Working-Memory Capacity and the Control of Attention: The Contributions of Goal Neglect, Response Competition, and Task Set to Stroop Interference

Michael J. Kane
University of North Carolina at Greensboro

Randall W. Engle
Georgia Institute of Technology

Individual differences in working-memory (WM) capacity predicted performance on the Stroop task in 5 experiments, indicating the importance of executive control and goal maintenance to selective attention. When the Stroop task encouraged goal neglect by including large numbers of congruent trials (*RED* presented in red), low WM individuals committed more errors than did high WM individuals on the rare incongruent trials (*BLUE* in red) that required maintaining access to the "ignore-the-word" goal for accurate responding. In contrast, in tasks with no or few congruent trials, or in high-congruency tasks that followed low-congruency tasks, WM predicted response-time interference. WM was related to latency, not accuracy, in contexts that reinforced the task goal and so minimized the difficulty of actively maintaining it. The data and a literature review suggest that Stroop interference is jointly determined by 2 mechanisms, goal maintenance and competition resolution, and that the dominance of each depends on WM capacity, as well as the task set induced by current and previous contexts.

Baddeley and Hitch (1974; Baddeley, 1986, 1993) proposed a "working memory" (WM) model that emphasized the dynamic interaction of memory maintenance and attention control in the service of complex cognition. In fact, individual-difference measures of WM capacity have turned out to be impressive predictors of a variety of cognitive abilities, including language comprehension, learning, and fluid reasoning (e.g., Daneman & Carpenter, 1980; Kyllonen & Christal, 1990; Shute, 1991; Süß, Oberauer, Wittman, Wilhelm, & Schulze, 2002). The present investigation explored the idea that the attentional, "executive" component of the WM system is specifically responsible for the covariation between measures of WM and higher order cognition (e.g., Conway & Kane, 2001; Engle, 2001, 2002; Engle, Kane, & Tuholski, 1999).

Our view is that simple memory ability, alone, cannot account for the strong association between WM span and complex cogni-

tion. Supporting evidence comes from structural-equation-modeling studies that tested participants in "span" tasks of WM and short-term memory (STM), as well as nonverbal tests of general fluid intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999). The WM span tasks required participants to maintain information on-line while they intermittently processed unrelated information. For example, the Operation Span (OSPAN) task (Turner & Engle, 1989) interleaved short series of words to be recalled with simple equations to be solved. The STM span tasks also required immediate recall of word lists, but unlike the WM tasks, they did not require any additional processing of unrelated information. WM and STM span tasks were strongly correlated with each other at the latent-construct level, reflecting that both involved the retention of unrelated words. However, only the WM construct correlated with a fluid-intelligence construct. The STM construct did not share unique variance with intelligence, suggesting that something other than simple retention processes must account for the relation between WM and complex cognitive ability. Moreover, partialing out the shared variance between the WM and STM constructs, which should reflect the common memorial demands of WM and STM tasks, had no effect on the correlation between the WM and intelligence constructs (which remained in the .60 range across studies).

Because pure memorial processes do not account for the covariation of WM capacity with higher order cognition, we have argued that WM = STM + controlled attention, and that the general, controlled-attention component of the WM system primarily drives the predictive utility of WM span tasks (Engle, 2001, 2002; Engle, Kane et al., 1999; Engle, Tuholski et al., 1999; Kane & Engle, in press). By our view, WM span tasks involve most of the processes required by STM span tasks, represented by the shared variance between the two constructs. In addition, WM tasks

make greater demands on attention control than do STM tasks, because WM tasks present a secondary task to interfere with the primary retention task. Information maintenance in the face of interference is the critical function of WM capacity, and so one may examine the attention-control capabilities involved in interference resistance to understand the covariation between WM capacity and complex cognition (e.g., Conway & Engle, 1994; Hasher & Zacks, 1988; Kane & Engle, 2000; Rosen & Engle, 1998).

Our use of the term *attention control* derives from Shallice's (1988; Norman & Shallice, 1986) conception of a "supervisory attention system" that is engaged during conflicts among task goals, external stimuli, and well-learned response schemas. In this vein, recent research has pointed to significant attention-control differences between high- and low-WM-span individuals outside of traditional memory tasks, specifically in contexts that present potent competition between task goals and habitual responses. For example, in an update of the "cocktail party" effect, Conway, Cowan, and Bunting (2001) tested high and low spans, on the basis of OSPAN scores, in a dichotic-listening task in which the participant's name and a yoked control name were spoken in the to-be-ignored channel. Previous research indicated that approximately 33% of individuals detect their name when it is presented in the "unattended" message (Moray, 1959; Wood & Cowan, 1995). Conway et al. (2001) found that only 20% of high spans detected their name when it was presented, but a full 65% of low spans did (no participants detected the control name). These findings indicate that low spans have more difficulty resisting the lure of a powerful orienting cue than do high spans.

Kane, Bleckley, Conway, and Engle (2001) reported similar span differences in an antisaccade paradigm, which presents a conflict between task goals and visual orienting cues (Hallett, 1978; Hallett & Adams, 1980). Participants were presented with a flashing visual cue in the periphery, and their task was to direct their eyes and attention away from the cue, in the opposite direction, to identify a pattern-masked letter. Thus, an automatic orienting response had to be prevented, or quickly recovered from, in favor of an opposing goal-directed response. High spans made fewer eye-movement errors toward the cue, they recovered from these eye-movement errors more rapidly, they initiated antisaccades more quickly, and they identified targets more quickly than did low spans. Thus, the OSPAN task predicted performance in a visual-attention task requiring the resolution of competition between a goal and a reflex.

Dichotic-listening and antisaccade tasks are clearly "attentional" tasks that have little surface similarity to either WM span tasks or other conventional tests of memory. Yet OSPAN, a marker of WM capacity, reliably predicts performance on such attention tasks in healthy, college-attending adults. Notions of WM capacity that emphasize number of "chunks" (e.g., G. A. Miller, 1956), managing simultaneous processing and storage (e.g., Daneman & Carpenter, 1980), or interdependence on knowledge, skill, or strategy (e.g., Ericsson & Delaney, 1999; McNamara & Scott, 2001) cannot easily account for these lower level attention findings. Rather, these data appear to be converging on the importance of WM capacity for maintaining access to even a single representation or intention in guiding action. This access is especially important (and, perhaps, is only important) when significant sources of interference or response competition are present.

Our view is related to Duncan's work on "goal neglect" (Duncan, 1990, 1993, 1995; for related views, see De Jong, 2000, 2001; De Jong, Berendsen, & Cools, 1999; Roberts & Pennington, 1996; West, 2001). Duncan has argued that in contexts lacking strong external cues for action, the production of organized behavior requires that hierarchically organized goal abstractions bias the selection of responses. This bias is effected through an attentional goal-weighting process that relies on intact prefrontal cortex (PFC) function and represents the essence of general fluid intelligence. Duncan and colleagues have supported this view by demonstrating that low-intelligence individuals, patients with PFC damage, and participants in dual-task conditions often fail to respond according to task goals when the environment lacks appropriate action prompts (e.g., Duncan, Burgess, & Emslie, 1995; Duncan, Emslie, Williams, Johnson, & Freer, 1996). Of interest, the failure to act according to goal is typically accompanied by an intact ability to articulate the goal when queried directly. This indicates that the goal was temporarily lost from active memory, that is, it was "neglected," and yet was still retrievable from long-term memory.

Roberts and Pennington (1996) extended such ideas by arguing that, in order for an attention-control mechanism to bias responding towards a goal state, it must have that goal information easily accessible. If hierarchical goal abstractions or intentions are not actively maintained (e.g., "look away from the cue" in the antisaccade task), the system may be momentarily captured by salient distractors and thus produce disorganized and inappropriate behavior. For example, maintaining a secondary memory load increases reflexive errors in the antisaccade task by over 50% (Roberts, Hager, & Heron, 1994). However, by maintaining activation of task goals, especially in the face of salient conflicting stimuli, executive processes assure that global goal states will not be lost in favor of more local action cues in the environment. By this view, then, WM capacity is not limited to maintaining representations of external stimuli, but it also keeps action plans, goal representations, and task-relevant information in an easily accessible state, allowing them to appropriately bias responding (see also E. K. Miller, 2000).

In the present study, we used the Stroop (1935) task to further examine the interaction of WM capacity, active goal maintenance, and the blocking or inhibition of competing stimulus representations and action plans. The Stroop task is a mainstay of research concerning selective attention and the external versus executive control of behavior (see MacLeod, 1991, for a review). In the Stroop task, participants report the colors in which words, or word-like stimuli, are presented. When the color and word are in conflict, such as the word *RED* appearing in blue, color naming is slower and less accurate than when the word is unrelated to the color (e.g., *PIN,* or *XXX,* in blue), or when the color and word match (e.g., *BLUE* in blue). The response slowing to conflict, or "incongruent," stimuli is known as the *Stroop effect* or *Stroop interference.* Stroop interference may be viewed as analogous to the dichotic-listening and antisaccade effects described above, insofar as it represents the failure of a goal ("ignore the word") to control behavior in the face of a conflicting habitual response (reading the word). We reasoned that if WM capacity were important to goal maintenance in the face of competition from overlearned responses, then WM capacity should predict Stroop interference.

Our framework is consistent with a computational model, grounded in the clinical and neuroscience literatures, in suggesting that a critical determinant of Stroop performance is the active maintenance of task goals in the face of competition from habit (Cohen, Dunbar, & McClelland, 1990; Cohen & Servan-Schreiber, 1992). According to Cohen's connectionist model, group differences in Stroop interference arise from the ability, or inability, to keep the task goal sufficiently active (i.e., "ignore the word and respond to color"). That is, Stroop performance deficits in schizophrenics are modeled by reducing the activity, or "gain," of nodes that represent the context, or task demands, and this is putatively analogous to decreased dopaminergic activity in prefrontal cortex pathways. These task-demand nodes, when normally active, bias the downstream system away from the dominant response by locally inhibiting the prepotent response pathway at intermediate layers of the model. Therefore, when these goal states are insufficiently active, they fail to block the habitual response, and greater interference effects are observed. Blocking is thus accomplished in part through active maintenance of "intention" (see also Braver & Cohen, 2000; O'Reilly, Braver, & Cohen, 1999).

The intelligence and neuropsychological literatures actually present some difficulties for our Stroop predictions, however. Although there is now substantial evidence that WM capacity is related to general fluid intelligence and PFC functioning, there is only equivocal evidence for a relation between these constructs and Stroop performance. Many studies indicate significant Stroop correlations with intelligence and significant Stroop deficits in patients with PFC damage (e.g., Jensen, 1965; Jensen & Rohwer, 1966; Perret, 1974; Richer et al., 1993; Spivack, Levine, & Sprigle, 1959). Other studies, however, find no such correlations or deficits (e.g., Ahola, Vilkki, & Servo, 1996; Jensen, 1965; Jensen & Rohwer, 1966; Shallice & Burgess, 1991; Spilsbury, 1992), although brain-imaging studies do consistently indicate PFC involvement in the Stroop effect (e.g., Bench et al., 1993; Pardo, Pardo, Janer, & Raichle, 1990; Zysset, Müller, Lohman, & von Cramon, 2001). By our reading of the relevant literatures, studies failing to link Stroop interference to intelligence or PFC generally minimized WM demands of the Stroop task by presenting only incongruent trials in a continuous sequence. Such blocked presentations of incongruent trials may minimize WM involvement because the task demands remain consistent across trials, making it easier for participants to keep the goal in mind. That is, once beginning the interference block, word information is never relevant to response, and so it must always be ignored on trial after trial. Every incongruent stimulus therefore reinforces the goal, to ignore the word, and so the task environment acts in the stead of the central executive (see also Cohen, Barch, Carter, & Servan-Schreiber, 1999).

We therefore predicted that the relation between WM capacity and Stroop interference would be most obvious if WM demands were increased by presenting congruent trials, in which the color and word match (e.g., the word *RED* printed in red), in addition to incongruent trials. A mixed-trial Stroop task that includes congruent stimuli puts more of a premium on keeping the unusual task requirements active and accessible in WM than does a blocked-trial task that presents only incongruent stimuli. This is because, on congruent trials, the word that is presented leads to the same response as the color, and so the stimulus does not reinforce the "ignore-the-word" task demands. A Stroop context including many

congruent trials will allow successful responding on most trials even if participants completely neglect the goal and respond habitually by reading the word. Thus, responses on the rare incongruent trials in such a context will be especially sensitive to successful goal maintenance; if the goal is momentarily lost, errors will be likely. Research outside of the individual-differences domain has indeed shown that the magnitude of Stroop and Stroop-like interference increases with the proportion of congruent trials in the design (e.g., Logan & Zbrodoff, 1979, 1998).

Important—if not explicit—to this view of Stroop interference is the idea that both "attention" and "memory" play a role in determining the effect. That is, resolution of the response competition between color and word dimensions in the Stroop task, an *attentional* process, will only be engaged when the goal to do so is sufficiently maintained in active *memory*. Stroop interference is thus dually determined. It may reflect a failure to maintain the task goal of ignoring the word dimension, which should lead to word reading and result in errors on incongruent trials (or some outlying slow responses, or both; see below). Interference may also, or instead, reflect the time-consuming process of resolving response competition in service of a successfully activated goal to do so, which should lead to a relatively consistent but modest slowing on incongruent trials relative to baseline trials.

