



Once established, goal reminders provide long-lasting and cumulative benefits for lower working memory capacity individuals.

Audrey V. B. Hood, Brooke Charbonneau, Keith A. Hutchison

© American Psychological Association, 2022-12-01. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: <https://doi.org/10.1037/xlm0001185>

Once established, goal reminders provide long-lasting and cumulative benefits for lower working memory capacity individuals

Audrey V. B. Hood, Brooke Charbonneau, and Keith A. Hutchison

Montana State University

Manuscript Accepted: JEP:LMC

Author Note

The authors would like to thank Abbigail Sparks, Ashley Quinto, Donovan Olson, Garrett Loders, Sara Snyder, Aubrey Williams, Bernard Lacey, Emily Blust, and Zoe Baker for their help with data collection. All R scripts, Data files, Output, and stimulus materials are available at <https://osf.io/3ej8b> (Hutchison et al., 2022).

Please Address Correspondence to:

Keith A. Hutchison
Department of Psychology
Montana State University, P.O. Box 173440
Bozeman, MT 59717-3440 USA
E-mail: khutch@montana.edu
Phone: 406.994.5528

Abstract

Previous research has shown that Stroop effects interact with working memory capacity (WMC) more strongly with lists of mostly congruent items. Although the predominant explanation for this relationship is goal-maintenance, some research has challenged whether list-wide effects truly reflect goal-maintenance abilities. The current study improved upon previous methodology by using both within-subject and between-subject manipulations of goal reminder, increasing both the number of trials between reminders and the total length of the task to allow for greater goal neglect, and more precisely maintaining congruency proportion within each block.

Participants completed the Automated Operation Span followed by a Stroop task in which they stopped every 24 trials to vocalize either a goal-reminder statement (“*name the color not the word*”) or a non-goal statement (“*This is part of my intro to psychology class*”). In the within-subject manipulation (Experiment 1), there was no consistent benefit for goal reminders over non-goal statements. However, in the between-subject manipulation (Experiment 2), results demonstrated a strong benefit of goal reminders, such that goal reminders eliminated the relation between WMC and Stroop effects, whereas that relation was robust following non-goal statements. Moreover, the benefit of receiving goal reminders lasted for at least 24 trials and accumulated across the course of the experiment. These data provide strong evidence that goal reminders eliminate the relationship between WMC and Stroop errors and suggest goal reminders can be a useful intervention for those suffering from lapses in controlled attention.

Keywords: cognitive control, Stroop interference, goal neglect, working memory capacity

Once established, goal reminders provide long-lasting and cumulative benefits for lower working memory capacity individuals

Working memory capacity (WMC) refers to the ability to maintain information in the face of distractions. Newer conceptualizations demonstrate that WMC reflects maintenance of information enabled by attention control, which is critical for maintaining attention on goal relevant information and avoiding external and internal distractions (Burgoyne & Engle; Shipstead et al., 2016). As a result, individuals higher in WMC are better at maintaining task goals in an active state, leading to selective attention of task-relevant stimuli and producing task-appropriate responses (Conway et al., 2001; Engle, 2002; Kane et al., 2001). In contrast, lower WMC individuals are often subject to goal neglect (Duncan et al., 1996), leading to more interference from task-irrelevant stimuli and automatic, but task-inappropriate responses. Such differences in the ability to control attention lead to WMC-related performance differences across many attention-related tasks, including dichotic listening (Conway et al., 2001), the antisaccade task (Kane et al., 2001; Unsworth et al., 2004), the Simon task (Miller et al., 2012), the Sustained Attention to Response task (McVay & Kane, 2009), and the AX-CPT (Redick, 2014).

Conflict tasks, such as the Stroop task (Stroop, 1935), provide a particularly strong demonstration of WMC-related differences in the ability to control attention because they require suppressing a habitual task-inappropriate response in order to give a more controlled task-appropriate response. In the Stroop task, participants are presented with color names presented in either congruent font (e.g., GREEN in green font) or incongruent font (GREEN in blue font). The task is to name the font color while ignoring the word itself. Thus, successful performance on the Stroop task requires suppressing the strong, habitual response to say the word (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).

Within the Stroop task, Kane and Engle (2003) found that the percentage of congruent trials moderates the degree to which WMC interacts with task performance. They argued that, in mostly incongruent (MI) Stroop lists, the response conflict triggered during each incongruent trial serves as a reminder of the task goal to name the color, rather than the word, providing external support for goal maintenance. As a result, WMC-related differences are not as strong and not consistently found in MI Stroop lists. In contrast, in mostly congruent (MC) Stroop lists, the infrequent number of incongruent trials requires participants to internally maintain the task goal in order to accurately respond to the rare incongruent trials. Here, higher WMC individuals outperform lower WMC individuals, presumably because they can maintain the task goal internally, whereas lower WMC individuals are more dependent upon external support (i.e., reminders in the form of frequent incongruent trials) to avoid goal neglect.

In addition to producing more errors, lower WMC individuals also typically have slower reaction times. This pattern of more errors and slower reaction times among lower WMC individuals can be explained by Kane and Engle's (2003) two-process model of goal maintenance and conflict resolution. For instance, when participants neglect or forget the task goal, Stroop effects manifest mainly in errors. In contrast, when participants attempt to resolve the competition between conflicting word and color responses on incongruent trials, Stroop effects manifest mainly in reaction times.

Although the goal-maintenance account of WMC differences in Stroop effects has 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 over time. Two alternative accounts include item-specific proportion congruency differences across lists (Jacoby et al., 2003) and temporal learning of response times (Schmidt, 2013).

The item-specific proportion congruency effect refers to the finding of reduced Stroop effects for specific words that usually appear in an incongruent font color, relative to words that usually appear in a congruent font color. If listwide PC effects are indeed due solely to item-level effects within these lists, rather than differences in goal maintenance, it would require a different explanation for why WMC differences only emerge in MC lists. Instead, an alternative possibility could be that MC items capture attention and/or trigger word reading more strongly for lower WMC individuals than higher WMC, perhaps because lower WMC individuals are less able to suppress the habitual word reading response triggered by these items (Kane et al., 2001; Kane & Engle, 2002). However, more recent studies have cast doubt on this alternative explanation by demonstrating that the listwide PC effect is not simply an item-specific PC effect in disguise. Specifically, list-level PC effects emerge even when item-specific PC levels are held constant (Bugg & Chanani, 2011; Bugg et al., 2011; Hutchison, 2011; Spinelli & Lupker, 2022). Moreover, Hutchison (2011) found larger WMC differences in MC lists even when item-specific PC effects were held constant. Nonetheless, it is still possible that part of enhanced WMC differences in Stroop performance within MC lists is due to the MC items within these lists reactively triggering word reading.

A second alternative account of listwide PC effects is that they reflect temporal learning of response times (Schmidt 2013; Schmidt, 2014). According to this account, MC Stroop lists can lead people to develop a rhythm of responding more quickly, which causes disruption when

they encounter the less frequent, incongruent items, producing larger Stroop effects in such lists. This would suggest that part of the list-PC effect is caused by temporal learning. Although, again, it is not clear how such temporal learning differences could explain larger WMC differences in MC lists, although it is possible that those lower in WMC are more influenced by temporal learning than those higher in WMC. However, as with the item-specific PC explanation, there is recent evidence against the temporal learning account of listwide PC effects in general. Specifically, Cohen-Shikora et al. (2019) examined the robustness of this account by conducting analyses on multiple published datasets in which list-wide PC effects were observed. Their comprehensive evaluation failed to find strong or consistent evidence of the temporal learning account. Further, they demonstrated that list-wide PC effects remained significant even after controlling for temporal learning (c.f., Spinelli et al., 2019, for similar findings). Nonetheless, in the absence of such manipulations, it is possible temporal learning can contribute to listwide PC effects and, in turn, could influence WMC differences across lists.

Because of the issues raised above, differences in Stroop effects between MC and MI lists probably reflect more than simply goal maintenance differences across lists. As a result, researchers need a more straight-forward measure of goal maintenance than simply using different PC lists and inferring maintenance differences across them. One such alternative approach uses precues to signal the probability of congruency for an upcoming Stroop trial or list (e.g., Bugg et al., 2015; Correa et al., 2009; Gratton et al., 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 proportion congruency constant (at 50%). Results demonstrated larger Stroop effects when participants were cued that an upcoming list was going to be MC, compared to when they were cued that an

upcoming list was going to be MI. However, this was only significant for the first half of the Stroop lists. Further, Hutchison et al. (2016) found that precuing participants eliminated both listwise and item-specific PC effects, suggesting that cuing participants to engage control prior to stimulus onset reduces the impact of list-based and item-based cues that direct attention toward or away from word reading.

Although such precuing studies can manipulate expectations of upcoming conflict, an important critique is that such cues are unnatural for the Stroop task. Typical instructions for the Stroop task instruct participants to respond to the color and ignore the word. Therefore, using precues that specify the likelihood of congruency of the distractor word directs participants' attention toward what is normally considered the "irrelevant" dimension in Stroop studies (word reading). Because of this, we believe that a more appropriate manipulation is to use explicit goal reminders, as they allow for a straight-forward alternative manipulation of goal maintenance. For this reason, we first provide a very brief review of developmental studies on the effectiveness of providing goal reminders to children and then discuss our lab's recent extension of this method for young adults.

Directly Testing the Goal Reminder Account

Although a goal reminder intervention has not been used for those lower in WMC, it has been used with children, who also struggle with AC and have poor executive function (Diamond, 2013). For instance, Gonthier et al. (2021) had preschoolers complete Stroop and Flanker tasks and manipulated listwise PC in both tasks while controlling for item-specific PC. Children's performance on these tasks mimicked that of lower WMC individuals. Specifically, they had larger Stroop and Flanker effects for MC lists, suggesting they are dependent on external support provided by frequent incongruent items within a list. Relatedly, Simpson and Riggs (2005; 2007)

found that, even though young children can accurately answer questions about task rules when prompted, they still neglect these rules when performing the tasks, especially during tasks that require suppressing well-practiced responses.

