In the current study, providing participants with a reminder of the task goal eliminated the typical WMC-related differences in Stroop performance. The form of external support in the current study allowed for a purer test of goal maintenance compared with previous studies, while providing further evidence for the goal-maintenance explanation of WMC-related differences in task performance.

**Author note** The authors would like to thank Dr. Emily Cohen-Shikora for her helpful review. We also thank Max Richey for his help with data collection.

**Open practices statement** The experiment was not preregistered. Complete data files are available at <http://www.montana.edu/atmmemlab/>.

## Appendix

**Table 3** Results of regression analyses (Stroop errors) after removing data from participants with 85% or more processing errors on the AOSPAN

			<i>B</i>	<i>SE</i>	Beta	<i>t</i>	<i>F</i>	Adj. <i>R</i> <sup>2</sup>
Goal Reminder vs. Combined Control	Step 1	AOSPAN	−.053	.025	−.152	−.215*	2.43*	.014
		Reminder Condition	.195	.453	.030	.429		
	Step 2	Reminder Condition × AOSPAN	.053	.027	.348	1.98*	2.94*	.028
True Control vs. Goal Reminder	Step 1	AOSPAN	−.054	.031	−.149	−.172	1.50	.008
		Reminder Condition	.071	.538	.012	.133		
	Step 2	Reminder Condition × AOSPAN	.068	.031	.462	2.19*	2.63*	.036
True Control vs. Nongoal Reminder	Step 1	AOSPAN	−.087	.030	−.247	−2.92*	4.32*	.048
		Reminder Condition	−.277	.519	−.045	−.534		
	Step 2	Reminder Condition × AOSPAN	.030	.030	.209	1.02	3.23	.048

Note. AOSPAN = Automated Operation Span. \* $p < .05$

**Table 4** Results of regression analyses (Stroop reaction times) after removing data from participants with 85% or more processing errors on the AOSPAN

			<i>B</i>	<i>SE</i>	Beta	<i>t</i>	<i>F</i>	Adj. <i>R</i> <sup>2</sup>
Goal Reminder vs. Combined Control	Step 1	AOSPAN	−.311	.261	−.084	−1.19	1.69	.007
		Reminder Condition	6.56	4.77	.097	1.38		
	Step 2	Reminder Condition × AOSPAN	.397	.283	.249	1.40	1.79	.012
True Control vs. Goal Reminder	Step 1	AOSPAN	−.355	.330	−.094	−1.08	.792	−.003
		Reminder Condition	3.42	5.69	.052	.601		
	Step 2	Reminder Condition × AOSPAN	.547	.329	.355	1.66	1.46	.010
True Control vs. Nongoal Reminder	Step 1	AOSPAN	−.573	.299	−.165	−1.92	2.48	.022
		Reminder Condition	−6.47	5.23	−.106	−1.24		
	Step 2	Reminder Condition × AOSPAN	.295	.300	.205	.985	1.98	.022

Note. AOSPAN = Automated Operation Span

## References

- Balota, D. A., Tse, C. S., Hutchison, K. A., Spieler, D. H., Duchek, J. M., & Morris, J. C. (2010). Predicting conversion to dementia of the Alzheimer's type in a healthy control sample: The power of Stroop performance. *Psychology and Aging, 25*, 208–218.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624–652.
- Bugg, J. M., & Chanani, S. (2011). List-wide control is not entirely elusive: Evidence from picture-word Stroop. *Psychonomic Bulletin & Review, 18*, 930–936.
- Bugg, J. M., Diede, N. T., Cohen-Shikora, E. R., & Selmeczy, D. (2015). Expectations and experience: Dissociable bases for cognitive control? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1349–1373.
- Bugg, J. M., McDaniel, M. A., Scullin, M. K., & Braver, T. S. (2011). Revealing list-level control in the Stroop task by uncovering its benefits and a cost. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1595–1606.
- Bugg, J. M., & Smallwood, A. (2014). The next trial will be conflicting! The effects of explicitly pre-cueing congruency on cognitive control. *Psychological Research, 80*, 16–33.
- Cohen-Shikora, E. R., Diede, N. T., & Bugg, J. M. (2018). The flexibility of cognitive control: Age equivalence with experience guiding the way. *Psychology and Aging, 33*, 924–939.
- Cohen-Shikora, E. R., Suh, J., & Bugg, J. M. (2019). Assessing the temporal learning account of list-wide proportion congruence effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 1703–1723.
- Conway, A. R., Cowan, N. A., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Bulletin and Review, 8*, 331–335.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, Z. D., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769–786.