We do not claim that this dual-mechanism account of Stroop interference is entirely new, but simply that it is not always explicitly recognized. As discussed above, the role of goal maintenance in resolving interference is implemented computationally in the Cohen model, where competition resolution—through inhibition—depends on activated goal representations (e.g., Cohen et al., 1990; Cohen & Servan-Schreiber, 1992).[1] A dual-component account of Stroop interference may also be implicit in the interpretation of response-time distributions in the task (e.g., Heathcote, Popiel, & Mewhort, 1991; Spieler, Balota, & Faust, 1996). These distributions show that one component of the latency increase on incongruent trials relative to baseline trials is a regular shifting of the entire incongruent-trial distribution. This shift may reflect the response competition elicited by every incongruent trial. Another component of the distribution, in the tail, reflects a small subset of very slow responses on incongruent trials, and this skewing may reflect momentary lapses of goal maintenance (i.e., lapses that were "caught" and subsequently corrected before an overt error was produced).

A dual-component view of interference is also suggested by research exploring the attentional deficits of the elderly and schizophrenics. Healthy older adults show equivalent interference to young adults when analyses isolate the response-competition

---

[1] Strictly speaking, the Cohen model is not a dual-mechanism model, because goal maintenance determines the degree of response competition that is present to be resolved; that is, the extent to which maintenance is successful determines the extent to which competition will disrupt performance, and so maintenance and competition cannot be completely decoupled (this is true also of the somewhat parallel mechanism of interference control in Jacoby's Counter model; Jacoby et al., 1999; Trainham et al., 1997). Although current computational models may be characterized as having some distinction between goal maintenance and goal execution, insofar as even perfect maintenance or control may fail to completely block the influence of word information, competition is countered in these models through a single mechanism.

component of the latency distribution (Spieler et al., 1996), and schizophrenics show equivalent interference to controls when it is assessed in contexts that reinforce task goals, making goal neglect unlikely (Perlstein, Carter, Barch, & Baird, 1998). However, in measures and contexts more sensitive to goal maintenance and neglect, these participant groups, along with Alzheimer's patients (Spieler et al., 1996), show substantial increases either in errors or in the skew of the response-time distribution on incongruent trials. This pattern of results indicates that normal aging and some disease states produce particular difficulties with active goal maintenance in the Stroop task, and not with resolving response competition proper.

Answers to the questions of whether WM capacity is related to Stroop interference, and whether Stroop interference may be decomposed into attentional and memorial components, should have broad influence on general theory regarding WM, fluid intelligence, PFC functioning, and the Stroop effect. Moreover, the study of neurological disorders, psychopathology, and clinical practice is increasingly informed by research on both WM capacity (e.g., Brewin & Beaton, 2002; Klein & Boals, 2001; Park, Püschel, Sauter, Rentsch, & Hell, 1999; Swanson & Sachse-Lee, 2001) and Stroop effects (Buckley, Blanchard, & Hickling, 2002; Lundh, Wikstroem, & Westerlund, 2001; Seddon & Waller, 2000; Taylor & Russo, 2001). Advancing our understanding of the relation between WM and the Stroop task will therefore have substantial applied, as well as theoretical, implications.

In five experiments we demonstrate that WM span does, in fact, predict Stroop interference and that these span differences vary with list-wide proportions of congruent trials, consistent with a dual-mechanism view of the Stroop effect. The data indicate that different manifestations of Stroop interference are dissociable by their relation to WM capacity. That is, the locus of WM span differences in Stroop interference is systematically shifted between response latencies and error rates depending on the set induced by the current, and the previous, "congruency" context. These dissociations, along with other converging behavioral and neuropsychological findings we will review, are consistent with the idea that Stroop interference is a multiply determined effect. Specifically, interference involves the time-consuming action of a relatively stable competition-resolution mechanism, the failure of a more transient goal-maintenance mechanism, or both, depending on the task context. Moreover, WM capacity is related to the successful operation of both mechanisms. Our data also imply that the goal-maintenance mechanism is partly responsible for producing the Stroop "facilitation" effects that are sometimes observed on congruent trials compared with neutral trials (see MacLeod, 1998; MacLeod & MacDonald, 2000a, 2000b).

EXPERIMENT 1

Experiment 1 presented high- and low-WM-span participants with one of two computerized versions of the Stroop task differing in the proportion of congruent trials in the design, either 0% or 75%. We hypothesized that activation, or accessibility, of task-goal information (e.g., "ignore the word") is important to successful performance on Stroop incongruent trials. Therefore, increasing the number of congruent trials, where the goal could be neglected without negative consequence, will increase the magnitude of observed interference effects. Increasing the number of

congruent trials will make the task more sensitive to goal maintenance and neglect on incongruent trials, and so the task will also be more sensitive to individual differences in WM capacity.

Method

Participant Screening for WM Capacity

Participants were first screened for WM capacity with the OSPAN task in which they solved series of simple mathematical operations interleaved with unrelated words to memorize (for details, see La Pointe & Engle, 1990). A Micro Experimental Laboratory (MEL) 2.0 program (Schneider, 1988) presented the stimuli at the center of a color monitor. Participants were tested individually, and sat at a comfortable viewing distance from the monitor.

Participants saw one operation-word string at a time, and each set of operation-word strings ranged from two to six items in length. For example, a set of three strings might be the following:

$$IS \ (9 \ / \ 3) + 2 = 5 \ ? \ drill$$

$$IS \ (5 \times 1) - 4 = 2 \ ? \ beach$$

$$IS \ (2 \times 2) + 3 = 7 \ ? \ job$$

The participant read the equation aloud as soon as it appeared; pausing before reading was not permitted. After reading the equation, the participant verbally indicated whether the provided answer was correct and then immediately read the word aloud. The experimenter then presented the next operation by pressing a key, and the participant immediately read this operation aloud. This sequence continued until three question marks (*???*) cued the participant to recall all of the words from that set only. Participants wrote the words on an answer sheet in the order in which they had been presented.

The OSPAN score was the sum of the recalled words for only those sets completely recalled in correct order. Three sets of each length (from two to six operation-word pairs) were presented, allowing scores to range from 0 to 60. The different set sizes appeared in an unpredictable order (fixed across participants), and so the number of words to recall was unknown until recall.

Prior research has established the OSPAN task to be a reliable and valid marker of WM capacity. With respect to reliability, OSPAN demonstrates adequate internal reliability as indexed by Cronbach's alpha (approximately .65–.75; Conway et al., 2002; Engle, Cantor, & Carullo, 1992; Engle, Tuholski et al., 1999; La Pointe & Engle, 1990). OSPAN scores also remain stable over test–retest intervals of a few minutes ($r = .77$–.79; Turley-Ames & Whitfield, 2002), to 3 weeks (stability coefficient = .82; Klein & Fiss, 1999), to 3 months (stability coefficient = .76; Klein & Fiss, 1999). Moreover, in studies comparing versions of OSPAN that differ in arithmetic difficulty, Conway and Engle (1996) and Lehto (1996) found correlations among them to be in the range of .70–.80.

With respect to validity, the correlations among OSPAN and other WM measures, such as Reading Span (Daneman & Carpenter, 1980), approach the reliability estimates of OSPAN, itself, ranging between $r = .40$ and $r = .80$ (Conway et al., 2002; Engle, Tuholski et al., 1999; Lehto, 1996). Moreover, both exploratory and confirmatory factor analyses reveal that OSPAN fits a tight latent variable with WM tasks using quite different materials. This WM latent variable is correlated with, but separate from, a latent variable constructed from STM tasks, and it accounts for a great deal of the variance in a latent variable constructed from measures of fluid nonverbal reasoning (Conway et al., 2002; Engle, Tuholski et al., 1999). Finally, specific strategies or skills cannot account for the covariation among OSPAN and measures of complex cognition. Partialing out individuals' processing times on the equations does not reduce correlations between OSPAN and the Scholastic Aptitude Test (SAT), partialing out

individuals' study times on the target words does not reduce correlations between OSPAN and the SAT, and individually matching the OSPAN equations to each participant's math ability does not reduce correlations with the SAT (Conway & Engle, 1996; Engle et al., 1992).

Stroop Task

Participants

Eighty-seven undergraduates from Georgia State University participated in Experiment 1 as partial fulfillment of a course requirement. These participants were identified and invited from a pool of approximately 300–400 people who had participated in OSPAN during the time frame of this experiment: Volunteers were 40 participants from the top quartile of the distribution, with scores $\geq$ 19 (hereafter, "high spans"), and 47 from the bottom quartile, with scores $\leq$ 9 (hereafter, "low spans").[2] All had correctly solved at least 85% of the OSPAN operations, and all had normal or corrected-to-normal vision as indicated by self-report. Participation in the Stroop task followed OSPAN by as little as 1 day, and as many as 100 days.

Design

The design was a $2 \times 2 \times 3$ mixed-model factorial, with span (high, low) and congruency proportion (0%, 75%) as between-subjects variables, and trial type (congruent, neutral, incongruent) as a within-subjects variable (but participants in the 0% congruent condition did not see congruent trials). Participants were randomly assigned to one of the two proportion-congruent conditions.

Apparatus and Materials

A MEL 2.0 program presented the stimuli in standard font on a Dell-brand computer color monitor. A MEL response box (Model 200) collected vocal latency and accuracy data.

The stimulus list for each version of the Stroop task consisted of 288 total trials. Each trial presented a letter string representing one of three trial types. Congruent trials presented the capitalized word, *RED, BLUE,* or *GREEN,* in its matching hue (e.g., *RED* in red). Incongruent trials presented one of these color words in a mismatching hue (e.g., *BLUE* in red). Neutral trials presented the letter string, *JKM, XTQZ,* or *FPSTW,* in red, blue, or green. Neutral stimuli were chosen to match color words in length, but to share no letters with the color words, given that even a few shared letters between Stroop stimuli and color names may produce some small level of interference (see MacLeod, 1991).

Each participant saw 36 critical trials in each of the three trial types (except for participants in the 0% congruent condition, who saw zero congruent trials), and response times and accuracy were analyzed for only these critical trials. For critical congruent trials, each color word appeared 12 times in its matching color. For critical incongruent trials, each color word appeared 6 times in each of the two conflicting colors. For neutral trials, all of which were critical, each letter string appeared 4 times in each of the three possible colors. The remaining filler trials—which were not identifiably "filler" to the participant—presented congruent or incongruent stimuli to create the appropriate list-wide proportions of congruent trials. In the 0% congruent condition, all 216 filler trials were incongruent, and in the 75% congruent condition, all 180 filler trials were congruent. All letter strings and target colors were equally represented across filler and critical trials. Critical trials were coded as such in the computer program that presented stimuli and collected responses, but they were otherwise identical to the filler trials. For each participant, the computer program randomly determined the presentation order of all 288 critical and filler trials.

Procedure

All participants were individually tested. During the Stroop task, the overhead lights were turned off, and a 50-watt lamp was turned on and placed on the floor. The experimenter remained in the room for the entire session.

The basic requirement of the Stroop task was to name aloud the color in which a letter string was presented onscreen, as quickly and accurately as possible. The task began with two practice blocks. The first block of 12 trials was designed to familiarize participants with the microphone. Participants simply read aloud each color word (*RED, BLUE,* or *GREEN*) that was displayed in white. The second block of 36 trials was designed to familiarize participants with color naming and with the neutral letter strings. Here, participants named the color of each neutral stimulus (*JKM, XTQZ,* or *FPSTW*) that was presented in red, blue, or green. Each letter string and color appeared equally often in Blocks 1 and 2. The final block of 288 trials consisted of the critical and filler trials described above.

We hoped to motivate participants to consistently ignore word information during the task, especially those in the 75% congruent condition who might find it easier to purposely read the word on most trials. The experimenter therefore explicitly instructed all participants to ignore the word on every trial, even if the word and color matched on some (or many) trials. The experimenter further explained that we were only interested in their performance on incongruent trials, and not congruent trials, and that they would perform much better on the important, incongruent trials if they always tried to ignore the word in favor of the color. As well, the experimenter instructed all participants that they would be observed to determine that they never looked away from the word or squinted their eyes during the task. Finally, all participants were encouraged to respond quickly while being accurate.

A voice-activated relay recorded response latencies to each letter string, and these latencies reflected time from stimulus onset to response onset. The experimenter recorded accuracy by keying in the word spoken by the participant ("red," "blue," "green"), with each word represented by a button on the response box. A fourth button was used to indicate an error of speech, such as stuttering when naming the color (e.g., "b-blue") or slurring two names together (e.g., "gre-blue").

Each trial began with the presentation of a yellow *READY?* signal at the center of the screen against a black background. The ready signal remained on screen until the participant pressed the keyboard's space bar, and was followed by a 1,000-ms blank screen. A centered, white fixation signal (******) then appeared for 200 ms, followed immediately by the target letter string that remained in the center of the screen until response (or 30 s, whichever came first). The experimenter then logged the response through key-press, and then after a 500-ms delay, the next ready signal appeared.

Results

Participants

The mean OSPAN scores for the high- and low-span participants, respectively, were 23.25 (SD = 6.36, range = 19–54) and 6.53 (SD = 2.05, range = 2–9).

---

[2] Across many years of testing undergraduate participants in the OSPAN task at comprehensive state universities in different parts of the country, our laboratories and others have found that the typical quartile cut-offs are 9 and 19, with some variation from year to year and from location to location (e.g., cut-offs of 8 and 20). The "distributions" discussed in the text refer to this modal distribution, as opposed to the specific distributions acquired from the particular participants tested in Experiment 1.