Because of such early deficits in goal maintenance, providing goal reminders can be especially helpful for children. Barker and Munakata (2015) gave three-year-old children goal orienting reminders in a go/no-go task and found that these reminders benefited cognitive control, such that children in the reminder conditions had fewer no-go intrusion errors, suggesting greater suppression of habitual responding. Additional studies by Blaye and Chevalier (2011, see also Chevalier & Blaye, 2009) have demonstrated that such benefits are stronger when reminders are more descriptive and clearer about the goal of the given task. Together, these results demonstrate that goal reminders can improve task performance in young children by improving explicit control over performance and therefore may help those lower in WMC.

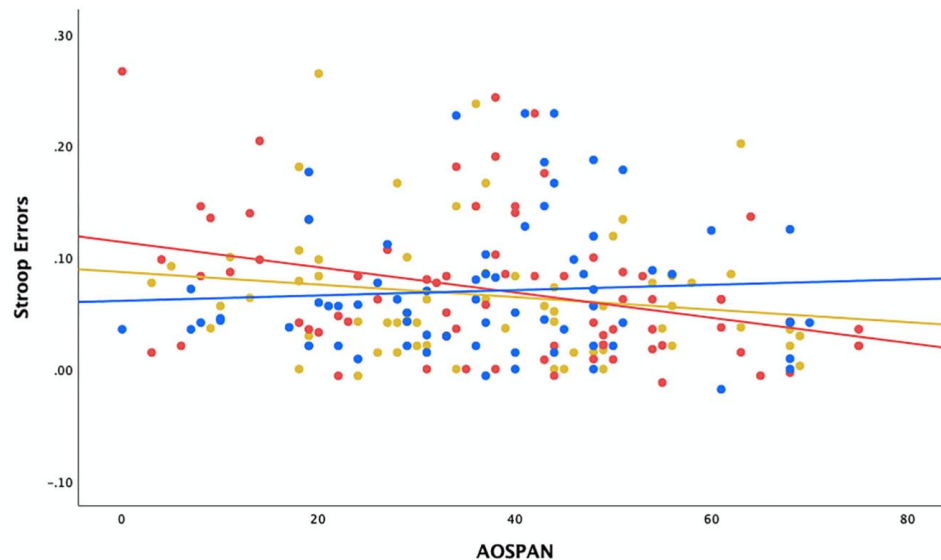
Our lab recently extended the use of goal reminders within a young adult population (Hood & Hutchison, 2021). In that study, participants first completed the Automated Operation Span (AOSPAN) followed by a Stroop task with a MC list in which all items were 75% congruent. At the beginning of the Stroop task, all participants were given instructions stating that *“the goal of this task is to name the color (not the word itself)”* and to *“please respond both quickly and accurately.”* However, during the Stroop task, participants in the goal-reminder condition stopped every 12 trials to vocalize the task goal of *“the goal is to name the color, not the word”* when given the prompt *“the goal is to name the ----- not the -----.”* In contrast, participants in the non-goal statement condition also stopped every 12 trials, but instead vocalized a non-goal statement of *“this task is for my intro to psychology class”* when given the prompt *“this task is for my ----- to ----- class.”* Lastly, participants in the true control

condition did not rehearse any statement. Instead, they simply received a rest break every 60 trials, as in typical Stroop paradigms (Cothran & Larson, 2008; Entel & Tvelgov, 2019; Hutchison 2007; Hutchison et al. 2013; Hutchison et al. 2016; Naber et al., 2016).

The main finding of this study is shown in Figure 1. We first examined the effect of giving participants a goal reminder by comparing the goal reminder condition to the other two conditions (collapsing across control conditions). As can be seen in the figure, we replicated the typical finding that individuals higher in WMC demonstrated fewer Stroop errors (Hutchison, 2007; 2011; Kane & Engle, 2003; Long & Prat, 2002). Also, importantly, having participants vocalize the task goal every 12-trials eliminated the relationship between WMC and Stroop errors.

Figure 1.

Stroop Errors as a Function of Goal Reminder Condition and WMC. Figure taken from Hood and Hutchison (2021) with permission.



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

When testing each reminder condition relative to the true control, we found that only the goal reminder condition significantly differed from the true control. Specifically, there was a

significant $WMC \times$ Goal Reminder interaction when comparing the goal reminder (blue line in Figure 1) to the true control (red line in Figure 1), with WMC predicting Stroop effects in the true control condition, but not in the goal reminder condition. In contrast, when comparing the non-goal reminder (gold line in Figure 1) to the true control, there was only a main effect of WMC and no interaction with reminder condition.

In addition to this main analysis, we examined whether effects of goal reminders would be stronger for the first half of trials after a break (relative to the same yoked six trials in the true control condition) than for the second half of the trials after a break. We found some evidence the benefit of goal reminding might be stronger for 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 (although the 3-way $WMC \times$ Reminder Condition \times Half interaction was not significant). This suggests possible temporal dynamics of goal reminders in initially overcoming the goal neglect experienced by individuals lower in WMC, but potentially failing to have a sustained effect. Thus, it is possible the effect of receiving goal reminders is short-lived. However, because the three-way $WMC \times$ Reminder Condition \times Half interaction was not significant, this possibility remains speculative.

Current Study

The purpose of the current study is to address the above-mentioned concerns of Hood and Hutchison (2021), with the aim of providing a cleaner and more powerful test of the goal-maintenance explanation for WMC-related differences in Stroop performance. Specifically, we wanted to address two ambiguous results regarding Hood and Hutchison's findings. First, although only the Goal Reminder condition significantly differed from the true control in its relation to WMC, there was no difference between the relation of WMC to Stroop effects

between the Goal Reminder and Non-Goal Statement conditions when directly compared. This suggests the temporal breaks or vocalized statements present in the Non-Goal condition might also have reduced the relation between WMC and Stroop errors. Second, the lack of a three-way i Reminder Condition \times WMC \times Half interaction leaves open the question of how long these goal reminders last in helping to maintain cognitive control across trials.

To address these issues, we made four main changes to the Hood and Hutchison methodology. First, given the possibility that goal reminders are short-lived, we implemented a more powerful within-subject manipulation of reminder condition, with each participant receiving both goal reminder and non-goal statements throughout the experiment. Second, we included more trials within blocks so that goal reminders appeared less often, as this allows for more consecutive, uninterrupted congruent trials, which should increase Stroop correlations with WMC (Meier & Kane, 2013). Third, we included more blocks of trials to allow additional time for goal neglect over the course of the experiment, as mind wandering tends to increase across the duration of experiments (Cunningham et al., 2000; Krimsky et al., 2017; McVay & Kane, 2012; Smallwood et al., 2004; Thomson et al., 2014). Fourth, we held item PC at 75% within every 12 trials, rather than simply across the overall experiment, as this should reduce noise created by varying levels of PC following each reminder. Specifically, in re-examining the Hood and Hutchison data, although their overall PC was 75%, the actual PC for the 12 trials following each reminder varied considerably ($SD = 11.7\%$), ranging between 33% and 100% across subjects and blocks. If we assume that each incongruent trial also serves as an external reminder (Hutchison, 2007; Kane & Engle, 2003), providing goal reminders should have less impact when followed by mostly incongruent trials than when followed by mostly congruent trials.

As with Hood and Hutchison (2021), we hypothesized that providing goal reminders should eliminate the relationship between WMC and Stroop effects. Further, with the within-subject manipulation of goal condition, the longer duration between reminders, the longer duration of the experiment, and the more precise manipulation of PC, we expect a significant Reminder Condition \times WMC \times Congruency interaction, with the relationship between WMC and Stroop effects (particularly errors) still intact in the non-goal statement condition, but greatly eliminated in the goal reminder condition. Finally, if the effects of goal reminders are short-lived, this interaction should be significant in the first 12 trials following the reminder, but not in the second 12 trials following the reminder.

EXPERIMENT 1

Methods

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. This study was not pre-registered. Based on data from Hutchison (2011) and Hood and Hutchison (2021) examining the correlation between AOSPAN performance and MC Stroop list errors and reaction times, we predicted a medium-sized correlation ($r = .30$). Using G-power, the *a priori* power analysis indicated a sample size of 93 to have 85% power for detecting this correlation when employing the traditional .05 criterion of statistical significance. Although we aimed for at least 93 participants, 105 undergraduates from Montana State University ($M_{\text{age}} = 19.81$ $SD_{\text{age}} = 3.12$, 41 male, 61 female, 1 other, 2 no response) had participated in the study for partial course credit by the end of the semester. Because one participant was missing scores from the AOSPAN task, the final analysis is based on 104 participants.

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. The E-prime experimental lists for both studies are uploaded to OSF and available at <https://osf.io/3ej8b>.

Automated Operation Span (AOSPAN; Unsworth et al., 2005). Participants first completed the Automated Operation Span (AOSPAN; Unsworth, et al., 2005). During this task, participants were asked to make “true” or “false” decisions via a mouse click to simple math problems (e.g., $4/2 + 3 = 6?$). After each math decision, 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 shown letters in the correct order that they were presented. The task was composed of three blocks, with each block 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 and colleagues (Conway et al., 2005).¹

Stroop (Stroop, 1935). Participants completed a 75% congruent version of the Stroop task similar to the goal reminder and non-goal statement conditions in Hood and Hutchison (2021). However, as discussed above, there were four important modifications from the Hood and Hutchison procedure. First, reminders were given every 24 trials, rather than every 12 trials. Second, the goal reminder and non-goal statement conditions were manipulated within-subjects,

¹ As in Hood and Hutchison (2021), 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 Richmond et al., 2021; Unsworth et al., 2009). **Nonetheless, we repeated all our analyses when using the 85% criterion and included these in the supplemental material. Of note, the critical three-way Reminder Condition × WMC × Congruency interaction in Experiment 2 remained significant when an 85% criterion was used.**

rather than between, and were randomly presented twice each within each block of 96 trials to maintain a similar ratio of goal and non-goal conditions throughout the duration of the experiment while still allowing condition to be randomized within each block. This also meant that neither the goal nor non-goal conditions could occur more than four times in a row (e.g., last two conditions of one block and first two of the next). Third, 75% congruency was maintained within every 12 trials, rather than simply experiment-wide. To accomplish this while keeping the congruency at 75%, it was necessary to switch from four words/colors to three so that, within each 12 trials, each word could be presented three times in the congruent color and once in an incongruent color. Fourth, we increased the number of total trials from 216 to 384 trials.