- Correa, A., Rao, A., & Nobre, A. C. (2009). Anticipating conflict facilitates controlled stimulus–response selection. *Journal of Cognitive Neuroscience*, *8*, 1461–1472.
- Cothran, D. L., & Larsen, R. (2008). Comparison of inhibition in two timed reaction tasks: The color and emotion Stroop tasks. *The Journal of Psychology: Interdisciplinary and Applied*, *142*, 373–385.
- Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, *145*, 508–535.
- 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.
- Engle, R. W. (2001). What is working memory capacity? In H. L. Roediger, III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *Science conference series: The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 297–314). Washington, DC: American Psychological Association.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working capacity, and a two-factor theory of cognitive control. *The Psychology of Learning and Motivation*, *44*, 145–199.
- Entel, O., & Tzelgov, J. (2019). When working memory meets control in the Stroop effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <https://doi.org/10.1037/xlm0000790>
- Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*, 480–506.
- 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.
- Hutchison, K. A., Balota, D. A., & Duchek, J. M. (2010). The utility of Stroop task switching as a marker for early stage Alzheimer’s disease. *Psychology and Aging*, *25*, 545–559.
- Hutchison, K. A., Bugg, J. M., Lim, Y. B., & Olsen, M. R. (2016). Congruency precues moderate item-specific proportion congruency effects. *Attention, Perception, & Psychophysics*, *78*, 1087–1103.
- Hutchison, K. A., Smith, J. L., & Ferris, A. (2013). Goals can be threatened to extinction: Using the Stroop task to clarify working memory depletion under stereotype threat. *Social Psychological and Personality Science*, *4*, 74–81.
- Jacoby, L. J., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*, 638–644.
- JASP Team. (2019). JASP (Version 0.11.1) [Computer software]. Retrieved from <https://jasp-stats.org/>. Accessed 22 May 2020.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.
- Kane, M. J., Bleckley, K. M., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, *130*, 169–183.
- Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeyns, M., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science*, *18*(7), 614–621. <https://doi.org/10.1111/j.1467-9280.2007.01948.x>
- 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.
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. New York, NY: Cambridge University Press.
- Logan, G. D., & Zbrodoff, N. J. (1982). Constraints on strategy construction in a speeded discrimination task. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 502–520.
- Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition*, *30*, 294–301.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163–203.
- 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: Learning, Memory, and Cognition*, *35*, 196–204.
- Meier, M. E., & Kane, M. J. (2013). Working memory capacity and Stroop interference: Global versus local indices of executive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 748–759.
- Miller, A., Watson, J. M., & Strayer, D. L. (2012). Individual differences in working memory capacity predict action monitoring and the error-related negativity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 757–763.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.
- 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*, 250–260.
- Naber, M., Vedder, A., Brown, S., & Nieuwenhuis, S. (2016). Speed and lateral inhibition of stimulus processing contribute to individual differences in Stroop-task performance. *Frontiers in Psychology*, *7*, 1–13.
- Redick, T. S. (2014). Cognitive control in context: Working memory capacity and proactive control. *Acta Psychologica*, *145*, 1–9.
- Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276–1284.
- Schmidt, J. R. (2013a). Temporal learning and list-level proportion congruency: Conflict adaptation of learning when to respond? *PLOS ONE*, *8*, 1–12.
- Schmidt, J. R. (2013b). Questioning conflict adaptation: Proportion congruent and Gratton effects reconsidered? *Psychonomic Bulletin & Review*, *20*, 615–630.
- Schmidt, J. R. (2014). List-level transfer effects in temporal learning: Further complications for the list-level proportion congruent effect. *Journal of Cognitive Psychology*, *26*, 373–385.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.
- Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer’s type. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 461–479.
- Spinelli, G., Perry, J. R., & Lupker, S. J. (2019). Adaptation to conflict frequency without contingency and temporal learning: Evidence from the picture-word interference task. *Journal of Experimental Psychology: Human Perception and Performance*, *45*, 995–1014.

- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavioral Research Methods, 37*, 498–505.
- Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*, 635–654.
- Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*, 392–406.
- Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology, 47A*, 631–650.
- Verguts, T., & Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive control. *Trends in Cognitive Sciences, 13*, 252–257.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.