Stroop Task

For all analyses reported hereafter, the alpha level is set at .05; we present *p* values for effects and interactions that do not reach this criterion.

Response Times

For all response-time analyses reported hereafter, group means were taken across individual participants' trimmed-mean latencies for correct responses in each condition. In trimming, any naming trial that exceeded the participant's mean response time for that condition by more than three standard deviations was deleted, as was any trial with a latency less than 200 ms. In all experiments reported, fewer than 2% of trials were trimmed for both high- and low-span participants. Mean response times in Experiment 1 by span group, congruency proportion, and trial type are presented in Table 1.

Interference effects. Interference effects and facilitation effects are not independent, because they are both calculated against the same baseline (interference contrasts incongruent with neutral trials, and facilitation contrasts congruent with neutral trials). However, as is normative in Stroop research, we analyze these effects separately, most importantly because the 0% and 75% congruent conditions can only be compared directly on neutral and incongruent trials (the 0% condition presented no congruent trials). By inspection of the means for Experiment 1, incongruent trials yielded substantially slower responses than did neutral trials, and this interference effect was larger for the 75% congruent condition than for the 0% condition. Low spans showed only somewhat larger interference effects than did high spans.

A 2 (span group) $\times$ 2 (congruency proportion) $\times$ 2 (trial type) mixed-model analysis of variance (ANOVA), with trial type as a repeated measures variable, tested these observations. Responses on incongruent trials were slower than on neutral trials, indicating a significant Stroop interference effect, $F(1, 83) = 240.64$, $MSE = 2,669.56$, and this Stroop effect was significantly larger for participants in the 75% congruent condition than for those in the 0% condition (i.e., congruency proportion interacted with trial

type), $F(1, 83) = 48.23$, $MSE = 2,669.56$. However, there were no significant effects involving span: High- and low-span participants showed similar naming latencies overall, $F(1, 83) = 1.16$, $MSE = 24,695.07$, $p = .28$, and this was true for both the 0% and 75% congruent conditions (with $F < 1$ for the interaction of span and congruency proportion). Most important for present purposes, low spans did not show larger interference effects than did high spans, either overall, $F(1, 83) = 2.22$, $MSE = 2,669.56$, $p = .14$, or in consideration of congruency proportion: The three-way interaction of span, congruency proportion, and trial type indicated that the span groups showed equivalent interference across congruency conditions ($F < 1$).

Facilitation effects. Of secondary interest, we examined congruent versus neutral response times in the 75% congruent condition. Increasing the list-wide proportion of congruent trials should have made goal neglect (and so, overt word reading) more likely in low spans. We thus expected low spans to show more benefit from a color–word match on congruent trials than would high spans.

By inspection of the response-time means in Table 1, congruent trials yielded faster responses than did neutral trials, reflecting response-time "facilitation." Moreover, low spans appeared to show larger facilitation effects than did high spans in the 75% congruent condition. A 2 (span) $\times$ 2 (trial type) mixed-model ANOVA confirmed these impressions: Congruent-trial latencies were significantly shorter than neutral-trial latencies, $F(1, 42) = 64.61$, $MSE = 889.97$, and this facilitation was greater for low spans than for high spans, $F(1, 42) = 3.63$, $MSE = 889.97$, $p = .06$ (facilitation was significant for both high- and low-span groups individually, both $Fs > 35$). As indicated by facilitation, then, low spans appeared less likely than high spans to ignore word information during the 75% congruent task. Low spans thus benefited more when this information was congruent with the goal.

Errors

Participants' mean proportion of color-naming errors, by WM span, congruency proportion, and trial type, are displayed in Ta-

Table 1

Mean Response Latencies (in Milliseconds) and Error Rates in Color Naming, With Standard Deviations, by Working-Memory Span Group, Congruency Proportion, and Trial Type in Experiment 1

Participant group	Trial type											
	Congruent				Neutral				Incongruent			
	Latency		Error rate		Latency		Error rate		Latency		Error rate	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
0% congruent												
High span[a]					646	68	0.01	0.01	699	94	0.02	0.03
Low span[b]					659	118	0.01	0.02	742	186	0.04	0.05
75% congruent												
High span[c]	620	106	0.02	0.03	660	104	0.02	0.02	828	122	0.10	0.07
Low span[d]	611	73	0.01	0.02	675	66	0.02	0.02	860	132	0.18	0.14

[a] $n = 21$.   [b] $n = 22$.   [c] $n = 19$.   [d] $n = 25$.

ble 1. Anecdotally, we can report that virtually every error reflected reading the word, either in part (as in "gre-blue" for the intended response "blue") or outright.

Interference effects. Inspection of the means suggests that, unlike response times, error rates were highly sensitive to span differences in interference, but this was true for the 75% congruent condition only. (As with response times, interference was measured by subtracting neutral error rates from incongruent error rates.) A 2 (span) $\times$ 2 (congruency proportion) $\times$ 2 (trial type) mixed-model ANOVA was conducted on error rates for neutral and incongruent trials. As with response times, a significant Stroop effect emerged, with more errors committed on incongruent than on neutral trials, $F(1, 83) = 74.20$, $MSE = 0.003$. Moreover, this interference effect was larger in the 75% congruent condition than in the 0% condition (i.e., congruency proportion interacted with trial type), $F(1, 83) = 34.14$, $MSE = 0.003$.

In contrast to the latency data, high and low spans performed differently. Low spans committed more errors overall than did high spans, $F(1, 83) = 5.73$, $MSE = 0.004$, and, critically, low spans showed a larger Stroop interference effect than did high spans, $F(1, 83) = 7.56$, $MSE = 0.003$. Finally, span interacted with trial type and congruency proportion, $F(1, 83) = 3.94$, $MSE = 0.003$, $p = .05$, indicating that the span difference in interference was larger in the 75% congruent condition than in the 0% condition.

To assess whether span differences in interference were significant in each congruency condition, we examined them separately. In the 0% congruent condition, the Stroop effect was significant overall, $F(1, 41) = 14.19$, $MSE = 0.001$, but high and low spans did not differ in interference, $F(1, 41) = 1.08$, $MSE = 0.001$, $p = .31$. In contrast, the 75% congruent condition produced an overall Stroop interference effect, $F(1, 42) = 61.63$, $MSE = 0.005$, as well as a significant difference in interference between high and low spans, $F(1, 42) = 6.55$, $MSE = 0.005$ (interference in the 75% condition was significant for both the high- and low-span groups individually, both Fs > 17). A full 40% of low-span participants committed more than 20% errors on incongruent trials, whereas only 11% of high spans did; error rates over 30% were demonstrated by 28% of low spans and no high spans. Clearly, in a high-congruency context, which allowed for goal neglect without frequent cost, low spans made more errors than did high spans on those rare trials where the goal ("ignore the word") and habit ("read the word") came in direct conflict.

Facilitation effects. As with response times, we examined facilitation effects in errors by comparing the proportion of errors committed on congruent versus neutral trials for participants in the 75% congruent condition. Error rates in these two conditions were near floor, and a 2 (span) $\times$ 2 (proportion congruent) $\times$ 2 (trial type) ANOVA on neutral and congruent trials indicated no significant facilitation effect and no span difference in facilitation (both Fs < 1).

Discussion

In Experiment 1, participants differing in WM capacity, as measured by the OSPAN task, were tested in one of two versions of the Stroop task that differed in the list-wide proportion of congruent trials (i.e., on which the color and word matched). We hypothesized that WM capacity should predict vulnerability to Stroop interference, but especially as the proportion of congruent trials in the context increased. In 0% congruent tasks, active goal maintenance should be least demanding because the task context repeatedly reinforces the goal, to ignore the word: Nearly every trial in the 0% condition presents an incongruent stimulus in which the word must be ignored for accurate responding. In a 75% congruent task, however, the context does not often reinforce the goal. Most trials present congruent stimuli, for which the habitual response of word reading is not detrimental.

Indeed, span differences in interference were robust in only the 75% congruent condition, suggesting span differences in goal neglect. But contrary to prototypical Stroop findings, these span differences arose in accuracy, and not in response time.[3] Although low spans demonstrated numerically larger response-time interference effects than did high spans in both congruency conditions, suggestive of span differences in resolving response competition, these span differences were not statistically significant. Thus, in the 0% congruent context, no WM-related differences were seen in either response time or error interference. The 75% congruent task, in contrast, appears to have promoted goal neglect, particularly for low-WM-span individuals. Low-span participants demonstrated significantly larger response-time facilitation effects, as well as larger error interference effects, and we would interpret both these effects as arising from periodic, overt word reading.

EXPERIMENT 2

We presented accuracy feedback in Experiment 2 in part to make closer contact with the Stroop literature. In healthy, adult participants, Stroop interference is more typically seen in latencies than in errors, and so here we tried to force any WM-related interference differences out of errors and into response times by emphasizing accuracy. Here we presented all high- and low-WM-span participants with two versions of the Stroop task, differing in the list-wide proportion of congruent trials. In the first task, 75% of the trials were congruent, and in the subsequent task, 0% of the trials were congruent. Thus, in contrast to Experiment 1, congruency was manipulated within subjects and between blocks. The 75% congruent task was identical to that used in Experiment 1, except that accuracy feedback was presented after every trial. This feedback was also presented during the 0% congruent task, which further differed from Experiment 1 in presenting half the number of trials to accommodate both tasks within a single session.

To preview, the data from the 75% task in Experiment 2 closely replicated those from Experiment 1, with span differences in interference arising in error rates, and not in response times, despite the presentation of accuracy feedback throughout. Also as in Experiment 1, span groups differed significantly in response-time facilitation.

Method

The OSPAN and Stroop task methods for Experiment 2 were identical to those of Experiment 1 with the following exceptions.

---

[3] As we noted in the introduction, research with patient groups, such as those with schizophrenia (e.g., Perlstein et al., 1998) and Alzheimer's disease (Spieler et al., 1996), indicate exaggerated interference effects primarily in errors, as opposed to response times.

Participants

We tested 88 undergraduates (40 high spans, 48 low spans, using the same participation criteria as in Experiment 1) from Georgia State University and Georgia Institute of Technology, in return for partial fulfillment of a course requirement or for extra credit. All had normal or corrected-to-normal vision by self-report. Participants were drawn from a larger pool that had completed the OSPAN task approximately 30 min before beginning the Stroop task. The larger pool and these specific participants were different from those tested in Experiment 1.

Design

The design was a 2 × 2 × 3 mixed-model factorial, with span (high, low) as a between-subjects variable, and congruency proportion (75%, 0%) and trial type (congruent, neutral, incongruent) as within-subjects variables.

Apparatus and Materials

The stimulus list for the 75% congruent task in Experiment 2 was identical to that in Experiment 1. The 0% list was cut from 288 to 144 total trials, but the critical trials were identical to those from Experiment 1. Here, only 72 filler trials were presented in the 0% congruent task, all of which were incongruent. Each word and color was equally represented across filler and critical trials.

Procedure

The experiment began with the same two blocks of practice trials used in Experiment 1. Following practice, all participants completed the 75% congruent task and then the 0% congruent task. Before beginning each of the tasks, participants were encouraged to consistently ignore the words, even though many trials might present matching color–word combinations, because we were only interested in performance on incongruent trials. Participants were also encouraged to respond as quickly as possible while remaining accurate.

Experiment 2 helped participants monitor accuracy by providing visual and auditory feedback after every trial. The message "Correct Response" or "Incorrect Response" was presented for 500 ms after the experimenter keyed in each response. Errors were additionally signaled by a tone, coincident with the visual feedback.

Results

Participants

The mean OSPAN scores for the high and low spans, respectively, were 22.88 (SD = 3.78, range = 19–35), and 6.06 (SD = 2.46, range = 0–9).

Stroop Task

Response Times

Mean response times in Experiment 2 by span group, proportion-congruent condition, and trial type are presented in Table 2.

Interference effects. Inspection of the means suggested that, in contrast to the 0% and 75% congruent conditions from Experiment 1, low spans had longer latencies overall than did high spans. However, the span differences in interference in both the 0% and 75% conditions appeared to be comparable with those in Experiment 1, where they were nonsignificant. A 2 (span group) × 2 (congruency proportion) × 2 (trial type) mixed-model ANOVA indicated that, although high spans had shorter naming latencies than did low spans overall, $F(1, 86) = 11.40$, $MSE = 45,504.65$, span did not interact significantly with either congruency ($F < 1$) or trial type, $F(1, 86) = 2.88$, $MSE = 3,713.32$, $p = .09$, and the three-way interaction of span, congruency, and trial type was also nonsignificant, $F(1, 86) < 1$, all indicating comparable response-time interference effects for high- and low-span participants across congruency conditions. There was, however, a significant Stroop effect overall, $F(1, 86) = 444.41$, $MSE = 4,023.06$, which was larger for the 75% congruent condition than for the 0% condition, $F(1, 86) = 102.03$, $MSE = 1,585.50$.