Stimuli consisted of one of three words (RED, GREEN, BLUE) presented upon a black background. Within every 12 trials, each word was presented 3 times in its congruent color (e.g., the word "GREEN" in green font) and one time in one of the two incongruent colors (e.g., the word "GREEN" in blue font). The choice for which incongruent color was randomized for the first 12 trials in each block of 24 and the alternate color was chosen for the second 12 trials in that block. Within every sub-block of 12 trials, stimuli were presented randomly and shown in the center of the screen in 18-point Courier New font for 3,000 ms or until a response.

Participants were instructed prior to the practice trials to name the color of the written word while ignoring the word itself.

Instructions described the two different goal statements and emphasized both accuracy and speed when responding to Stroop stimuli. Participants responded to the goal and non-goal prompts and Stroop stimuli by speaking into a microphone. Following each color naming response, the 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 inter-trial 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 followed by the Stroop task. At the beginning of the Stroop task, all participants were given instructions stating that “the goal of this task is to name the color (not the word itself)” and to “please respond both quickly and accurately.” During the Stroop task, participants stopped every 24 trials to vocalize a rehearsed statement. The rehearsed statements appeared on the computer screen and the participants were instructed to read them aloud. For example, in goal reminder trials, participants saw the phrase “*the goal is to name the -----, not the ----*” and had to vocally respond by saying “*the goal is to name the color, not the word.*” On non-goal statement trials, participants saw the phrase “*this is part of my ----- to ----- class*” and had to respond by saying “*this is part of my intro to psychology class.*” No participants failed to accurately verbalize the goal or non-goal statements. Overall, the task contained a total of 384 trials and was preceded with 12 practice trials with the same stimuli and proportion congruency.

Results

Stroop Errors

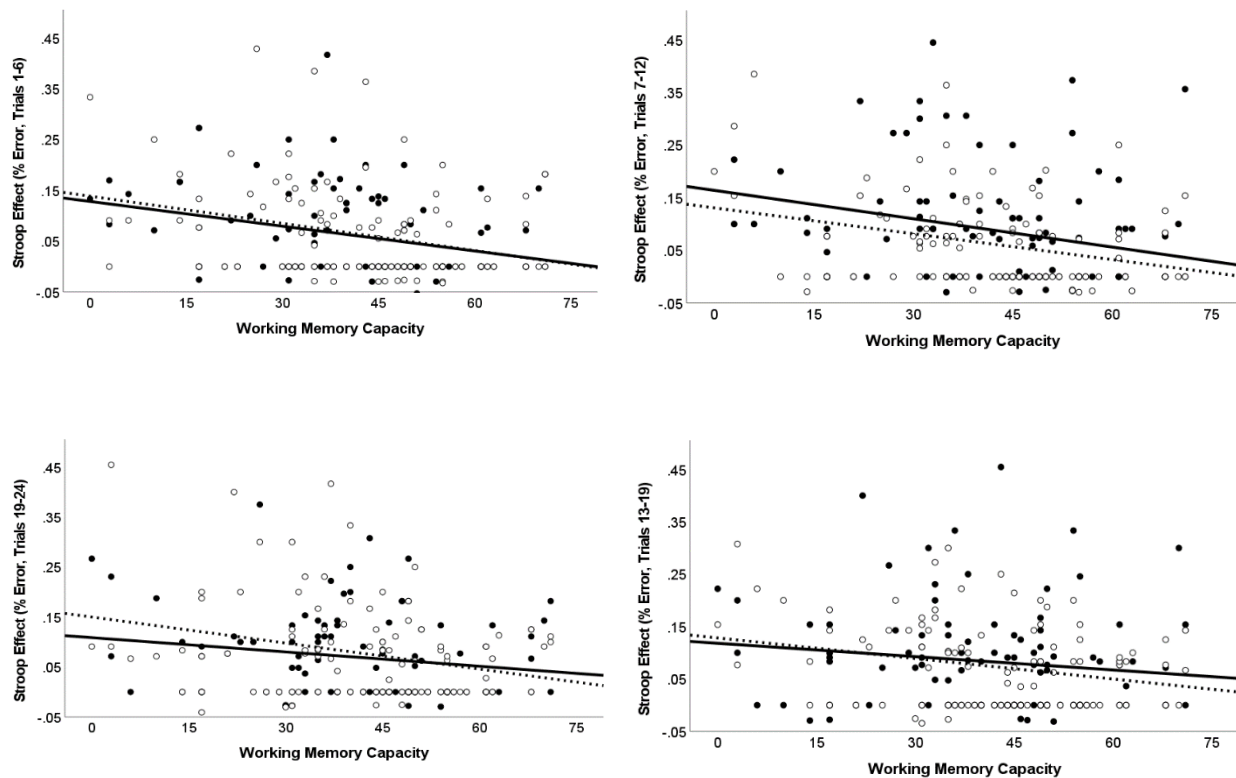
We conducted linear mixed model analyses on the pre-aggregated data using the lme4 package (Bates et al., 2015) in R (Version 4.2.1; R Core Team, 2021). These data, R code, and output are available at <https://osf.io/3ej8b>. Stroop effects for trials 1-6 (top left), 7-12 (top right), 13-18 (bottom left) and 19-24 (bottom right) following a reminder are shown in Figure 2 as a function of Reminder Condition and WMC. In each analysis, Congruency (incongruent vs.

congruent), Trial (1-24 following goal/non-goal statements), and Reminder Condition are within-subject variables and WMC (AOSPAN) is a between-subject variable. Trial number and WMC (AOSPAN) were allowed to remain continuous. We also included subject intercepts, but did not include subject slopes because this is the variance we were trying to predict with WMC. Finally, we used the *parameters* package in R to refit the model by first standardizing all measures.

Therefore, all our reported beta weights are standardized.

Figure 2.

Stroop Errors for trials 1-6 (top left), trials 7-12 (top right), trials 13-18 (bottom left) and trials 18-24 (bottom right) following a goal reminder (solid line, filled circle) or non-goal statement (dotted line, open circle) as a Function of WMC in Experiment 1.



There was a significant main effect of Congruency ($b = .22$, $SE = 0.007$, $t = 31.43$, $p < .001$) and WMC ($b = .04$, $SE = 0.01$, $t = 3.24$, $p = .001$), demonstrating more errors for incongruent trials than congruent trials (i.e., a Stroop effect) and more errors among those lower

in WMC. There was also a WMC \times Congruency interaction ($b = .06$, $SE = 0.007$, $t = 8.23$, $p < .001$), indicating increasing Stroop effects among those with lower WMC. In addition, there was a WMC \times Congruency \times Trial interaction ($b = .02$, $SE = 0.007$, $t = 2.45$, $p = .014$). Follow up analyses examining each set of six trials separately revealed that the relation between WMC and Stroop effects, although significant in all cases, was greater within trials 1-6 ($b = .08$, $SE = 0.01$, $t = 5.51$, $p < .001$) and trials 7-12 ($b = .07$, $SE = 0.01$, $t = 5.00$, $p < .001$) following a goal/non-goal statement than within trials 13-18 ($b = .04$, $SE = 0.01$, $t = 3.00$, $p = .003$) and trials 19-24 ($b = .04$, $SE = 0.01$, $t = 3.14$, $p = .002$) following a goal/non-goal statement. Finally, the predicted 3-way Reminder Condition \times WMC \times Congruency interaction obtained by Hood and Hutchison (2021) did not reach significance ($b = .02$, $SE = 0.01$, $t = 1.85$, $p = .064$). To determine the strength of evidence for this null effect, a Bayes Factor was computed comparing the linear mixed effects model with the Reminder Condition \times WMC \times Congruency three-way interaction to a model with no three-way interaction using the *BayesFactor* package (Morey et al., 2015) in R. The resulting Bayes Factor (BF_{10}) shows the ratio of how much improvement the model with the component predicts the data over the null model missing that component. According to the classification scheme from Lee and Wagenmakers (2013; adjusted from Jeffreys 1961), a BF_{10} of 1 = no evidence, 1-3 = anecdotal (weak), 3-10 = moderate, 10-30 = strong, 30-100 = very strong, and >100 = extreme evidence. [Note that values < 1 equal evidence for the null, such that .33, .10, and .033 equal moderate, strong, and very strong evidence for the null hypothesis, respectively.] Our $BF_{10} = .157$, indicating strong evidence for the null hypothesis. No other effects were significant.

Although our power analysis was based on the ability to detect a Pearson r correlation of .30, our critical prediction was that the typical WMC \times Congruency correlation of around .30

would be greatly reduced or eliminated following the goal reminders only, producing a 3-way WMC \times Reminder Condition \times Congruency interaction. Because of this, we next ran a sensitivity analysis to determine the actual size of the difference of correlation that we were able to detect. Our sensitivity analysis compared two dependent Pearson r 's with a common index for the correlation of Stroop interference for both the goal and the non-goal condition ($r = .648, p < .001$). At 80% power, we were able to detect a difference in correlations of at least .23 (e.g., hypothetical $r = -.30$ for nongoal and $r = -.07$ for goal condition). At 90% power, we were able to detect a difference in correlation of at least .26 and at 95% power we were able to detect a difference of .29. Thus, we had over 95% power to detect the predicted pattern of approximately $r = -.30$ following non-goal statements and $r = .00$ following goal reminders. However, rather than this predicted effect, we obtained statistically equal correlations following non-goal statements ($r = -.42$) and goal reminders ($r = -.30$).²

Finally, because of the possibility of carry-over effects reducing the effect of reminders, we conducted an exploratory between-subject analysis testing the critical 3-way Reminder Condition \times WMC \times Congruency interaction within the first 24 trials, which allowed us to include all participants and treat Reminder Condition as a between-subject variable. In this analysis, although in the right direction, with numerically larger correlation between WMC and Stroop effects among those in the non-goal statement ($r = -.354, p = .010$) than the goal statement ($r = .313, p = .024$), the critical 3-way interaction still did not reach significance ($b = .06, SE = 0.04, t = 1.65, p = .098$).

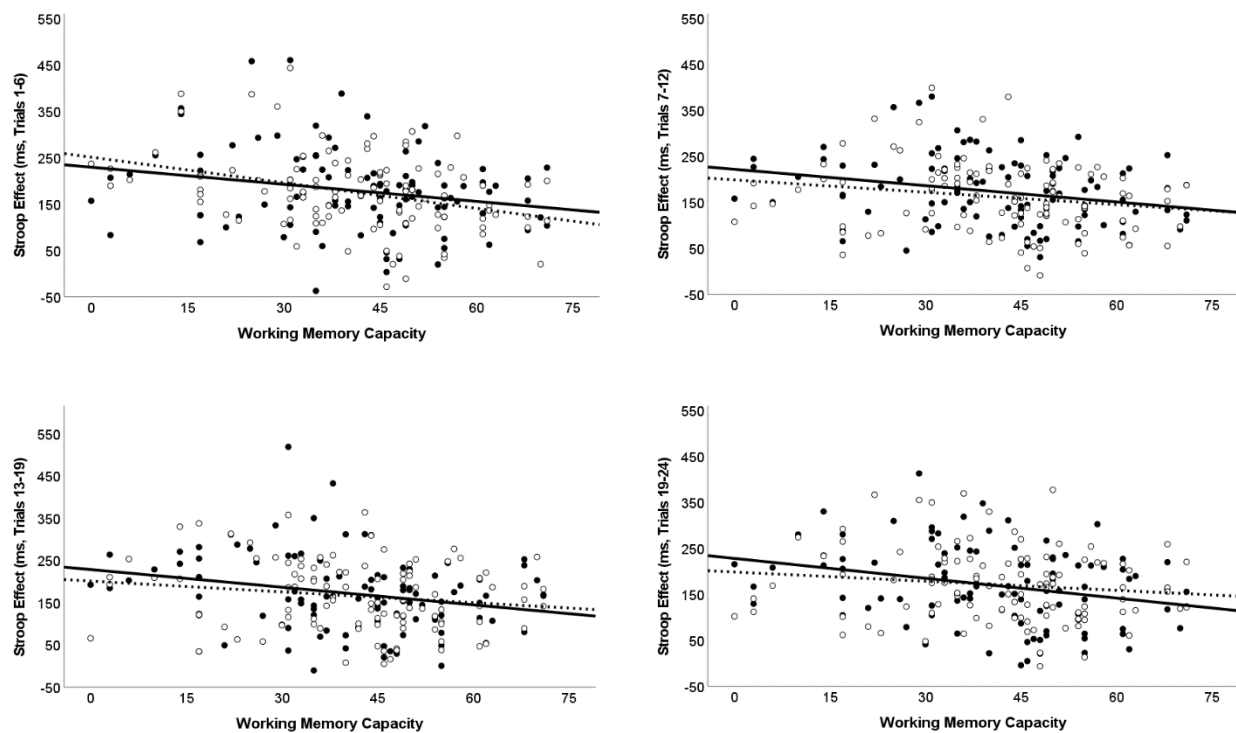
Reaction Times (RTs)

² We thank Dr. Michael Kane for suggesting this sensitivity analysis.

We used SPSS for initial data processing prior to reading this data into R. The RT analysis was conducted on only responses greater than 50 ms, which removed 0.9% of RTs. We then selected only accurate responses, which removed another 2.4% of the trials. Next, we used Van Selst and Jolicoeur's (1994) nonrecursive method to remove outliers, which resulted in the removal of an additional 2.3% of RTs. To fit the assumptions of LME, we log-transformed all RTs prior to running the analyses.³ The Stroop effects based on trimmed arithmetic RTs for trials 1-6 (top left), 7-12 (top right), 13-18 (bottom left) and 19-24 (bottom right) following a reminder (solid line, filled circle) or non-goal statement (dotted line, open circle) as a function of WMC are shown in Figure 3 as a function of Reminder Condition and WMC.

Figure 3.

Stroop RT effects for trials 1-6 (top left), trials 7-12 (top right), trials 13-18 (bottom left) and trials 18-24 (bottom right) following a goal reminder (solid line, filled circle) or non-goal statement (dotted line, open circle) as a Function of WMC in Experiment 1.



³ In addition to the LME analysis, we also conducted GLMM models (Lo & Andrews, 2015) on the untransformed data. However, several of these models failed to converge.

There was a significant main effect of Congruency ($b = .42$, $SE = 0.005$, $t = 76.78$, $p < .001$) and Reminder Condition ($b = .02$, $SE = 0.008$, $t = 2.10$, $p = .036$), demonstrating longer RTs for incongruent trials than congruent trials (i.e., a Stroop effect) and longer RTs following non-goal statements. There was also a WMC \times Congruency interaction ($b = .05$, $SE = 0.006$, $t = 8.64$, $p < .001$), with increased Stroop effects among those lower in WMC. The critical three-way Reminder Condition \times WMC \times Congruency interaction was not significant ($b = .01$, $SE = 0.008$, $t = 1.40$, $p = .161$), with moderate evidence for the null hypothesis ($BF_{10} = .036$). This demonstrates no difference in the correlations between WMC and Stroop effects among those in the non-goal statement ($r = -.24$) and goal reminder ($r = -.30$) conditions. No other effects were significant.

A post-hoc sensitivity analysis for Stroop interference in reaction times, in which the goal and non-goal conditions were highly correlated ($r = .871$, $p < .001$), indicated that our study could detect relatively small differences in the size of WMC and Stroop effect correlations across conditions. Specifically, we had 80%, 90%, and 95% power to detect difference in the size of the correlations across non-goal statement and reminder conditions of at least .14, .16, and .18, respectively.

Discussion

In Experiment 1, we tested the effect of providing periodic goal reminders during a mostly congruent Stroop task using a within-subjects design, as this should provide increased power to detect different goal conditions. The expected results based on Hood and Hutchison (2021) were that WMC differences in Stroop performance should disappear in the goal condition but remain in the non-goal condition. However, although marginal and in the predicted direction, the 3-way WMC \times Congruency \times Reminder Condition did not reach significance. When

attempting to artificially create a between-subjects design by examining only the first 24 trials, again, the WMC \times Reminder interaction did not reach significance.

Given this failure to replicate Hood and Hutchison (2021), we believe this pattern suggests that either (1) there is in fact no difference between goal reminder and non-goal statement conditions, or (2) there is a carry-over effect of goal reminders. In terms of the first possibility, we replicated the null difference between goal and non-goal conditions found in Hood and Hutchison. However, the pattern is different. In that study, neither condition showed a significant correlation between WMC and Stroop effects ($r = .066, p = .59$ and $r = -.170, p = .166$ for the goal and non-goal conditions, respectively), whereas in the current study this correlation was robust for both conditions ($r = -.303, p = .002$ and $r = -.416, p < .001$, respectively). This suggests Hood and Hutchison's lack of correlation between WMC and Stroop effects in the previous goal and non-goal conditions may have been due simply to giving participants breaks every 12 trials, rather than every 24 trials in the current study. Alternatively, the current failure to find a significant effect of reminder condition in the current study could be due to a carry-over effect. However, such a carry-over effect would have to be asymmetric, with non-goal statements eliminating the effectiveness of goal reminders, rather than goal reminders carrying-over to prevent mind wandering even following non-goal statements. Given this asymmetry, it's possible that intermixing goal reminders and non-goal statements caused participants to disregard both preventing the task goal from ever becoming established. The current null effect of Reminder Conditions in Experiment 1 data are consistent with either of these explanations.

EXPERIMENT 2

The goal of Experiment 2 was to determine whether carryover effects may be diminishing the effects of the goal reminder condition or if instead the effect of goal reminders eliminating the relation between WMC and Stroop effects obtained by Hood and Hutchison (2021) was spurious. To test this, Experiment 2 modified Experiment 1 to be between-subjects, rather than within-subjects, by having participants state either the goal reminder or non-goal statement prior to each block of 24 trials.

Method

Participants and Design

As per Experiment 1, we wanted at least 93 participants in each goal condition to achieve 85% power to detect a correlation of .30; therefore, we estimated that we would need at least 186 participants for Experiment 2. A total of 203 undergraduates from Montana State University participated for partial course credit. Data from three participants were not included in the analysis due to either missing Stroop data (two participants), missing AOSPAN data (one participant), or for the participant not following instructions and instead naming the word on every trial (one participant). This left 100 and 99 participants in the non-goal and goal conditions, respectively, final $N = 199$, $M_{age} = 19.62$, $SD_{age} = 2.91$, 73 males, 125 females, 1 other).

Measures, Apparatus, and Procedures

This study was approved by the Montana State University Institutional Review Board. The same measures, apparatus, stimuli, and procedures were used for Experiment 2 as in Experiment 1, except that the participants were randomly assigned to vocalize either the goal reminder or non-goal statements during the Stroop task. Participants first completed the ASOPAN followed by the Stroop task based on condition. No participants failed to accurately

verbalize the goal/non-goal statements. As with Experiment 1, the task contained 12 practice trials and 384 experimental trials.