Facilitation effects. Inspection of the means in Table 2 suggests that response-time facilitation was greater for low spans than for high spans in the 75% congruent condition. A 2 (span) × 2 (trial type) mixed-model ANOVA confirmed that latencies for congruent trials were significantly shorter than those for neutral trials, $F(1, 86) = 138.64$, $MSE = 1,008.31$, and that low spans

Table 2

Mean Response Latencies (in Milliseconds) and Error Rates in Color Naming, With Standard Deviations, by Working-Memory Span Group, Congruency Proportion, and Trial Type in Experiment 2

	Trial type											
	Congruent				Neutral				Incongruent			
	Latency		Error rate		Latency		Error rate		Latency		Error rate	
Participant group	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
0% congruent												
High span[a]					646	86	0.02	0.02	736	109	0.03	0.03
Low span[b]					707	105	0.02	0.03	818	144	0.04	0.05
75% congruent												
High span[a]	640	92	0.01	0.02	681	92	0.02	0.03	855	120	0.09	0.09
Low span[b]	679	92	0.02	0.04	751	95	0.03	0.04	949	157	0.14	0.08

[a] $n = 40$.   [b] $n = 48$.

showed significantly larger facilitation effects than did high spans, $F(1, 86) = 10.10$, $MSE = 1,008.31$ (again, the facilitation effect was significant for each span group, with $Fs > 40$). As in Experiment 1, the 75% congruent condition was quite sensitive to span differences in response-time facilitation, but not in interference.

Errors

Participants' mean color-naming error rates in Experiment 2, by WM span, congruency proportion, and trial type, are displayed in Table 2.

Interference effects. Inspection of the means in Table 2 suggests that low spans showed a larger error interference effect than did high spans in the 75% congruent condition only, as they had in Experiment 1. A 2 (span) × 2 (congruency proportion) × 2 (trial type) mixed-model ANOVA indicated that low spans had a higher error rate overall than did high spans, $F(1, 86) = 6.17$, $MSE = 0.004$, that span interacted with congruency proportion, $F(1, 86) = 7.87$, $MSE = 0.002$, and that span interacted with trial type, indicating span differences in error interference, $F(1, 86) = 4.87$, $MSE = 0.003$. The three-way interaction of span, congruency proportion, and trial type approached conventional significance, $F(1, 86) = 2.88$, $MSE = 0.002$, $p = .09$.

To explore this predicted interaction, we tested for span differences in the 0% and 75% congruent conditions separately. In the 0% congruent condition, the span groups showed equivalent error rates overall ($F < 1$), and despite a significant Stroop effect, $F(1, 86) = 12.95$, $MSE = 0.001$, high and low spans showed equivalent interference, $F(1, 86) = 1.05$, $MSE = 0.001$, $p = .31$. In contrast, the 75% congruent condition elicited a higher overall error rate for low spans than for high spans, $F(1, 86) = 8.34$, $MSE = 0.005$, a significant interference effect overall, $F(1, 86) = 109.12$, $MSE = 0.004$, and most importantly, a larger interference effect for low spans than for high spans, $F(1, 86) = 5.09$, $MSE = 0.004$ (the error interference effect was significant for each span group, with $Fs > 26$). The significant span difference in interference clearly replicates that from Experiment 1.

Facilitation effects. For the 75% congruent condition, error proportions on congruent and incongruent trials were near floor and did not significantly differ from each other, $F(1, 86) = 1.81$, $MSE = 0.002$; span groups did not differ in error facilitation (or lack thereof) either ($F < 1$).

Discussion

Experiment 2 presented high- and low-span participants with the 75% congruent condition used in Experiment 1, but with accuracy feedback. Despite this salient feedback provided after every trial, low spans still committed more errors on incongruent trials than did high spans in the 75% congruent task. Low spans also showed larger response-time facilitation effects than did high spans. Both findings replicated Experiment 1. The reduction in the span difference in errors in Experiment 2 versus Experiment 1 suggests that the presentation of accuracy feedback helped low spans reduce their goal neglect somewhat, but obviously it did not eliminate it entirely. In Experiment 1, 40% of low spans made ≥ 20% errors, whereas in Experiment 2, only 21% of low spans committed this many errors. It should be noted that high spans' error rates in the

75% congruent condition remained remarkably constant between Experiments 1 and 2, suggesting that they were better able than low spans to monitor their performance without the presentation of overt feedback. In Experiments 1 and 2, 11% and 13% of high spans, respectively, made ≥ 20% errors.

Subsequent to the 75% congruent task, participants encountered a 0% congruent Stroop task that was shortened relative to that in Experiment 1. Presenting the 0% task after the 75% task did not appreciably change the results. In both Experiments 1 and 2, there was a nonsignificant tendency for low spans to show greater response-time interference than did high spans in the 0% congruent condition. However, that tendency did not approach significance in Experiment 1 ($p = .18$) and fell short of conventional significance levels in Experiment 2 ($p = .09$).

EXPERIMENTS 3A AND 3B

Although we predicted larger WM span differences in interference in the 75% congruent condition than in the 0% congruent condition, we expected span differences in the 0% condition to be more robust than they were. Presumably some degree of goal maintenance is necessary in even the 0% congruent Stroop task, but even if goal maintenance is almost entirely unnecessary in our 0% task, prior research on memory interference suggests that WM capacity is related to the effective resolution of response competition in service of the task goal (e.g., Conway & Engle, 1994). In fact, low spans' response-time interference effects in the 0% congruent task were over 50% larger than high spans' in Experiment 1, and over 25% larger in Experiment 2, despite the lack of statistically significant differences between groups. Perhaps span differences would have been more substantial in the 0% congruent task after less task practice, before the goal of ignoring word information had been practiced. In Experiment 1, all three proportion-congruent conditions presented the same number of total trials, so the 0% congruent task presented far more incongruent trials to each participant than did the 75% congruent task (252 vs. 36). In Experiment 2, the 0% congruent task was presented after 288 trials of the 75% task. It is therefore conceivable that span differences might be larger when the 0% congruent task is more novel. Moreover, the large span difference in errors in the 75% congruent condition may have reflected a lack of practice with incongruent stimuli as much as it reflected a failure of goal maintenance. Perhaps with more incongruent-trial practice, high and low spans would become equivalently able to withhold habitual responses in favor of goal-directed ones.

Experiments 3A and 3B presented high- and low-WM-span participants with the 0% congruent task before the 75% congruent task. Here we tested whether practice with incongruent trials—and not the probability of goal neglect—was responsible for the different pattern of WM span effects between the 0% and 75% tasks in Experiments 1 and 2. If lack of practice with incongruent trials is critical to producing span differences in interference, then reducing the number of incongruent trials encountered prior to the 0% congruent task should make it more sensitive to WM differences. Likewise, presenting the 75% congruent task after substantial practice on the 0% task should make the 75% task less sensitive to WM differences.

The two experiments differed only in their trial pacing (participant- vs. computer-paced in Experiments 3A and 3B, re-

spectively), and because this variable did not appreciably affect the results, their data were combined for all analyses. To preview, although practice did affect the pattern of span differences in interference, it did so in an unexpected way.

Method

The OSPAN and Stroop task methods for Experiments 3A and 3B were identical to those of Experiment 2 with the following exceptions.

Participants

We tested 75 undergraduates (40 high spans, 35 low spans) in Experiment 3A and 63 undergraduates (28 high spans, 35 low spans) in Experiment 3B, all from Georgia State University and Georgia Institute of Technology, in return for partial fulfillment of a course requirement or for extra credit. Participants were drawn from a larger pool that had completed the OSPAN task approximately 30 min before beginning the Stroop task. These participants were different from those tested in Experiments 1 and 2.

Design

The designs for both Experiment 3A and 3B were 2 × 2 × 2 mixed-model factorials, with span (high, low) as a between-subjects variable, and congruency proportion (0%, 75%) and trial type (congruent, neutral, incongruent) as within-subjects variables.

Procedure

In both experiments, all participants completed the 0% congruent task and then the 75% congruent task. In Experiment 3A, trials were participant-paced, as they were in Experiments 1 and 2. In Experiment 3B, trials were computer-paced: The participant initiated each trial block by pressing a key at the presentation of a yellow *READY?* signal. After a 600-ms blank screen, a white fixation signal (******) appeared for 200 ms, followed immediately by the target letter string that remained until response (or 30 s, whichever came first). The experimenter then keyed in the response as in Experiment 1, and then after a 600-ms delay, the next fixation signal automatically appeared.

Results

Participants

The mean OSPAN scores for the high and low spans, respectively, were 22.98 (*SD* = 3.72, range = 19–34), and 5.66 (*SD* = 1.57, range = 2–8) in Experiment 3A, and 25.61 (*SD* = 5.43, range = 20–38), and 6.23 (*SD* = 2.64, range = 0–9) in Experiment 3B.

Stroop Task

Response Times

Mean response times in combined Experiments 3A and 3B, by span group, congruency proportion, and trial type, are presented in Table 3.

Interference effects. From inspection of the means, low spans appeared to show larger interference effects than did high spans in both the 0% and 75% proportion-congruency conditions. Because a 2 (span group) × 2 (experiment) × 2 (congruency proportion) × 2 (trial type) mixed-model ANOVA indicated that the only effect of experiment was a marginally significant interaction with congruency proportion, $F(1, 134) = 3.25$, $MSE = 3,473.08$, $p = .07$ (with most other Fs < 1), we did not analyze this variable further.

High-span participants responded more quickly than did low spans overall, $F(1, 134) = 6.58$, $MSE = 36,660.95$, and unlike the previous experiments, span group interacted with trial type, indicating significant span differences in response-time interference, $F(1, 134) = 12.74$, $MSE = 3,414.53$ (interference was significant for each span group, with Fs > 310). As the means suggest, the nonsignificant interaction of span, trial type, and congruency ($F < 1$) demonstrated that span differences in interference were significant and comparable in both congruency conditions. Thus, in contrast to Experiments 1 and 2, low spans in Experiments 3A and 3B demonstrated greater interference in latencies than did high spans.

Table 3
Mean Response Latencies (in Milliseconds) and Error Rates in Color Naming, With Standard Deviations, by Working-Memory Span Group, Congruency Proportion, and Trial Type in Combined Experiments 3A and 3B

Participant group	Congruent				Neutral				Incongruent			
	Latency		Error rate		Latency		Error rate		Latency		Error rate	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
0% congruent												
High span[a]					628	78	0.01	0.02	710	105	0.03	0.04
Low span[b]					651	81	0.01	0.02	767	122	0.04	0.04
75% congruent												
High span[a]	614	86	0.01	0.02	653	81	0.01	0.02	781	117	0.09	0.08
Low span[b]	629	93	0.01	0.01	674	90	0.01	0.02	839	148	0.10	0.07

[a] $n = 68$.   [b] $n = 70$.

The only remaining significant results were a main effect of trial type, indicating a significant Stroop effect overall, $F(1, 134) = 601.61$, $MSE = 3,414.53$, and a Congruency Proportion × Trial Type interaction, $F(1, 134) = 73.95$, $MSE = 1,009.21$, indicating greater interference in the 75% than the 0% congruent condition.

Facilitation effects. From inspection of the means in Table 3, high and low spans demonstrated equivalent facilitation effects in the 75% congruent condition, in contrast to the findings from Experiments 1 and 2, and despite the greater power here. A 2 (span) × 2 (trial type) mixed-model ANOVA revealed that congruent trials yielded faster responses than did neutral trials overall, $F(1, 136) = 167.55$, $MSE = 717.95$, but the lack of a Span × Trial Type interaction indicated that WM span did not affect response-time facilitation ($F < 1$).

Errors

Participants' mean color-naming error rates in Experiments 3A and 3B combined, by WM span, congruency proportion, and trial type, are displayed in Table 3.

Interference effects. Inspection of the means suggests that, unlike response times, error rates were not sensitive to span differences in interference in either the 0% or 75% congruent conditions. We conducted a 2 (span) × 2 (experiment) × 2 (congruency proportion) × 2 (trial type) mixed-model ANOVA on error rates, and found significant effects only for trial type, $F(1, 134) = 152.13$, $MSE = 0.003$, indicating a Stroop interference effect, and for the interaction of congruency proportion and trial type, $F(1, 134) = 121.25$, $MSE = 0.001$, indicating greater interference for the 75% congruent condition than for the 0% condition (all other $Fs < 1.05$). Unlike the previous experiments, then, no span differences were seen in error interference in any congruency condition, despite other effects being statistically detectable.

Facilitation effects. As in Experiments 1 and 2, a 2 (span) × 2 (trial type) mixed-model ANOVA indicated that errors in the 75% congruent condition were equivalent, and near floor, on congruent and neutral trials, and the span groups were equivalent in facilitation (or lack thereof; with all $Fs < 1$).

Discussion

Presenting the 0% congruent task before the 75% congruent task produced span differences in response-time interference in both tasks, and it did so in both Experiments 3A and 3B. Thus, congruency-task order appears to have a robust effect on the pattern of span differences in Stroop interference. Low spans showed greater error interference than did high spans in the 75% congruent conditions when they were presented as a first task (Experiments 1 and 2), but low spans showed exaggerated response-time interference in the 75% condition presented after a 0% condition here. The lack of a span difference in error interference for a 75% congruent task that followed a 0% task (despite significant interference overall) suggests that practice in a 0% congruent context reduced the subsequent probability or severity of goal neglect in low-span participants.