Results

WMC

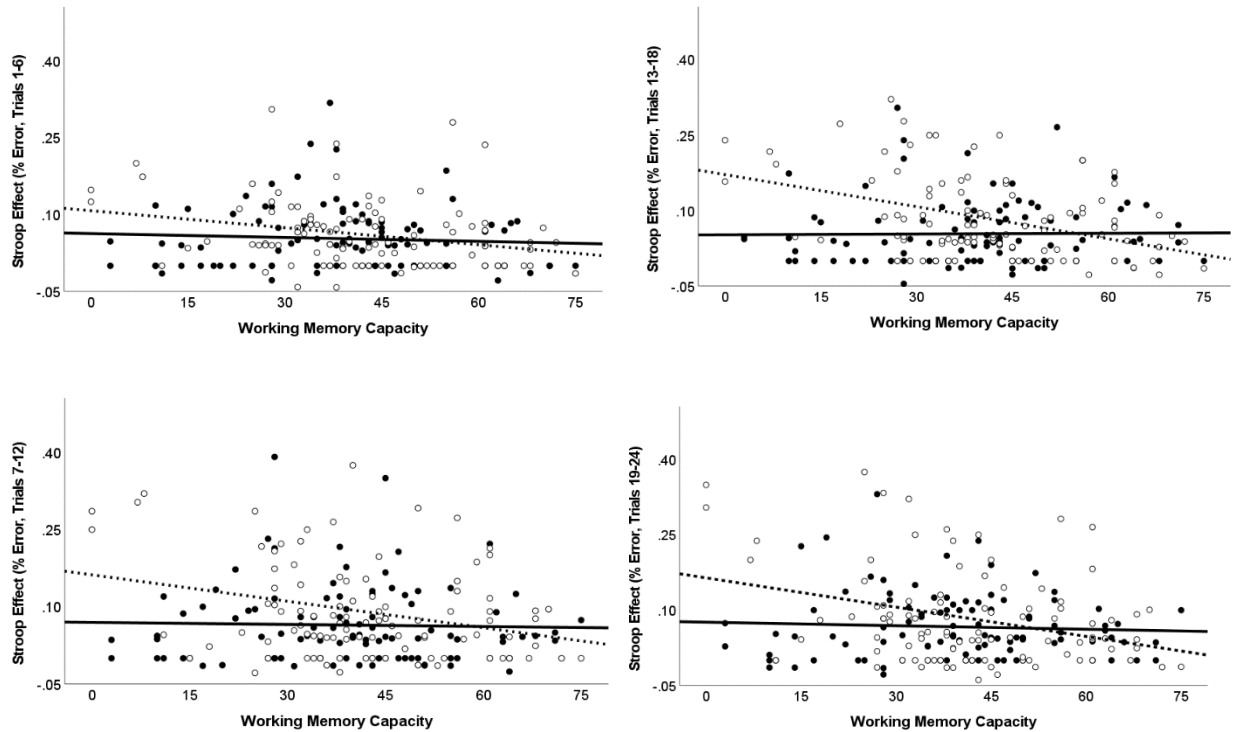
To ensure that there were no preexisting differences in WMC between our two randomly assigned Reminder Conditions, we conducted an independent-groups t -test. The effect of Reminder was nonsignificant [$t(1, 197) = .882, p = .379$], indicating no significant difference in WMC between the goal ($M = 39.54, SE = 1.62$) and the non-goal conditions ($M = 41.52, SE = 1.56$).

Stroop Errors

Stroop effects for trials 1-6 (top left), 7-12 (top right), 13-18 (bottom left) and 19-24 (bottom right) following a reminder are shown in Figure 4 as a function of Reminder Condition and WMC. As with Experiment 1, we conducted linear mixed model analyses using the lme4 package (Bates et al., 2015) in R (Version 4.0.5; R Core Team, 2021). These data, R scripts, and output are available at <https://osf.io/3ej8b>. In each analysis, Congruency (incongruent vs. congruent) and Trial (1-24 following goal/non-goal statement) are within-subject conditions, Reminder Condition is a between-subject condition, and standardized WMC (AOSPAN) is a continuous between-subject variable. All reported beta weights are again standardized. As in Experiment 1, we included subject intercepts, but not subject slopes, because this is the variance that we are trying to predict with WMC.

Figure 4.

Stroop Errors for trials 1-6 (top left), trials 7-12 (top right), trials 13-18 (bottom left) and trials 18-24 (bottom right) following a goal reminder (solid line, filled circle) or non-goal statement (dotted line, open circle) as a Function of WMC in Experiment 2.



There were significant main effects of Congruency ($b = .17$, $SE = .01$, $t = 33.82$, $p < .001$), Reminder Condition ($b = .04$, $SE = .02$, $t = 2.46$, $p = .014$), and Trial ($b = .01$, $SE = .00$, $t = 2.57$, $p = .010$), demonstrating more errors for incongruent trials, more errors for participants receiving the non-goal statement, and more errors on later trials, respectively. A 2-way Reminder Condition \times Congruency interaction was qualified by the predicted 3-way Reminder Condition \times WMC \times Congruency ($b = .07$, $SE = .01$, $t = 10.39$, $p < .001$). Our Bayesian analysis revealed extreme evidence for this critical three-way interaction ($BF_{10} = 2.48 \times 10^{21}$). As shown in Figure 4, the benefit of goal reminders in reducing Stroop errors was more pronounced as WMC decreased. In addition, three other significant interactions (Trial \times Congruency, Reminder Condition \times WMC, and Reminder Condition \times WMC \times Trial) were qualified by a 4-way Reminder Condition \times WMC \times Congruency \times Trial interaction ($b = .01$, $SE = .0012$, $t = 2.13$, $p = .033$). In order to decompose this 4-way interaction, we analyzed the 3-way Reminder Condition \times WMC \times Congruency interactions separately for each set of 6 trials following a reminder.

Although the 3-way Reminder Condition \times WMC \times Congruency was significant within each set of 6 trials following a Reminder, it was smaller during the initial 6 trials ($b = .04$, $SE = .01$, $t = 3.11$, $p = .002$) than during trials 7-12 ($b = .07$, $SE = .01$, $t = 4.98$, $p < .001$), trials 13-18 ($b = .10$, $SE = .01$, $t = 7.25$, $p < .001$), and trials 19-24 ($b = .07$, $SE = .01$, $t = 5.08$, $p < .001$). This demonstrates that the effect of goal reminders is long-lasting. Specifically, as can be seen in Figure 4, Stroop errors among those lower in WMC increased across trials, but only among those receiving the non-goal statements. In contrast, among those receiving the goal reminders, there was no relation of WMC to Stroop errors across all 24 trials. This observation that lower WMC individuals' Stroop errors increased across trials only among those in the non-goal condition is confirmed by a significant WMC \times Trial \times Congruency interaction among those receiving the non-goal statements ($b = .01$, $SE = .00$, $t = 2.55$, $p = .011$), but not among those receiving the goal reminders ($b = .00$, $SE = .01$, $t = 0.33$, $p = .744$).

The 4-way interaction above demonstrates that the difference in WMC \times Stroop effect interactions between the goal reminder and non-goal statement conditions is larger during later trials than during the first 6 trials. Rather than quickly dissipating, the effect of goal reminders (vs. non-goal statements) grew over time, as lower WMC individuals receiving goal reminders remained vigilant in naming the color whereas lower WMC receiving non-goal statements began to neglect the goal over time.⁴

Our sensitivity analysis showed that at we had an ability to detect a difference between independent Pearson r correlations of .40 with 80% power, .47 with 90% power, and .52 with 95% power when conducting a two-tailed test. If instead we used a one-tailed test consistent with our directional hypothesis, we had the sensitivity to detect a difference in correlations of .36 at

⁴ This 4-way interaction remains significant ($b = .02$, $SE = .01$, $t = 2.31$, $p = .021$) if Half is included as a dichotomous variable (first 12 vs. second 12) rather than using the continuous variable of trial (1-24).

80% power, .42 at 90% power, and .47 at 95% power. Our actual difference in correlation between the goal ($r = -.045, p = .665$) and the non-goal ($r = -.401, p < .001$) was .36, indicating that we had 70% power for a two-tailed test and 80% power for a one-tailed test.

Time Course of Reminder \times WMC Interaction across the Experiment

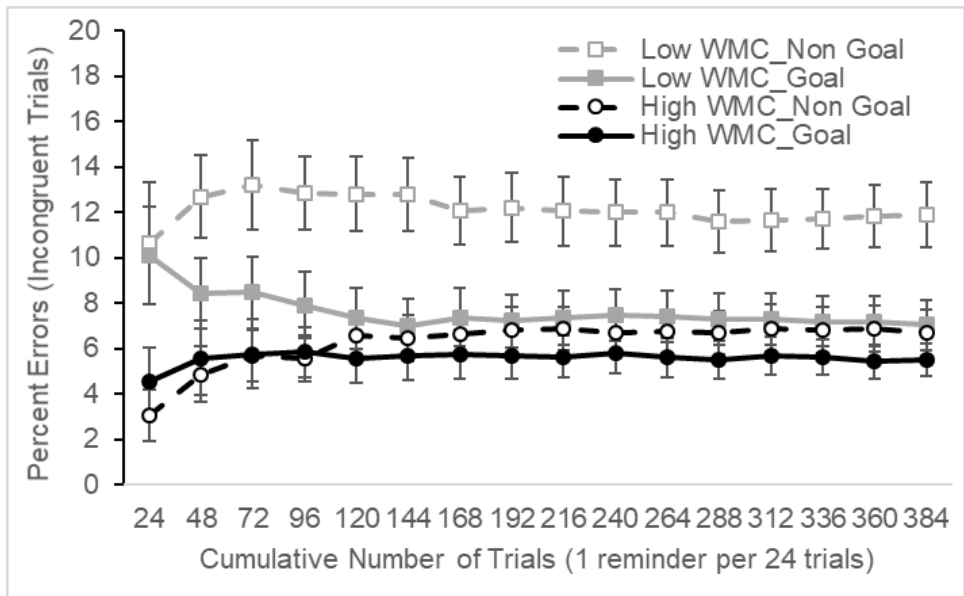
Given the significant 3-way Reminder Condition \times WMC \times Congruency interaction, we examined the time course of this effect across the duration of the experiment. Because reminders were given every 24 trials, we tested the critical 3-way interaction separately after each cumulative reminder (i.e., after 24 total trials, 48 total trials, 72 total trials, etc.). When examining just the first 24 trials only, this interaction was not significant ($b = .03, SE = 0.03, t = 1.22, p = .233$). However, the interaction was significant [$b = .07, SE = 0.02, t = 3.46, p < .001$] after just the second reminder (48 trials) and remained significant throughout the rest of the experiment [all subsequent p -values $< .001$]. The fact that the interaction was not significant until after the second reminder and then remained significant throughout suggests that multiple initial goal reminders are needed for those lower in WMC to establish and maintain the task goal.

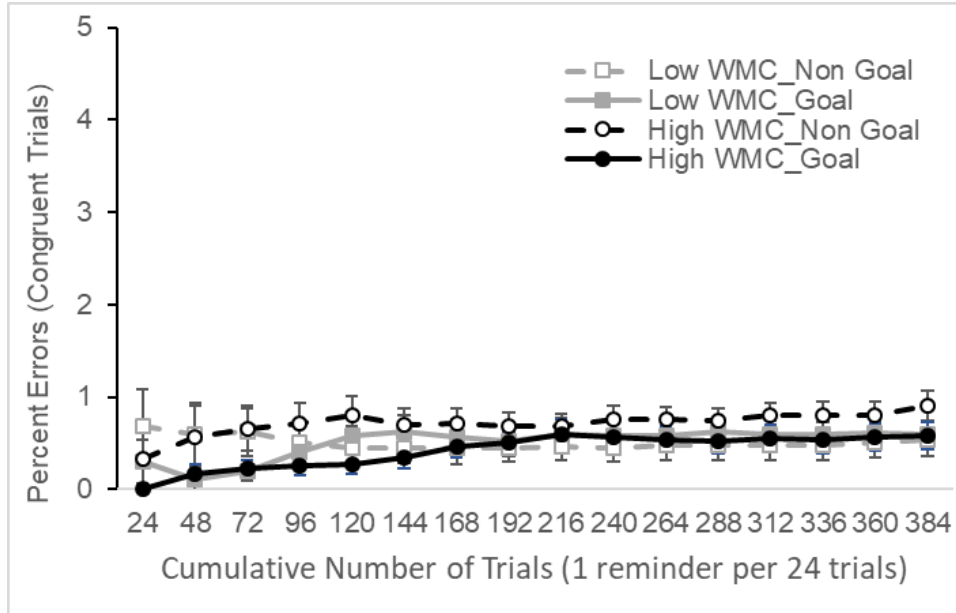
For illustrative purposes only, Figure 5 shows performance on incongruent and congruent trials separately for those in the upper tertile (black lines) and lower tertile (grey lines) of WMC and by Reminder Condition (solid lines = goal, dashed lines = non-goal). There are three important patterns to notice in Figure 5. First, note that Reminder Condition affected errors solely on incongruent trials, as there were virtually no errors on congruent trials. Second, consistent with the cumulative analyses above, the pattern starts to emerge after the second reminder and stabilized around the 3rd-4th reminder, with greater effects of Reminder Condition among lower WMC individuals. Third, the pattern shows constant, and large WMC differences throughout the experiment in the non-goal condition whereas, in the goal condition, lower WMC

individuals initially made more errors, but their performance then improved near the level of high WMC individuals after the 4th goal reminder.

Figure 5.

Stroop Errors for Low WMC (grey lines) and High WMC (black lines) Tertiles in the Goal Reminder (solid lines, filled marker) and Non-Goal Statement (dashed lines, transparent marker) Conditions on Incongruent (top) and Congruent (bottom) Trials in Experiment 2. Errors bars reflect standard error of the mean.



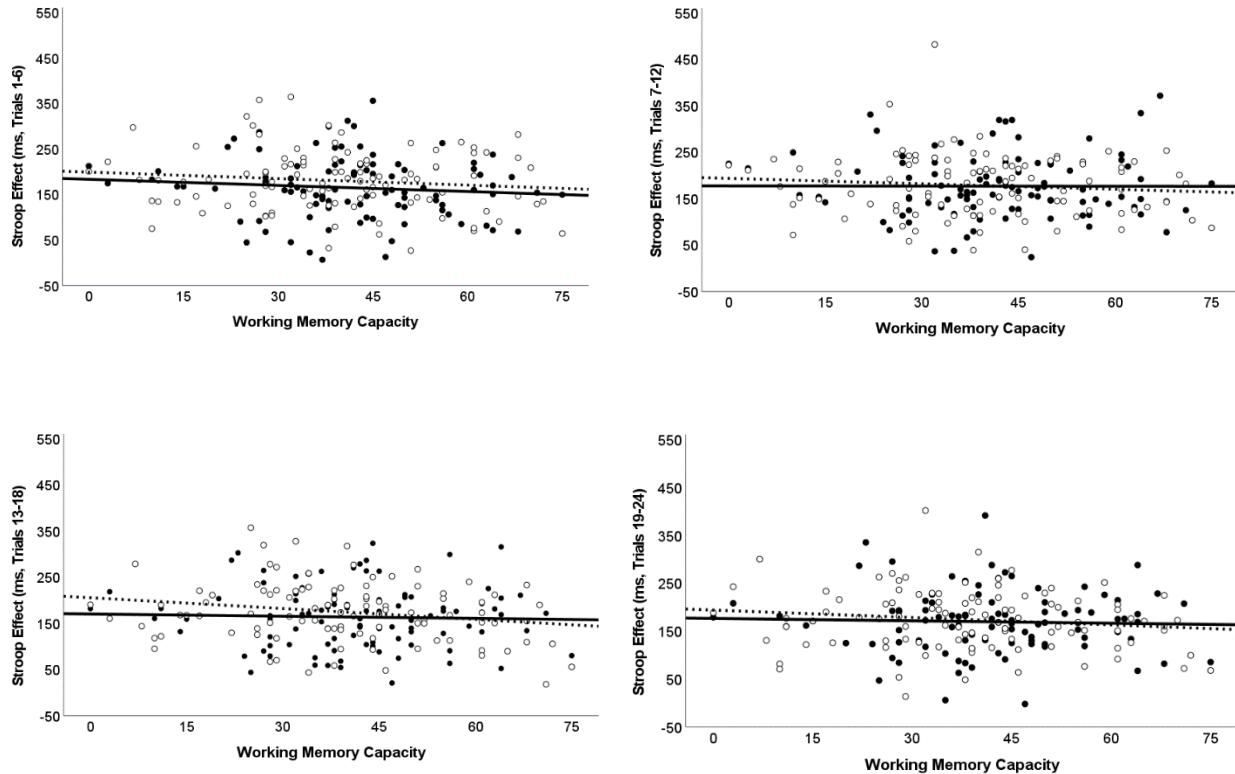


Reaction Times

The RT analysis was conducted on only accurate responses and RTs of 50 ms or more (which removed .02% of RTs). Van Selst and Jolicoeur's (1994) nonrecursive method was then used, which resulted in an additional .02% of RTs. As was done for Experiment 1, we then log-transformed all remaining RTs for the analyses. Stroop effects in the trimmed RTs for trials 1-6 (top left), 7-12 (top right), 13-18 (bottom left) and 19-24 (bottom right) following a reminder are shown in Figure 6 as a function of Reminder Condition and WMC.

Figure 6.

Stroop RT effects for trials 1-6 (top left), trials 7-12 (top right), trials 13-18 (bottom left) and trials 18-24 (bottom right) following a goal reminder (solid line, filled circle) or non-goal statement (dotted line, open circle) as a Function of WMC in Experiment 2.



There was a significant main effect of Congruency ($b = .42$, $SE = 0.004$, $t = 108.78$, $p < .001$), demonstrating longer standardized RTs for incongruent trials. There were also significant interactions between Reminder Condition and Congruency ($b = .03$, $SE = 0.001$, $t = 4.90$, $p < .001$) and between WMC and Congruency ($b = .01$, $SE = 0.004$, $t = 3.99$, $p < .001$), with reduced Stroop effects for those receiving the goal reminder and among those higher in WMC. As was the case for errors, the predicted 3-way Reminder Condition \times WMC \times Congruency interaction was significant ($b = .02$, $SE = 0.005$, $t = 2.77$, $p = .006$). However, despite being significant and in the predicted direction, our Bayesian analysis revealed moderate evidence for the null hypothesis ($BF_{10} = .09$). Therefore, we will treat this interaction as nonsignificant. Finally, there was a significant 3-way interaction of Reminder Condition \times Trial \times Congruency ($b = .000$, $SE = 0.000$, $t = 2.098$, $p = .036$). To decompose this 3-way interaction, we analyzed Reminder Condition \times Congruency interactions separately for each set of six trials following a reminder.

This 3-way interaction was caused by the fact that the 2-way Reminder Condition \times Congruency interaction was significant within trials 1-6 ($b = .05$, $SE = 0.01$, $t = 4.28$, $p < .001$) and trials 13-18 ($b = .03$, $SE = 0.01$, $t = 3.01$, $p = .003$) following a reminder, with larger Stroop effects in the non-goal statement condition, but was not significant within trials 7-12 ($b = .01$, $SE = 0.01$, $t = 0.93$, $p = .355$) or trials 19-24 ($b = .01$, $SE = 0.01$, $t = 1.23$, $p = .218$). No other effects were significant.

As with our sensitivity analysis above, we were able to detect a correlation difference of .40, .47, and .53 with 80%, 90%, and 95% power, respectively, when conducting a two-tailed test and a difference of .36, .42, and .47 at 80%, 90%, and 95% power, respectively, when conducting a one-tailed test. Our actual difference in correlation between the goal reminder ($r = -.162$, $p = .109$) and the non-goal statement ($r = -.234$, $p = .019$) was only .072.

Discussion

The purpose of Experiment 2 was to test whether the significant WMC \times Stroop interaction in Hood and Hutchison (2021) was spurious or whether instead the null WMC \times Stroop interaction in Experiment 1 was due to carryover effects. Results demonstrated that goal reminders indeed eliminated the relationship between WMC and Stroop errors. This not only replicates Hood and Hutchison's main finding of goal reminders eliminating WMC effects on Stroop, but the cleaner manipulation of PC within each 12 trials, the addition of more trials per reminder, and longer study duration adds even stronger evidence for the goal maintenance explanation for WMC-related differences in MC Stroop lists. Specifically, we now found a significant elimination of WMC differences in Stroop errors only in the goal reminder condition ($r = -.045$, $p = .655$), whereas those receiving the non-goal statements continued to show the typical WMC \times Stroop interaction ($r = -.401$, $p < .001$). We also found that goal reminders

reduced Stroop RT effects as well. However, although the three-way interaction with WMC was also significant, the Bayesian analysis gave strong evidence for the null. This suggests the reminder was not selectively more helpful to those lower in WMC in reducing Stroop RTs like it was for reducing Stroop errors. This is likely because errors in MC lists are a more sensitive measure of goal neglect than RT effects (Hood & Hutchison, 2021; Kane & Engle, 2003).