In addition, here the response-time interference differences between high and low spans in the 0% congruent tasks were significant, where they had not been in Experiments 1 and 2. As we expected, this may have resulted from the more limited practice

with incongruent trials in Experiments 3A and 3B compared with Experiments 1 and 2. In Experiment 1, the 0% congruent task consisted of twice as many trials as in Experiments 3A and 3B, and in Experiment 2, the 0% task was presented after 288 trials of practice with the 75% congruent task. However, because the combined Experiments 3A and 3B had included more participants than did either Experiment 1 or 2, the significant span differences seen here may reflect improved power rather than the effects of practice. To test this idea we conducted a 2 (span) × 2 (trial type) × 2 (experiment) ANOVA on the 0% congruent response-time data from Experiments 1 and 2 together. The key finding from this analysis was that, with the improvement in power, low spans' interference effects in the 0% condition were significantly larger than high spans' effects, $F(1, 127) = 4.90$, $MSE = 1,881.43$, and this span difference in interference did not interact with experiment ($F < 1$). Thus, it appears that whether the 0% congruent task was completed as a first task or as a second task, low spans showed larger response-time interference than did high spans across experiments; however, the effect is not especially robust, and so large samples are required to detect it.

With respect to the 75% congruent task, span differences in interference were evident in Experiments 3A and 3B, but in response times rather than in errors. Practice on a 0% condition did not eliminate span differences in interference on a subsequent 75% condition, but it did shift the effect from errors to response times. To ensure that the lack of a response-time difference between high and low spans in the 75% congruent conditions from Experiments 1 and 2 had not been due to insufficient power, we conducted a 2 (span) × 2 (trial type) × 2 (experiment) ANOVA on the combined 75% congruent data from both experiments. Unlike the analysis of the 0% congruent data reported above, here the increase in power did not reveal significant span differences. Low spans' 75% task interference effects in latency were not larger than high spans', $F(1, 128) = 1.78$, $MSE = 3,632.56$, $p = .18$, and this lack of a span difference in interference did not interact with experiment ($F < 1$). Thus, when the 75% congruent task was encountered first, WM span predicted error interference only, and when the 75% task was encountered after the 0% congruent task, WM span predicted latency interference only.

Thus, Experiments 1–3B indicated that the locus of WM-span-related differences in Stroop interference depends not only on the congruency proportions within the task, but also on the order in which congruency proportions are encountered. Experiments 1 and 2 demonstrated that 75% congruent tasks that occur as a first task produce span differences in error interference, whereas Experiments 3A and 3B both showed that these same tasks occurring subsequent to 0% congruent tasks produce span differences in response-time interference. Although 0% congruent tasks produced small and inconsistent span effects in response-time interference within individual experiments, combined-experiment analyses reported above indicated that low spans showed greater response-time interference in these tasks, regardless of task order. For a summary of results across Experiments 1–3B, see Figures 1 (response-time interference) and 2 (error-rate interference).
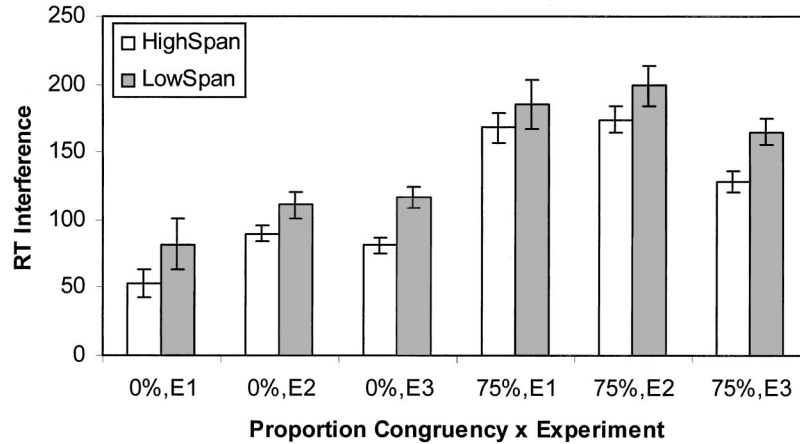
Figure 1. Mean response-time interference effects (in milliseconds) for high- and low-span participants, by congruency proportion (0%, 75%) and experiment (1–3). Interference effects were calculated by subtracting participants' mean neutral-trial latency from incongruent-trial latency. Error bars depict standard errors of the means. E1 = Experiment 1; E2 = Experiment 2; E3 = Experiment 3; RT = response time.

EXPERIMENT 4

We conducted Experiment 4 to extend the generality of the findings summarized above. Our primary interest was in whether the differences we observed between the 0% and 75% congruent tasks, and their order effects, were tied to these specific congruency proportions. In particular, we tested whether there was something unique about a 0% congruent Stroop task, that is, whether the inclusion of any congruent trials would affect interference and span differences therein, or whether any low-congruency proportion would lead to results similar to those we had found. One could imagine that, with some minimal proportion of congruent trials in the design (e.g., 10% or 20%), some low-span participants might periodically fail to maintain the task goal. If this were true, the 0% congruent condition, which presents no congruent trials, would be a special case.

We therefore tested high- and low-span participants in high- and low-congruency Stroop tasks, but here the percentages of congruent trials were 80% and 20%, respectively. We also manipulated congruency order within this single experiment, with half of the participants encountering the 20% congruent task first, to make contact with the full picture of results we obtained in Experiments 1–3B (see Figures 1 and 2). Our predictions for the 80% congruent condition were straightforward—like the 75% condition, we expected span differences in error interference when it was encountered first, and in response-time interference when it followed the 20% task. For the 20% congruent task, we predicted findings to follow those of the 0% congruent task, with span differences in response-time interference regardless of task order.

However, we also thought it possible that the inclusion of congruent trials in the 20% congruent task might lead to some
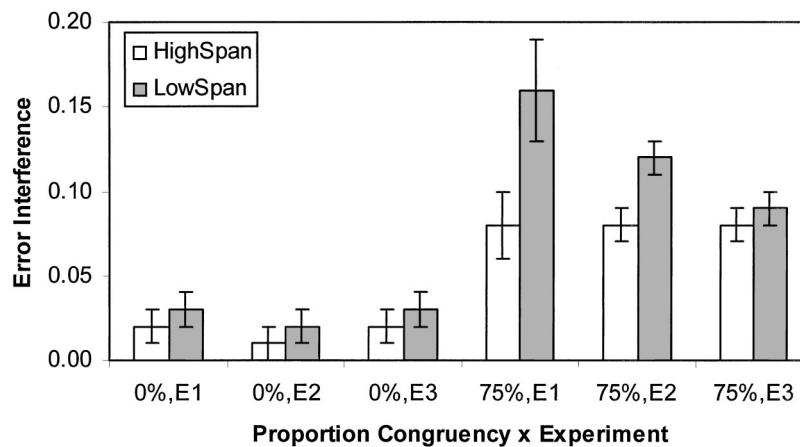

Figure 2. Mean error-rate interference effects for high- and low-span participants, by congruency proportion (0%, 75%) and experiment (1–3). Interference effects were calculated by subtracting participants' mean neutral-trial error rate from incongruent-trial error rate. Error bars depict standard errors of the means. E1 = Experiment 1; E2 = Experiment 2; E3 = Experiment 3.

discrepancies with the 0% data. For example, the 20% task might lead at least some low-span participants to intermittently neglect the goal, leading to span differences in error interference like those we observed in the 75% congruent task. Or, more subtly, the 20% task might fail to establish the task set as effectively as did the 0% task. Even in the absence of detectable error effects within the 20% congruent task, the subsequent 80% task might be more susceptible to goal-maintenance failures than it would be after following the 0% task. Thus, span differences in errors, rather than in response time, might arise in the subsequent 80% congruent task.

To measure responses to a suitable number of trials per condition while still keeping the length of the experiment manageable for participants, we eliminated the neutral trials here. Thus, in Experiment 4, we calculated all interference effects as the difference between congruent and incongruent trials, rather than between neutral and incongruent trials.

Method

The OSPAN and Stroop task methods for Experiment 4 were identical to those of the previous experiments with the following exceptions.

Participants

We tested 274 participants (136 high spans, 138 low spans) in Experiment 4. These participants were either undergraduates from the University of North Carolina at Greensboro or Georgia Institute of Technology, completing the experiment in return for partial fulfillment of a course requirement or for extra credit, or young-adult volunteers from the Atlanta, Georgia, metro area, completing the experiment for pay. Participants were drawn from a larger pool that had completed the OSPAN task between 1 and 240 days before beginning the Stroop task. None of these participants had been tested in Experiments 1–3B.

Design

The design for Experiment 4 was a 2 × 2 × 2 × 2 mixed-model factorial, with span (high, low) and congruency-proportion order (20%-first, 80%-first) as between-subjects variables, and congruency proportion (20%, 80%) and trial type (congruent, incongruent) as within-subjects variables.

Apparatus and Materials

The stimulus list for the 20% congruent task consisted of 180 total trials, with 36 critical congruent trials, 36 critical incongruent trials, and 108 filler incongruent trials. The list for the 80% congruent task consisted of 180 total trials, with 36 critical congruent trials, 36 critical incongruent trials, and 108 filler congruent trials. Each word and color was equally represented across filler and critical trials, which were identical to one another except for their coding in the computer program that presented the stimuli and recorded responses.

Procedure

All participants in Experiment 4 completed a practice color-naming block identical to those in Experiments 1–3, followed by the 20% and 80% congruent blocks. Participants were randomly assigned to a congruency-task order, completing the 20% congruent condition before the 80% condition or vice versa. Trials were participant-paced, as they were in Experiments 1, 2, and 3A, but no accuracy feedback was provided, as in Experiment 1. On each trial, a ready signal remained on screen until the participant pressed the keyboard's space bar, followed by a 750-ms blank screen. A centered, gray fixation signal (i.e., a plus sign) then appeared for 250 ms, followed immediately by the target word that remained in the center of the screen until response. The experimenter then logged the response through key-press, and then after a 250-ms delay, the next ready signal appeared.

Results

Participants

The mean OSPAN scores for the high and low spans, respectively, were 24.28 (SD = 5.09, range = 18–38) and 5.25 (SD = 1.99, range = 0–9) in the 20%-first condition, and 23.80 (SD = 5.38, range = 18–42) and 5.37 (SD = 1.75, range = 0–9) in the 80%-first condition.

Stroop Task

Response Times

Mean response times in Experiment 4, by span group, congruency-proportion order, congruency proportion, and trial type, are presented in Table 4. From inspection of the means, the 80% congruent tasks produced larger interference effects than did the 20% congruent tasks overall. However, low spans appeared to show greater interference than high spans in all conditions, that is, in both congruency proportions and in both proportion orders.

A 2 (span group) × 2 (congruency-proportion order) × 2 (congruency proportion) × 2 (trial type) mixed-model ANOVA confirmed these observations. Consistent with our findings from 0% and 75% tasks, Stroop interference effects were larger in the 80% congruent conditions than in the 20% congruent conditions, $F(1, 270) = 670.28$, $MSE = 1,419.43$. In addition, the significant span difference in latency, $F(1, 270) = 4.01$, $MSE = 38,533.42$, was moderated by significant interactions of span with congruency proportion, $F(1, 270) = 6.31$, $MSE = 4,642.86$, and with trial type, $F(1, 270) = 13.49$, $MSE = 3,517.11$, the latter indicating greater interference for low spans than for high spans overall (interference was significant for both span groups, with $Fs > 722$). However, the nonsignificant interactions of span, congruency proportion, and trial type, $F(1, 270) = 2.23$, $MSE = 3,161.81$, $p = .14$, of span, congruency-proportion order, and trial type, $F(1, 270) = 1.42$, $MSE = 3,517.11$, $p = .23$, and of span, congruency-proportion order, congruency proportion, and trial type ($F < 1$) indicated that span differences in interference did not differ as a function of congruency proportion, or the order in which congruency proportions were encountered. Thus, span differences in latency interference were seen here in both 20% and 80% congruent conditions, for both 20%-first and 80%-first participants. (For archival purposes, the full ANOVA results for response times are presented in Appendix A.)

Errors

Participants' mean proportion of color-naming errors in Experiment 4, by WM span, congruency order, congruency proportion, and trial type, are displayed in Table 4. From inspection of the means, 80% congruent conditions produced greater interference effects than did 20% conditions, and low spans showed larger interference effects than did high spans in only these 80% con-

Table 4
Mean Response Latencies (in Milliseconds) and Error Rates in Color Naming, With Standard Deviations, by Working-Memory Span Group, Congruency-Proportion Order, Congruency Proportion, and Trial Type in Experiment 4

Participant group	Trial type							
	Congruent				Incongruent			
	Latency		Error rate		Latency		Error rate	
	M	SD	M	SD	M	SD	M	SD
20% first								
20% congruent								
High span[a]	663	112	0.04	0.04	735	125	0.06	0.05
Low span[b]	700	124	0.04	0.04	782	125	0.07	0.07
80% congruent								
High span[a]	619	86	0.03	0.05	786	105	0.08	0.07
Low span[b]	632	89	0.03	0.04	824	107	0.13	0.13
80% first								
20% congruent								
High span[c]	646	94	0.03	0.03	710	103	0.05	0.04
Low span[d]	657	107	0.04	0.04	750	132	0.07	0.06
80% congruent								
High span[c]	643	106	0.03	0.04	834	119	0.11	0.08
Low span[d]	624	79	0.02	0.03	856	122	0.16	0.10

[a] $n = 72$. [b] $n = 68$. [c] $n = 64$. [d] $n = 70$.

gruent conditions. However, the span difference in 80% congruent errors appeared to be present in both task orders (i.e., for both 20%-first and 80%-first participants).