We were also interested in examining whether any potential effect of goal reminders is short-lived and whether the effectiveness of reminders accumulates across the duration of the experiment. First, the effect of the goal reminders was actually stronger, relative to the non-goal statements, in trials 13-24 following a reminder than it was in trials 1-12, demonstrating that the effect of goal reminders is long-lasting. This is because lower WMC individuals in the non-goal statement condition increased errors in later trials, whereas errors remained flat across all 24 trials for individuals receiving goal reminders. Second, although lower WMC individuals begin the experiment making more errors than higher WMC individuals, those in the goal reminder condition improve to the level of higher WMC individuals after the 4th goal reminder. This demonstrates that the benefits of goal reminders accumulate across the duration of the experiment. Such accumulation supports the asymmetric carry-over effect explanation for the null effect of reminders in Experiment 1, such that intermixing non-goal statements disrupts the establishment and subsequent benefit of goal reminders, perhaps by causing participants to ignore both statements.

Regarding reaction times, Stroop effects were reduced for participants receiving goal reminders. Further, as was found for errors, the typical $WMC \times Congruency$ interaction was nonsignificant in the goal reminder group but remained robust for the non-goal group. However, evidence for the 3-way interaction favored the null hypothesis. Thus, although goal reminders

reduce Stroop effects relative to the non-goal condition, there is no evidence within the RT data that they do so more strongly for those lower in WMC. Lastly, goal reminders were more strongly effective for trials 1-6 and trials 13-18 following reminders, again demonstrating that effects of goal reminders are long-lasting.

General Discussion

There are individual differences in the ability to regulate behavior in a goal-directed manner to overcome automatic, habitual responses. Goal reminders are one effective way to increase performance for those who struggle with maintaining such explicit, proactive control. Evidence for these benefits stems from work with children (e.g., Gonthier et al., 2021) as well as recent work from our lab with young adults (Hood & Hutchison, 2021). Given the effectiveness of goal reminders on task performance, the purpose of the current two experiments was to provide a more powerful experimental design compared to the Hood and Hutchison study. Specifically, the current experiments provided a cleaner test of the benefits of goal reminders by including more trials within blocks, including more blocks of trials, and holding congruency at 75% within each set of 12 trials. Further, these experiments allowed for an examination of whether goal reminders are short-lived and whether their benefits accumulate over the course of the experiment.

Following Hood and Hutchison (2021), we predicted that the relationship between WMC and Stroop errors would be greatly reduced or eliminated in trials containing goal reminders, as the explicit reminder of the task goal should provide external support for those who cannot maintain such preparatory control across the course of an experiment. In Experiment 1, the goal-reminder and non-goal statements were manipulated within-subjects. Although the results showed the typical WMC \times Stroop interaction such that those lower in WMC produced more

Stroop errors, the critical three-way interaction between Reminder Condition, WMC, and Stroop effects did not reach significance and a Bayesian analysis suggested strong evidence for the null hypothesis.

To further investigate this null interaction, we examined the first 24 trials only, as this allowed us to create a between-subjects variable. Once again, the relationship between WMC and Stroop errors was not statistically different among participants who received a goal reminder and participants who instead received a non-goal statement.

Given the null effect of goal reminders in eliminating the $WMC \times \text{Stroop}$ interaction in Experiment 1, we deemed a replication of Hood and Hutchison (2021) was needed to determine if their findings were spurious or if the results from Experiment 1 were instead due to carryover effects in which intermixing non-goal statements reduces the effectiveness of goal reminders. Therefore, in Experiment 2, we used a between-subjects design. Importantly, the Experiment 2 results showed that goal reminders indeed eliminated the relationship between WMC and Stroop effects in both errors and RTs, whereas non-goal statements did not. Examining Figures 2 and 4, the slope of WMC on errors in the Non-Goal Condition was remarkably similar across Experiments 1 and 2, whereas the slope in the Goal Reminder Condition fattened across the two experiments. This pattern demonstrates that the within-subject mix of both goal-reminders and non-goal statements in Experiment 1 diminished the effectiveness of the goal-reminder cues. These results not only provide stronger evidence for the effectiveness of goal reminders than what was found in Hood and Hutchison but provide important insights into the implications of using goal reminders as a manipulation. Specifically, if goal reminders are to be used as an intervention to help improve task performance, they should be used in isolation, as simultaneously including non-goal statements appears to dilute their effectiveness.

Goal Reminders are Long-Lasting and Accumulate in Strength

In Hood and Hutchison (2021), the main effect of WMC on Stroop effects was significant in the second 6 trials, but not the first 6 trials following the reminder, whereas the WMC \times Reminder Condition interaction on Stroop effects (the current three-way Reminder Condition \times WMC \times Congruency interaction) was significant in the first 6, but not the second 6. This suggested that, although goal reminders may allow lower WMC individuals to initially overcome goal neglect, they may not have a sustained effect on performance. However, this interpretation remained speculative, given their null three-way WMC \times Reminder Condition \times Half interaction. In the current study, the effect of goal reminders was actually stronger later trials for the error analysis and equally strong across the first and second 12 trials for the RT analysis. This demonstrates that goal reminders are long-lasting and have a sustained effect on preventing the increased goal neglect experienced by lower WMC individuals over time.

Additionally, time course analyses showed that the interaction between WMC and goal reminders on Stroop effects was not significant until after the second reminder (48 trials), but then persisted until the end of the task. This is illustrated in Figure 5 in that incongruent errors among lower WMC individuals in the goal-reminder condition continually decrease across each successive reminder until about trial 144 (5th reminder) and then remain equally low throughout the rest of the task. Thus, the beneficial effects of goal reminders for lower WMC individuals accumulate across the duration of the experiment. Given that mind wandering tends to increase across the duration of experiments (Cunningham et al., 2000; Krimsky et al., 2017; McVay & Kane, 2012; Smallwood et al., 2004; Thomson et al., 2014), this finding is important, as it demonstrates that goal reminders can help lower WMC individuals (who are more prone to experiencing mind wandering during challenging tasks; Kane et al., 2007) remain focused

throughout the course of the experiment. Figure 5 further demonstrates that presenting goal reminders to lower WMC individuals equated their performance to that of higher WMC individuals under normal (non-goal reminder) conditions. Thus, presenting goal reminders effectively turned lower WMC individuals into higher WMC individuals by reducing the reliance of task performance on goal maintenance abilities.

On the other hand, there was no effect of goal reminders in Experiment 1. In fact, the goal reminder condition in Experiment 1 looked similar to the non-goal statement conditions in both experiments, suggesting that the non-goal statements disrupted goal maintenance, potentially causing participants to ignore both goal and non-goal statements. Thus, although the effectiveness of goals appears to be long lasting and cumulative, this can only occur if the goal is firmly established through multiple goal reminders without interruption from non-goal statements. It would be interesting for future studies to test if 4-5 consecutive initial goal reminders could sufficiently establish the goal to withstand future interruptions from non-goal statements later in the experiment.

Stroop Effects and Examinations of WMC versus Cognitive Load on Performance

Recently, it has been argued that attention control, in terms of inhibitory processing, might not exist as a psychometric construct, as it does not correlate with measures of WMC and general fluid intelligence once memory demands are controlled (Karr et al., 2018; Rey-Mermet et al., 2018; 2019). Our results in the Stroop task are consistent with arguments against “inhibition” as an important individual difference. Indeed, once the demand of keeping the task goal in mind is externally provided with reminders, Stroop performance in our studies no longer correlates with WMC.

However, contrary to the strong claims above, we do not suggest that this indicates the absence of executive or attention control as a reliable individual differences construct. Instead, we agree with past descriptions of attentional control as the coordination of attention and memory processes in service of a goal (Balota & Faust, 2001; Hutchison et al., 2010; Miller & Cohen, 2001). Both attention and memory processes are critical for implementing selective and sustained attention over time to enhance task relevant stimuli and suppress task-irrelevant stimuli. It is therefore no more surprising that memory plays an important role in selective attention tasks than it is for attention to play an important role in memory tasks. In more recent models, Engle and colleagues (Draheim et al., 2020; Shipstead et al., 2016; Tsukahara et al., 2020) have demonstrated that the ability to focus attention, as measured through attention control tasks, is critical for both maintaining information in memory (as captured in WMC tasks) and disengaging from no-long-useful information (as captured in fluid intelligence tasks).

A second issue concerns recent mixed results on the impact of imposed cognitive loads on Stroop performance. Specifically, if Stroop performance indeed requires working memory, then not only should Stroop effects be increased for those low in trait-level WMC, but they should also be increased for participants placed under a cognitive load. However, results from these studies are mixed. In terms of positive evidence, Kalanthroff et al. (2015) had participants perform the Stroop task while also performing an *n*-back task. They manipulated the number of letters back participants needed to recall during the *n*-back portion of the experiment, creating high and low working memory load conditions. Results revealed increased Stroop interference with high working memory load. In addition, Entel and Tzelgov (2019) demonstrated that including incongruent trials during practice can reduce later Stroop effects in a MC Stroop list,

consistent with incongruent trials serving as goal reminders. However, this benefit was eliminated when participants were put under a high working memory load.