These impressions were confirmed by a 2 (span) × 2 (proportion-congruency order) × 2 (proportion congruency) × 2 (trial type) mixed-model ANOVA on error rates. Stroop interference was significant overall, $F(1, 270) = 224.87$, $MSE = 0.004$, and it was larger for the 80% congruent condition than for the 20% condition, $F(1, 270) = 132.15$, $MSE = 0.002$. Low spans committed more errors than did high spans, $F(1, 270) = 11.87$, $MSE = 0.007$, but this span effect was modified by a significant interaction of span and trial type, $F(1, 270) = 16.32$, $MSE = 0.004$, and a significant interaction of span, trial type, and congruency proportion, $F(1, 270) = 16.76$, $MSE = 0.002$. Thus, as expected, span differences in error interference were larger in the 80% congruent condition than in the 20% congruent condition (span differences in interference in the 80% condition were significant for both high- and low-span groups, with Fs > 99): A follow-up analysis indicated that, although the Stroop effect was significant in the 20% congruent condition, $F(1, 270) = 57.52$, $MSE = 0.002$, it did not differ in magnitude between high- and low-span participants, $F(1, 270) = 1.63$, $MSE = 0.002$, $p = .20$.

In contrast to our prior findings, but obvious from inspection of the means, span differences in the high-congruency condition did not depend on the presentation order of congruency conditions here. Span did not interact significantly with congruency-proportion order and trial type, $F(1, 270) = 1.47$, $MSE = 0.002$, $p = .23$, or with congruency-proportion order, congruency-proportion, and trial type ($F < 1$). Thus, low spans' exaggerated vulnerability to error interference in 80% congruent conditions was present whether the 80% task was completed first or following

the 20% task. (For archival purposes, the full ANOVA results for error rates are presented in Appendix B.)

As a final analysis of errors, we examined their latencies. It should be noted that if a majority of errors in the 80% congruent condition were the result of goal neglect, then they should have had short latencies; that is, goal neglect should lead to overt word reading and therefore faster responses than accurate color naming. Experiment 4 was the first of our experiments to have a large enough sample for such an analysis. First, we identified participants who committed at least six errors (a 16.7% error rate) in the incongruent condition from the 80% congruent task, regardless of task order, to obtain minimally stable mean latencies for each participant. We then divided the errors into two types: "clean" errors represented trials on which the participant clearly said a color–word aloud that did not match the color of the stimulus, and "ambiguous" errors represented trials on which the participant stuttered, slurred words together, said a noncolor word, or made any other kind of incorrect or ambiguous response.

Our principal concern was with clean errors, as these should most clearly represent word-reading responses reflecting goal neglect. We hypothesized that, not only would more low spans reach the error criterion than high spans (as suggested by the error analyses above), but of these errors a higher proportion would be clean for low spans than for high spans; that is, among error-prone individuals, low spans should make disproportionately more word-reading errors than other kinds of errors. With respect to latencies, we predicted that clean errors should be considerably faster than ambiguous errors and correct responses on incongruent trials. In contrast, clean-error latencies should be similar to congruent-trial latencies, where word reading and color naming may both contribute to mean response time (see MacLeod, 1998). Finally, we

predicted that span groups should not differ significantly in clean-error latency. Once a person neglects the goal and reads the word, there is no reason to think that WM span should predict the speed of word reading.

Our hypotheses were confirmed. Of importance, only 22 high spans met our inclusion criterion of ≥ 6 incongruent errors in the 80% congruent condition, whereas 47 low spans did, which is obviously more than twice the number of high spans. Moreover, even among this highly error-prone group, low spans had a larger error interference effect than did high spans ($Ms = +.23$ and $+.17$, respectively), $F(1, 67) = 4.40$, $MSE = 0.01$. These numbers provide further dramatic indication of the robustness of span difference in error interference. With respect to the kinds of errors committed by error-prone participants, clean errors made up a mean of 58% of high spans' errors and 68% of low spans' errors. Although this span difference was in the predicted direction, with low spans making proportionally more clean errors, it did not reach conventional significance, $F(1, 67) = 2.00$, $MSE = 0.07$, $p = .16$.

Error-prone participants' response times on incongruent errors and on correct congruent and incongruent trials in the 80% congruent condition are presented in Table 5. A 2 (span group) × 2 (error type) mixed-model ANOVA indicated that, as expected, clean errors were committed much more quickly than ambiguous errors, $F(1, 57) = 58.46$, $MSE = 792,484.57$. Moreover, high- and low-span participants had equivalent error latencies, and span group did not interact with error type (both $Fs < 1$). Thus, participants who committed goal-neglect errors did so quickly, regardless of WM capacity. We also compared response times on clean errors with those on accurate congruent and incongruent trials, expecting them to be similar to congruent trials and substantially shorter than incongruent trials. A 2 (span group) × 3 (response type) ANOVA indicated no response-time differences between span groups, nor an interaction of span and response type (both $Fs < 1$), but a significant overall difference among response types, $F(2, 134) = 144.17$, $MSE = 6,352.46$. As expected, clean-error latencies were not different from congruent latencies, $F(1, 67) = 2.31$, $MSE = 6,688.63$, $p = .13$, but much shorter than those on accurate incongruent trials, $F(1, 67) = 131.04$, $MSE = 9,318.31$.

Discussion

Experiment 4, which used 20% and 80% congruent conditions, demonstrated both similarities and differences to Experiments 1–3B, which had used 0% and 75% congruent conditions. In our prior experiments, low-congruent conditions yielded small but significant span differences in response-time interference, regardless of whether they were presented before or after high-congruent conditions (i.e., in the collapsed data from Experiments 1 and 2, and from Experiments 3A and 3B). In contrast, high-congruent conditions produced span differences in error interference when completed as a first task (Experiments 1 and 2), while producing span differences in latency interference when completed after a low-congruent task (Experiments 3A and 3B).

Here, the data from the low-congruent condition, itself, closely matched our prior findings, despite the fact that we changed the congruency proportion rather dramatically, from 0% to 20%. Like the 0% task across our experiments, the 20% congruent task yielded significant span differences in response-time interference, regardless of task order. Unexpectedly, it was the 80% congruent condition, modified only from a 75% condition, which produced anomalous results. In short, the 80% congruent task produced span differences in all manifestations of interference, rather than selectively eliciting it in errors in high-congruency-first conditions and in latencies in high-congruency-last conditions. That is, low spans produced greater interference than did high spans in both latencies and error rates, for both the 20%-first and 80%-first groups. How might these discrepant results be explained?

First, with respect to the significant span differences in response-time interference, regardless of congruency order, the discrepancy is more apparent than real. Although Experiments 1 and 2 failed to find a significant span difference in response-time interference in a 75% congruent task encountered as a first task (instead, finding differences in error rates), interference was calculated there against a neutral-stimulus baseline (see Figure 1). In Experiment 4, however, it was calculated against a congruent-trial baseline because we eliminated the neutral trials there. When we compared the response-time interference effect in Experiment 4 with recalculated effects from Experiments 1–3B, now using the congruent-trial baseline (see Figure 3), the pattern across experiments was consistent: Low spans showed greater response-time interference than did high spans against a congruent baseline. The response-time data from Experiment 4 thus provide no real inconsistency with Experiments 1 and 2.

However, these data do pose the question of whether neutral or congruent trials are the proper baseline from which to assess

Table 5

Mean Response Latencies (in Milliseconds), With Standard Deviations, for "Clean" Errors, "Ambiguous" Errors, Correct Congruent Responses, and Correct Incongruent Responses in the 80% Congruent Condition From Experiment 4, for High and Low Working-Memory-Span Participants Who Met the Error Criterion of ≥ 6 Incongruent-Trial Errors

| | Response type | | | | | | | |
| | Clean errors | | Ambiguous errors | | Correct congruent | | Correct incongruent | |
Participant group	M	SD	M	SD	M	SD	M	SD
High span[a]	628	142	1,902	1,426	614	98	826	133
Low span[b]	622	144	1,976	1,211	591	90	827	140

[a] $n = 22$. [b] $n = 47$.
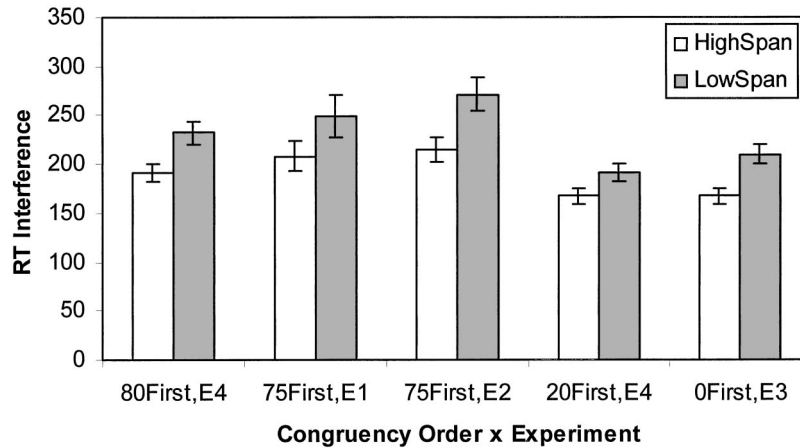
Figure 3. Mean response-time interference effects (in milliseconds) for high- and low-span participants, in high-congruency conditions, by congruency-task order and experiment (1–4). Interference effects were recalculated by subtracting participants' mean congruent-trial latency from incongruent-trial latency. Error bars depict standard errors of the means. 80First = participants completing the 80% congruent task first; 75First = participants completing the 75% congruent task first; 20First = participants completing the 20% congruent task first; 0First = participants completing the 0% congruent task first; E1 = Experiment 1; E2 = Experiment 2; E3 = Experiment 3; E4 = Experiment 4; RT = response time.

Stroop interference (see Lindsay & Jacoby, 1994; MacLeod, 1991). We believe our neutral baseline and the conclusions we draw from it in Experiments 1–3B are defensible because the pattern of span differences in response-time interference and facilitation across experiments cannot be attributed to differences in the neutral-baseline latencies. That is, in Experiments 1 and 2, where span groups differed in facilitation but not interference, span differences in neutral response times occurred only in Experiment 2. In Experiments 3A and 3B, where span groups differed only in interference and not in facilitation, span differences in neutral response times occurred only in Experiment 3A. Thus, span differences in response-time interference and facilitation, as measured against a neutral-stimulus baseline, varied systematically with congruency-condition order, but not with neutral-baseline differences.

Returning to Experiment 4, the real discrepancy it produced with Experiments 1–3B is the significant span difference in error interference in the 80% congruent conditions, regardless of task order. Of course, we did expect span differences in errors when the 80% condition was encountered first, replicating the robust findings from Experiments 1 and 2, which presented a 75% congruent condition first. However, in Experiment 4, the 80% task that followed the 20% task also produced findings indicative of overt word reading (i.e., failed goal maintenance and subsequent errors) in low-span participants. In contrast, in Experiments 3A and 3B, a 75% congruent task that followed a 0% task elicited span differences only in response-time interference, not in errors, indicating less difficulty in goal maintenance for the low spans there.

We believe that this difference between experiments is meaningful, because there was no hint of a span difference in error interference in either Experiments 3A or 3B, and the effect here in Experiment 4 was substantial. As foreshadowed in the introduction to this experiment, our provisional explanation for the effect is that the preceding 20% congruent task did not as firmly establish the

"ignore the word" task set as did the 0% task used in the other experiments. The presence of congruent trials did not lead to actual, or overt, error/goal-neglect differences between high- and low-span participants in the 20% congruent condition, itself. However, it may have subtly interfered with low spans' ability to overlearn the task set over the course of 180 trials. Whereas a 0% congruent task may help low-span participants transfer the "ignore the word" goal to a subsequent high-congruency task, a 20% task may not. Such an explanation is sensible but speculative, and so further work will be required to confirm it.

Finally, Experiment 4 allowed us to examine incongruent-trial errors from high-congruency conditions more closely than in prior experiments, because of the relatively large *N* allowed by the similarity of the 20%-first and 80%-first results. We found that for a subsample of participants with high error rates in the 80% congruent condition, low spans were over twice as likely to reach our "error prone" criterion than were high spans, and furthermore, error-prone low spans showed greater error interference than did error-prone high spans. For both span groups, most errors were "clean," that is, they represented a clear color-naming response that did not match the target color, and so most likely reflected word reading (low spans' errors were slightly more skewed toward clean errors than were high spans'). Also regardless of WM capacity, clean errors were committed very quickly, further supporting the proposal that they reflected overt word reading that resulted from failed goal maintenance.

GENERAL DISCUSSION

In five experiments we investigated the relation between WM capacity, as measured by the OSPAN task, and Stroop interference. Experiments 1, 2, and 4 demonstrated in a straightforward manner that WM span is related to goal maintenance and neglect in the Stroop task. Low spans committed many more errors on

incongruent trials than did high spans when congruent trials made up 75% or 80% of the Stroop trials (and Experiment 4 indicated that most of these errors were committed very quickly). Without the external goal reinforcement provided by frequent incongruent trials, low spans were more likely to lose access to the goal and respond according to habit, by reading the word in error. In contrast, in the 0% and 20% congruent tasks from all of our experiments, low spans showed somewhat larger response-time interference effects than did high spans, despite changes in the statistical significance of span difference across individual experiments. Here, low spans did not show evidence of overt word reading, suggesting that the presentation of mostly incongruent trials helped participants to keep track of the task goal of ignoring the word information. Moreover, in the 75% congruent task in Experiments 3A and 3B, where it followed the 0% task, span differences in interference appeared in latencies, and not in errors. Thus, in 0% congruent task contexts that provided substantial goal reinforcement by requiring color naming on every trial, and in 75% congruent contexts that followed these 0% tasks, low spans did not often lose complete track of the task goal. Instead, they were slower to respond accurately, according to goal, than were high spans.

Our results clearly demonstrate that span differences in error versus latency interference were strongly influenced by the current and previous task contexts. In particular, when the prior 0% congruent context reinforced the goal of the task by repeatedly presenting incongruent trials, it appeared to establish a task set that transferred to the subsequent 75% congruent task. Even though the subsequent 75% task did not externally reinforce the task goal, it yielded span differences in response-time interference, rather than in errors, just as in the 0% tasks. Viewing these transfer effects another way, high spans' error interference effects in the 75% congruent task were strikingly stable (and low) across experiments, regardless of task order. In contrast, low spans' effects varied systematically by whether the 75% task followed the 0% task. Low spans showed large error effects when the 75% congruent condition was encountered first, and small error effects when it was presented after the 0% congruent condition. At least with respect to accuracy, then, low spans were more affected by set transfer than were high spans, responding in the 75% task as they had in the previous 0% task. This pattern is reminiscent of our own antisaccade findings that low spans had more difficulty than high spans switching set from one blocked task to another (Kane et al., 2001).[4]

In contrast to the 75% congruent task, the 0% and 20% tasks were not particularly sensitive to task set, producing latency interference of about the same modest magnitude across experiments, regardless of task order. This differential sensitivity of low- and high-congruency tasks to set transfer was first reported by Lowe and Mitterer (1982, Experiment 3). They tested two groups of participants who differed in the progression of congruency conditions encountered across six sessions, with one group exposed to the progression 0%, 50%, 75%, 75%, 50%, 0% and the other exposed to the progression 75%, 50%, 0%, 0%, 50%, 75%. For the group starting with the 0% congruent task, response-time interference was stable across sessions, with effects ranging from 20 ms to 32 ms. As in our experiments, then, the 0% task established a strong set that affected subsequent tasks, effectively reducing subsequent effects of congruency over sessions. For the group

starting with the 75% congruent task, in contrast, interference varied widely over sessions. Effects ranged from 15 ms to 47 ms, and the magnitude of interference was closely tied to the proportion of congruent trials in the current session. Moreover, the interference effects seen in the 75% tasks differed substantially between the task-order groups ($M = 20$ ms for the 0%-first group and $M = 47$ ms for the 75%-first group). Similar to our findings, then, the 75% congruent task yielded smaller effects following the 0% task than preceding it.

Our broad conclusions about set-transfer effects must be tempered, however, by our findings from Experiment 4, which involved a 20% congruent task rather than a 0% task. Here, when the 80% task followed the 20% task, it produced span differences in error interference effects, just as in the 80% and 75% congruent tasks that appeared as first tasks. As we suggested in our discussion of Experiment 4, this finding suggests that the presence of even a relatively small proportion of congruent trials may have subtle effects on the establishment and transfer of the task set. Whereas the 0% congruent tasks in Experiments 3A and 3B appear to have established a set that transferred to the subsequent 75% task, making goal neglect less likely there, the 20% task does not seem to have done so. Given that the effects of the 0%-to-75% sequence were replicated across experiments 3A and 3B, we are confident that they are real, and so further work will be required to establish whether the 20% congruent effects are similarly reliable and important.

Dual Mechanisms of Stroop Interference?

WM capacity and task context interact to moderate the strength and the locus of Stroop interference. Moreover, we suggest that the specific patterns of span effects observed here suggest that Stroop interference has two causes, as we described in the introduction. One mechanism is a rather consistently time-consuming process of competition resolution, which is reflected primarily in latencies and prevailing in contexts that minimize the need for goal maintenance. The other, additional mechanism is a more transient failure of goal maintenance, reflected most obviously in accuracy and prevailing in contexts that maximize the need for goal maintenance. By our view, the former competition resolution mechanism can only be properly engaged when the goal to do so is actively maintained, or is reinforced by the task context (see also Cohen et al., 1990). Although our data suggest that both mechanisms rely on WM capacity, the congruency context determines the relative influence of the goal-maintenance mechanism.

Specifically, we propose that low spans have a relatively enduring deficit in the competition-resolution mechanism, independent of goal maintenance. This deficit was largely responsible for the modest span differences in response-time interference in 0% congruent tasks, 20% tasks, and in 75% tasks that followed 0% tasks. Low spans were not more likely to commit word-naming errors in these conditions than were high spans, suggesting that they were

---

[4] Although we now have found evidence for WM-related differences in set switching between blocked conditions in both the antisaccade and Stroop tasks, other research suggests that WM capacity may be unrelated to trial-to-trial costs that are measured in currently popular task-switching paradigms (Miyake et al., 2000; Oberauer, Süß, Wilhelm, & Wittmann, in press).

not as often neglecting the goal of the task altogether. However, low spans could not resolve the conflict between word and color responses as quickly as could high spans. In contrast, the WM differences in error interference in 75% and 80% congruent tasks that were not preceded by 0% tasks resulted largely from momentary failures of goal maintenance in low spans. Here, low spans were more likely than high spans to name the words, in error, suggesting that low spans periodically lost the goal from active memory to ignore the word.[5]

Further evidence for the dual nature of interference is also suggested by our facilitation data. Only Experiments 1 and 2, where the 75% congruent conditions did not follow 0% conditions, produced span differences in both error interference and response-time facilitation. That is, only in contexts where low spans were induced to neglect the task goal and overtly read the words on incongruent trials were low spans also more likely to respond faster than high spans on congruent trials. To the extent that response-time facilitation reflects participants actually reading the word aloud on some proportion of congruent trials (MacLeod & MacDonald, 2000b; for a detailed discussion, see below), this finding suggests that both error interference and latency facilitation arise from the failure of the goal-neglect mechanism of Stroop interference.

Although our findings concerning span, context, and task set may not conclusively point to dual Stroop mechanisms, below we review converging data in the literature for this idea. We already discussed some of the relevant findings in our introduction, such as the fact that schizophrenics show low-span-like patterns of Stroop deficits that are linked to congruency contexts. That is, schizophrenics show exaggerated error interference and response-time facilitation effects in high-congruency contexts, indicative of goal neglect, and less exaggerated response-time interference in low-congruency contexts (e.g., Cohen et al., 1999; Perlstein et al., 1998). Analyses of Stroop latency distributions from normal adults are also suggestive. When the distributions of incongruent and neutral response times are compared, incongruent trials produce an overall shift in the normal part of the distribution, as well as an increase in positive skew resulting from a small proportion of very slow responses (e.g., Heathcote et al., 1991; Stroop, 1935). We argue that the overall shift reflects the regular process of resolving response competition on all incongruent trials. In contrast, the increased skew reflects some exaggerated response times resulting from periodic goal neglect. That is, this subset of very slow responses in the tail of the distribution represents trials on which the goal was lost but then recovered before an overt error was committed, thus leading to much slower responses than on trials on which the goal is maintained and implemented immediately.

If our arguments are correct, then variables thought to affect goal neglect should have a selective effect on skew, and not on the overall distribution shift. Indeed, skew on incongruent trials is selectively increased for participant populations prone to lapses of intention, such as in older adults (Spieler et al., 1996), and it is increased in young adults when trials are presented at a very slow pace, thus encouraging mind-wandering and goal neglect (De Jong et al., 1999). In contrast, skew is selectively decreased when trials are rapidly paced to limit lapses (De Jong et al., 1999), and also when color and word dimensions of Stroop stimuli are spatially separated (Spieler, Balota, & Faust, 2000), a manipulation that

should reduce the likelihood that participants would read the words overtly and lose track of the color-naming goal.

Additionally, evidence from cognitive neuroscience techniques suggests separable competition and goal-neglect mechanisms of Stroop interference. For example, research by West and colleagues has demonstrated that different event-related potentials from frontal cortex correspond to different manifestations of Stroop interference. Specifically, West and Alain (1999, 2000a) found that an increased negativity, probably generated by anterior cingulate cortex, begins between 400 ms and 500 ms after the onset of an incongruent stimulus (the "N450"; see also Liotti, Woldorff, Perez, & Mayberg, 2000). Given the timing of this wave, and the putative role of anterior cingulate in response selection and competition (e.g., Posner & Peterson, 1990; Posner & Raichle, 1994), we interpret the N450 as involved in the competition-resolution mechanism of Stroop interference. In contrast, West and Alain (2000b) discovered that a slow wave, originating in polar or dorsolateral prefrontal cortex (dPFC), and beginning 400–800 ms before a Stroop stimulus is presented, discriminated errors from correct responses on the upcoming stimulus. Moreover, this error-predicting wave was significantly larger in mostly congruent Stroop tasks than in mostly incongruent Stroop tasks. Given the association of this wave with errors, its sensitivity to congruency, and its likely source in a brain region that supports goal maintenance (e.g., Braver & Cohen, 2000; Duncan, 1995), we suggest that it is a marker of the goal-neglect mechanism of Stroop interference (see also West, 2001).

These conclusions are further bolstered by recent event-related functional magnetic resonance imaging data from a Stroop task in which the instructions to read the word or name the color were presented 11 s before each stimulus appeared (MacDonald, Cohen, Stenger, & Carter, 2000). Half the trials were word-reading trials and half were color-naming trials, with 50% congruent trials in each task. Over the 11-s delay, dPFC activity increased on color-naming trials only; no increase was seen on the more automatic, word-reading trials. Moreover, the correlation between delay-period dPFC activity and Stroop interference in color naming was strongly negative ($r = -.63$). These data indicate that dPFC acts to reduce Stroop interference by maintaining the "name the color/ ignore the word" demands in an active state. Despite the fact that many imaging studies indicate anterior cingulate activation in Stroop tasks (for a review, see MacLeod & MacDonald, 2000b), here there was no anterior cingulate activity related to task instructions over the cue-to-target delay. Instead, anterior cingulate activity was closely linked to the onset of incongruent stimuli, and it was positively correlated with interference ($r = .38$), with greater

---

[5] Alert readers will note that our view predicts that a consistently deficient competition-resolution mechanism in low spans should produce span differences in response-time interference in all contexts, including those in which a 75% congruent task is performed first. We speculate that span differences in competition resolution are present in these contexts (indeed, low spans show numerically larger interference effects in response-times than do high spans), but these span differences are not statistically detectable. Because low spans lose a substantial number of incongruent trials to errors in these contexts, the deleted trials cannot contribute to their response-time distributions, and so span differences in response time are masked. For a similar situation with Alzheimer's patients, see Spieler et al. (1996).

activity associated with larger interference effects. These findings indicate that anterior cingulate may not have a role in combating Stroop interference, but rather it may signal or recognize that conflict is present.

The findings discussed above, regarding response-time distributions and correlations between brain activity and interference, indicate that evidence for failed goal maintenance may sometimes be obtained even in the absence of frequent overt errors. Goal neglect may be indicated by positive skew in incongruent latency distributions in some task contexts, or by dPFC activity before the onset of Stroop stimuli. This is useful knowledge at a pragmatic level. Researchers working with participant groups who may not show the substantial error effects demonstrated by low WM individuals, schizophrenics, or Alzheimer's patients may still be able to evaluate hypotheses about goal maintenance and neglect by examining more subtle behavior or neural effects in Stroop latencies.[6]

In addition, the findings are important at a theoretical level, because they suggest that failures in goal maintenance may not be all-or-none. That is, participants may periodically lose complete access to the goal and read the word instead of naming the color, but in addition, they seem able to incrementally boost the activation of the task goal (as in the imaging studies), and also to suffer intermediate levels of goal neglect—monitoring for and catching errors before overt responses are made, and producing a small number of very slow but correct responses (as in the latency-distribution studies). Perhaps goals are not maintained in an all-or-none fashion, but rather are activated in a more graded fashion, below which overt errors are most likely to occur but above which some probability of retrieving the goal before error commission is retained. This idea seems consistent with the graded activation levels possible for the "task demand" goal units in the Cohen model (e.g., Cohen & Servan-Schreiber, 1992) and also with the cumulative gamma function used to model the attentional-control mechanism in the Stroop counter model (Jacoby, McElree, & Trainham, 1999; Trainham, Lindsay, & Jacoby, 1997); it is also analogous to the variety of "capture" errors and action slips regularly seen in the lab and in life, which are sometimes corrected before being completed (see Reason, 1990). The implications of this view for our data are that WM capacity and task context may mediate both the probability and the degree of active goal maintenance in the Stroop task. Some goal neglect probably occurs in all contexts and for all participants, but low WM capacity and high-congruency contexts elicit more episodes of goal neglect, less consistently high activation of goal states, or both, than do high WM capacity and low-congruency contexts.

Mechanisms of Stroop Facilitation

MacLeod (1998; MacLeod & MacDonald, 2000a, 2000b) has argued that Stroop interference and facilitation are independent phenomena, despite the fact that virtually every theory and model of the Stroop effect assumes that they are but two sides of the same coin. That is, interference is assumed to reflect competition between color and word dimensions on incongruent trials, whereas facilitation is assumed to reflect a convergence of the two dimensions on congruent trials. MacLeod, however, pointed out several empirical dissociations between facilitation and interference effects in the literature, and in new data, that challenge this idea (but

see Jacoby et al., 1999; Lindsay & Jacoby, 1994). Instead, MacLeod argued that responding on congruent trials reflects a mixture of slower, color-naming responses and faster, word-naming responses. Because "incorrect" word naming and "correct" color naming cannot be discriminated on congruent trials, the presence of fast word-naming responses in the distribution decreases the average latency. To restate this view in our terms, facilitation arises through goal neglect, as participants periodically read the distractor word aloud.