In contrast to such positive evidence that imposing loads can reduce Stroop performance and/or benefits of goal reminders, Moss et al. (2020) found little-to-no impact of cognitive load on Stroop performance. In their experiments, they included Stroop trials within a change detection task involving a low (1-2 items) or high (5 items) cognitive load. Specifically, on each trial, participants received a low or high WM load for the change detection task and were also presented with a Stroop stimulus during the maintenance period of the change detection task. Across all four of their experiments, they found limited evidence that the size of the WM load influenced performance, providing evidence against a need to maintain goals in WM during the Stroop task, as such goal maintenance should have been impaired in the high set size. However, results from the current study indicate that individual differences in goal maintenance abilities are not based on a single trial, but on what happens when there are multiple congruent trials in a row. Because Moss et al.'s design did not include multiple congruent trials in a row (and was broken up by a change detection response, followed by a new trial) it did not provide people with an opportunity for goal neglect across trials (e.g., Meier & Kane, 2013).⁵

Potential Limitations, Implications, and Future Directions

Although the current two experiments provide additional evidence for the goal maintenance explanation of WMC-related differences in MC Stroop lists, a potential limitation is

⁵ Another alternative argument to the goal-maintenance explanation for list-wide PC effects not discussed in the introduction is short-lived conflict adaptation, in which Stroop effects diminish following previous incongruent trials (e.g., Botvinick et al., 2001; Gratton et al., 1992). However, because the lists in the current study were equivalent across reminder conditions, the likelihood of any trial following a previous congruent or incongruent trial is equal. Thus, sequential conflict adaptation cannot explain the effect of goal reminders. Further, it should be noted that WMC is frequently unassociated with conflict adaptation effects. Therefore, these aren't likely to be a good explanation for WMC differences in high congruency Stroop contexts (e.g., Keye et al., 2009, Meier & Kane, 2013; 2015, Unsworth et al., 2012).

that our experiments did not include a true control condition. Thus, one might argue that the effects could be due to either goal reminders improving performance or instead to the non-goal statements impairing performance. However, there are three reasons that we do not believe this alternative argument. First, as can be seen in Figure 5, performance among lower WMC individuals who receive goal reminders improves across the first 5 reminders and then remains low and similar to that of higher WMC individuals, demonstrating that the goal reminders did in fact improve performance. Second, the default condition is for WMC to correlate with Stroop performance in MC lists. Indeed, we have consistently found this significant negative correlation between WMC and Stroop effects for MC lists across numerous studies in which goals are not used (Hutchison, 2011; Hutchison et al., 2013; Hood & Hutchison, 2021). Thus, the unusual pattern is not the negative correlation between WMC and Stroop effects seen in the non-goal statement condition in Experiment 2 and in both groups for Experiment 1, but the elimination of this correlation in the goal reminder condition in Experiment 2. Third, in Hood and Hutchison we did include both a true control and a non-goal statement condition and found that the non-goal condition did not impair performance relative to the true control. For these reasons, we are confident that the difference is due to improved performance following goal reminders, rather than due to non-goal statements impairing performance.

We hope the current study will foster extensions of this goal reminder intervention in future studies. First, given the accumulated benefit of reminders over time in the current study, we recommend future studies examine the effectiveness of such goal reminders in helping individuals who struggle with mind wandering during sustained attention tasks such as the Sustained Attention to Response Task and Psychomotor Vigilance Task. In addition, we plan to implement goal reminders within other populations that have difficulties controlling attention

(e.g., older adults, children, and ADHD individuals) and investigating differences in brain activity following goal reminder versus non-goal statement conditions. Finally, future research can determine how goal reminders affect the interactions found between multiple forms of top-down control and in interactions between top-down and bottom-up forms of control (Hutchison, 2011; Hutchison et al., 2016).

Conclusion

There are individual differences in the ability to regulate behavior in a goal-directed manner to overcome automatic, habitual responses. Goal reminders are one effective way to increase performance for those who struggle with maintaining such attention control. The current results demonstrate that, when explicit control is cued in the form of goal reminders, lower WMC individuals can perform just as well as higher WMC individuals. Thus, lower WMC individuals *can* maintain task goals, they just don't spontaneously do so. Additionally, as long as the effect of reminders are not dampened by interference from non-goal statements, the beneficial effects of goal reminders are long-lived and accumulate across the course of the experiment. These results build on recent work by providing stronger evidence that goal reminders eliminate the relationship between WMC and Stroop effects, providing additional evidence of the goal-maintenance explanation for WMC-related differences in MC Stroop lists. Further, and more broadly, these results add to newer conceptualizations of individual differences in WMC by providing evidence that WMC is not so much about capacity per se, but instead about the ability to control attention and avoid attentional lapses.

Open Practices Statement

The experiment was not preregistered. Complete data files, R-Script, Output, and all materials are available at: <https://osf.io/3ej8b>.

References

- Balota, D. A., & Faust, M.E. (2001). Attention in dementia of the Alzheimer's type. In F. Boller & S. Cappa (Eds.), *The handbook of neuropsychology: Aging and dementia* (2nd ed., pp. 51-80). Elsevier Science.
- Barker, J. E., & Munakata, Y. (2015). Time isn't of the essence: Activating goals rather than imposing delays improves inhibitory control in children. *Psychological Science*, *26*, 1898-1908.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.
- Blaye, A., & Chevalier, N. (2011). The role of goal representation in preschoolers' flexibility and inhibition. *Journal of Experimental Child Psychology*, *108*, 469-483.
- 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., & Selmecky, D. (2015). Expectations and experience: Dissociable bases for cognitive control? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *41*, 1349-1373.
- Bugg, J. M., & Hutchison, K. A. (2013). Converging evidence for control of color-word Stroop interference at the item level. *Journal of Experimental Psychology: Human Perception & Performance*, *39*, 433-449.

- 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 & 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.
- Burgonye, A. P., & Engle, R. W. (2020). Attention control: A cornerstone of higher-order cognition *Current Directions in Psychological Science*, *13*, 190-193.
- Chevalier, N., & Blaye, A. (2009). Setting goals to switch between tasks: Effect of cue transparency on children's cognitive flexibility. *Developmental Psychology*, *45*, 782-797.
- Cohen-Shikora, E. R., Suh, J., & Bugg, J. M. (2019). Assessing the temporal learning account of the list-wide proportion congruence effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *45*, 1703-1723.
- Conway, A. R. A, Cowan, N., A., & 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, 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 & Applied*, *142*, 373-385.

- Cunningham, S., Scerbo, W., & Freeman, F. (2000). The electrocortical correlates of daydreaming during vigilance tasks. *The Journal of Mental Imagery, 24*, 61-72.
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology, 64* (2013), 135-168.
- 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.
- Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2020). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General, 150*, 242-275.
- 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. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23.
- Entel, O., & Tzelgov, J. (2019). When working memory meets control in the Stroop effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. Advance online publication. doi: 10.1037/xlm0000790
- Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*, 777-799.
- Gonthier, C., Ambrosi, S., & Blaye, A. (2021). Learning-based before intentional cognitive control: Developmental evidence for a dissociation between implicit and explicit control. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. Advance online publication. <http://dx.doi.org/10.1037/xlm0001005>

- 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.
- Hood, A. V. B., & Hutchison, K. A. (2021). Providing goal reminders eliminates the relationship between working memory capacity and Stroop errors. *Attention, Perception, & Psychophysics*, *83*, 85-96.
- Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *33*, 645-662.
- 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, & 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 & 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., Hood, A. V., & Charbonneau, B. (2022, August 2). Data. Retrieved from osf.io/vba47
- 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 & Personality Science*, *4*, 74-81.

- Jacoby, L. L., 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 28 July 2021.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.
- Kalanthroff, E., Avnit, A., Henik, A., Davelaar, E. J., & Usher, M. (2015). Stroop proactive control and task conflict are modulated by concurrent working memory load. *Psychonomic Bulletin & Review*, *22*, 869-875.
- 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-Germeys, I., & 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*, 614-621.
- 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.
- Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*, 1147-1185.
- Keye, D., Wilhelm, O., Oberauer, K., & Ravenzwaaij, D. (2009). Individual differences in conflict-monitoring: testing means and covariance hypothesis about the Simon and the Eriksen Flanker task. *Psychological Research*, *73*, 762-776.

- Krimsky, M., Forster, D. E., Llabre, M. M., & Jha, A. (2017). The influence of time on task on mind wandering and visual working memory. *Cognition, 169*, 84-90.
- Lee, M. D., & Wagenmakers, E.-J. (2013). Bayesian cognitive modeling: A practical course. New York, NY: Cambridge University Press.
- Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 7*, 1171.
- Logan, G. D., & Zbrodoff, N. J. (1982). Constraints on strategy construction in a speeded discrimination task. *Journal of Experimental Psychology: Human Perception & 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, & Cognition, 35*, 196-204.
- McVay, J. C., & Kane, M. J. (2012). Drifting from slow to “d’oh!”: Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology, Learning, Memory, & Cognition, 38*, 525-549.
- 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, & Cognition, 39*, 748-759.

- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167-202.
- 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, & Cognition, 38*, 757-763.
- 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.
- Morey, R. D., Rouder, J. N., & Jamil, T. (2015). Bayesfactor: Computation of bayes factors for common designs. r package version 0.9. 12-2.
- Moss, M. E., Kikumoto, A., & Mayr, U. (2020). Does conflict resolution rely on working memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 46*, 2410-2426.
- 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.
- Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 44*, 501-526.
- Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General, 148*, 1335-1372.

- Redick, T. S. (2014). Cognitive control in context: working memory capacity and proactive control. *Acta Psychologica, 145*, 1-9.
- Richmond, L. L., Burnett, L. K., Morrison, A. B., & Ball, H. (2021). Performance on the processing portion of complex working memory span tasks is related to working memory capacity estimates. *Behavior Research Methods*. Advance online publication. doi: 10.3758/s13428-021-01645-y
- Simpson, A., & Riggs, K. J. (2005). Inhibitory and working memory demands of the day-night task in children. *British Journal of Developmental Psychology, 23*, 471-486.
- Simpson, A., & Riggs, K. J. (2007). Under what conditions do young children have difficulty inhibiting manual actions? *Developmental Psychology, 43*, 417-428.
- Schmidt, J. R. (2013). Temporal learning and list-level proportion congruency: Conflict adaptation or learning when to respond? *PLoS One, 8*, 1-12.
- 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.
- Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11*, 771-799.
- 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.
- Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness & Cognition, 13*, 657-690.

- Spinelli, G., & Lupker, S. J. (2022). Conflict-monitoring theory in overtime: Is temporal learning a viable explanation for the congruency sequence effect? *Journal of Experimental Psychology: Human Perception & Performance*. Advance online publication. <https://doi.org/10.1037/xhp0000996>
- 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 & Performance*, *45*, 995-1014.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662.
- Thomson, D. R., Seli, P., Besner, D., & Smilek, D. (2014). On the link between mind wandering and task performance over time. *Consciousness & Cognition*, *27*, 14-26.
- Tsukahara, J. S., Harrison, T. L., Draheim, C., Martin, J. D., & Engle, R. W. (2020). Attention control: The missing link between sensory discrimination and intelligence. *Attention, Perception, & Psychophysics*, *82*, 3445-3478.
- 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., Redick, T. S., Spillers, G. J., & Brewer, G. A. (2012). Variation in working memory capacity and cognitive control: Goal maintenance and microadjustments of control. *Quarterly Journal of Experimental Psychology*, *65*, 326-355.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*, 1302-1321.

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.