We believe that our data support MacLeod's (1998; MacLeod & MacDonald, 2000a, 2000b) "inadvertent reading" hypothesis. As already discussed, WM span predicted response-time facilitation only in Experiments 1 and 2, and not in Experiments 3A and 3B (see Figure 4). In Experiments 1 and 2, the 75% congruent task was performed first, and low spans showed evidence of increased goal neglect by committing errors on incongruent trials. That is, low spans only showed greater response-time facilitation than high spans when they were also more likely to neglect the goal and read words overtly. Moreover, low spans' facilitation effects were largest in Experiments 1 and 2, where their error rates on incongruent trials were also highest. High spans' facilitation effects, in contrast, remained constant across experiments, as did their incongruent error rates.

Correlational analyses of our data, independent of WM span, also support a connection between goal neglect and facilitation. If losing access to goal information—and subsequent word reading—leads to response-time facilitation, then response-time facilitation and *error* interference should be significantly correlated. Indeed, this is the pattern that emerges from our 75% congruent task data, which are presented in Table 6. Response-time facilitation correlated positively with error interference in Experiments 1, 2, 3A and 3B, with rs between .34 and .45. As in other reports (e.g., Lindsay & Jacoby, 1994; Vanayan, 1992, cited in MacLeod, 1998), we found no positive correlation between response-time facilitation and response-time interference (with rs between $-.15$ and .25). Although this may be taken to support our idea that error- and latency-interference effects are dissociable, it is more likely to reflect the statistical artifact of calculating facilitation and interference effects from the same neutral baseline (Watson, Brown, & Stott, 2001). In any case, the main point here is that the two key indices of goal neglect in our experiments, error interference and latency facilitation, were positively correlated with one another. Moreover, whereas error interference correlated with latency facilitation, error and latency interference did not significantly correlate with each other (rs $= -.02$, .17, and .10 in Experiments 1–3B, respectively), further suggesting that the two indices of interference may reflect different mechanisms.

We would therefore amend MacLeod's (1998; MacLeod & MacDonald, 2000a, 2000b) conclusion that Stroop facilitation is independent of Stroop interference. Facilitation may be independent of the competition-resolution process that largely contributes

---

[6] We did not examine response-time distributions in the present study for two reasons: (a) We were able to induce overt word-reading differences across span groups and congruency contexts, and so examination of any potentially more subtle goal-neglect effects in skew would be redundant; (b) Larger numbers of trials per participant are needed to appropriately analyze response-time distributions than we included in our design.
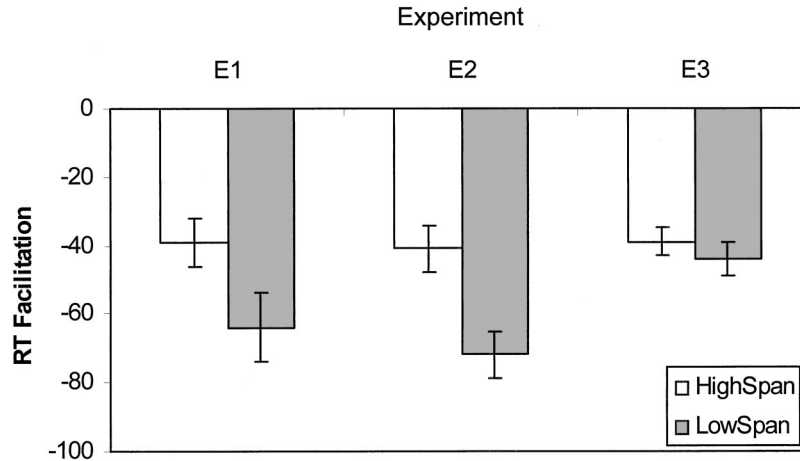
Experiment

Figure 4. Mean response-time facilitation effects (in milliseconds) for high- and low-span participants from the 75% congruent conditions from Experiments 1–3. Facilitation effects were calculated by subtracting participants' mean neutral-trial latency from congruent-trial latency. Error bars depict standard errors of the means. E1 = Experiment 1; E2 = Experiment 2; E3 = Experiment 3; RT = response time.

to interference in 0% congruent tasks and in other conditions that minimize the need for goal maintenance. However, facilitation is, in fact, at least partly dependent on the goal-neglect mechanism that largely contributes to interference in 75% congruent Stroop tasks (and in other conditions that put a premium on effective goal maintenance). Although our data cannot rule out a possible role of convergent color and word responses in producing Stroop facilitation (e.g., Jacoby et al., 1999), our data clearly do support a role for goal neglect in determining the effect.

An Attentional View of WM Capacity

We have recently argued that the mechanism responsible for the broad predictive power of WM span tasks is a domain-general, executive-attention mechanism that allows for the maintenance of stimulus representations, action plans, and goals in the face of interference (Engle, 2001, 2002; Engle, Kane, & Tuholski, 1999; Engle, Tuholski et al., 1999; Kane & Engle, in press). Our view is that, although the active maintenance of information may be useful in many contexts, it is most critical under conditions of interference. In the absence of interference, task-relevant information or goals can be retrieved on demand from inactive long-term memory. In the presence of interference, however, task-irrelevant information and action schemas are likely to be retrieved, resulting

Table 6
Correlations of Facilitation Effects With 75%-Task-Interference Effects (in Response Times and Errors) in Experiments 1, 2, 3A and 3B

Experiment	Response-time interference (*r*)	Error interference (*r*)
1	.25	.45*
2	.07	.35*
3A & 3B	−.15	.34*

* $p < .05$.

in errors or slowed responding (e.g., Norman, 1981; Reason, 1990). Thus, interference sets the occasion for reliance on active memory maintenance. Moreover, individual differences in such active, controlled maintenance determine the capability to prevent attentional focus from being captured by mental or environmental distractors—to avoid powerful distractors, one must actively maintain the intention to do so (see also Roberts & Pennington, 1996). By this view, then, active maintenance and distractor blocking are interdependent executive-attention processes that are central to WM capacity.

Our findings, here, clearly support this theoretical view. As in dichotic listening (Conway et al., 2001) and antisaccade tasks (Kane et al., 2001), low-WM-span individuals demonstrated worse performance than high spans in Stroop tasks that present conflict between habit and goal. Despite the surface differences among these three attention-control tasks, they all had in common the requirement that powerful stimuli for action had to be ignored for goal-appropriate responding: Prepotent, elicited responses had to be held in check. In all of these tasks, low spans perform significantly worse than do high spans, and in particular, low spans are deficient in their ability to prevent attentional "capture" and respond in a goal-directed manner (see Conway & Kane, 2001).

Our discovery that the proportion of congruent trials in the Stroop task drives WM span differences in error interference supports our idea that goal maintenance, and individual differences therein, are critical to action control and selective attention. At least, this is true under conditions where the prompts for action are compelling but in conflict with momentary intention. When the Stroop context promoted goal neglect by failing to reinforce task goals, low spans committed more errors than did high spans. In contrast to theoretical arguments that inhibitory capabilities constrain WM capacity (e.g., Dempster, 1991, 1992; Hasher & Zacks, 1988), then, our error interference findings suggest that WM capabilities constrain attentional inhibition, or at least the consistency of its application (see also De Jong, 2000, 2001; Engle, 1996; Roberts et al., 1994; Roberts & Pennington, 1996).

However, in the present experiments we also found WM-span differences in Stroop contexts that did not encourage goal neglect, such as in 0% and 20% congruent conditions, where low spans showed greater response-time interference than did high spans (see also Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). Here, goal maintenance was presumed to be less critical than in the 75% and 80% congruent contexts, because the repeated presentation of incongruent trials should have reinforced the ignore-the-word goal of the task often enough to aid goal-directed responding. Our finding of span differences in interference here, along with dissociations between response competition and goal neglect in the Stroop literature, indicates that WM capacity may be linked to additional attentional capabilities that are not obviously involved in active goal maintenance in the face of interference. Indeed, recent demonstrations that WM span correlates with negative priming (Conway, Tuholski, Shisler, & Engle, 1999), with counting from 1 to 12 visual objects (Tuholski, Engle, & Baylis, 2001), and with flexibly allocating the extent of visual attention (Bleckley, 2001), suggest that the relation between WM capacity and attentional functioning will require further specification.

Moreover, in a study published after we completed Experiment 4, Long and Prat (2002) found that WM span is differentially related to Stroop interference according to incongruency-proportion contexts. As in our studies, they found in two experiments that when incongruent trials made up the majority of trials in the task, low-span participants showed larger response-time interference effects than did high spans. However, they also found in one experiment that when incongruent trials were relatively rare in the task, high and low spans demonstrated equivalent interference in latencies and errors; this conflicts with our findings of span differences in error interference when incongruent trials were rare. Of importance, in the Long and Prat study, low proportions of incongruent trials accompanied high proportions of neutral trials (*XXXXX*) rather than congruent trials. A context that is high in neutral trials should not induce goal neglect because word reading is not a possible response on these trials, so this is almost certainly the source of the discrepancy between our findings. More work will need to be done to determine why a context that is very high in neutral trials would prevent high spans from more effectively resolving the response competition present on incongruent trials than low spans. However, our findings, along with Long and Prat's, clearly indicate that there are important interactions between WM capacity and task context in mediating Stroop interference.

CONCLUSIONS

Individuals of high- and low-WM capabilities, as indicated by scores on the OSPAN task, were tested in variations of the Stroop task. These variations differed in the proportion of congruent trials (e.g., *RED* presented in red) included in the design, with either low proportions (0%, 20%) or high proportions (75%, 80%). In all five experiments, high-WM-span participants demonstrated less interference than did low-span participants. These findings are consistent with the idea that WM capacity, as defined by complex span measures, is a valid predictor of attention control.

However, the locus of the WM-span difference in Stroop interference—in latency or accuracy—varied with congruency conditions and the order of congruency conditions. In Experiments 1, 2, and 4, where a high-congruency Stroop task was presented alone or as the first task, low spans demonstrated larger error interference effects than did high spans, with low spans more often naming the word when the color and word were in conflict. In contrast, tasks that presented 0% or 20% congruent trials in these experiments elicited consistently modest span differences in response-time interference. Low-congruency contexts generally allowed low spans to respond accurately, but they responded more slowly on incongruent trials than did high spans. Experiment 3A and 3B presented a 75% congruent task following a 0% task, and here low spans demonstrated significantly larger response-time interference than did high spans in both tasks. Low spans were not more likely to commit overt errors, but they were relatively slowed in accurately naming the color of incongruent words. Here, then, the 0% congruent task appeared to elicit a task set that transferred to the subsequent 75% task. Experiment 4 yielded different results using a 20% congruent task to precede an 80% congruent task. Here, the subsequent 80% task elicited greater error interference in low spans than in high spans, suggesting that tasks that include even a small proportion of congruent trials do not induce a task set that is strong enough to transfer to a subsequent high-congruency task.

Our findings and a review of the literature suggest that Stroop interference is jointly determined by two mechanisms, both sensitive to individual differences in WM capacity. One is a response-competition mechanism predominating in contexts with few congruent trials. The time-consuming operation of this mechanism is reflected primarily in response times, and more specifically, in the regular shift of the latency distribution between incongruent and neutral trials. The other mechanism is a more transient failure of goal maintenance that predominates in contexts with many congruent trials. Failures of this mechanism are reflected primarily in error rates, but also potentially in the tail of the latency distribution. We argue that examining individual differences in WM capacity may thus inform theories and mechanisms of the Stroop task, specifically, in addition to addressing theories of attention and action control more generally.

References

Ahola, K., Vilkki, J., & Servo, A. (1996). Frontal tests do not detect frontal infarctions after ruptured intracranial aneurysm. *Brain and Cognition, 31,* 1–16.

Baddeley, A. D. (1986). *Working memory.* London/New York: Oxford University Press.

Baddeley, A. D. (1993). Working memory or working attention? In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control* (pp. 152–170). Oxford, England: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Paulesu, E., Frackowiak, R. S. J., & Dolan, R. J. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia, 31,* 907–922.

Bleckley, M. K. (2001). *Individual differences in visual attention and working memory capacity: Further distinctions between where and what.* Unpublished doctoral dissertation, Georgia Institute of Technology, Atlanta, GA.

Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In

S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 713–737). Cambridge, MA: MIT Press.

Brewin, C. R., & Beaton, A. (2002). Thought suppression, intelligence, and working memory capacity. *Behavior Research and Therapy, 40,* 923–930.

Buckley, T. C., Blanchard, E. B., & Hickling, E. J. (2002). Automatic and strategic processing of threat stimuli: A comparison between PTSD, panic disorder, and nonanxiety controls. *Cognitive Therapy and Research, 26,* 97–115.

Cohen, J. D., Barch, D. M., Carter, C., & Servan-Schreiber, D. (1999). Context-processing deficits in schizophrenia: Converging evidence from three theoretically motivated cognitive tasks. *Journal of Abnormal Psychology, 108,* 120–133.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review, 99,* 45–77.

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin and Review, 8,* 331–335.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30,* 163–183.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123,* 354–373.

Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory, 4,* 577–590.

Conway, A. R. A., & Kane, M. J. (2001). Capacity, control and conflict: An individual differences perspective on attentional capture. In C. Folks & B. Gibson (Eds.), *Attraction, distraction and action: Multiple perspectives on attention capture* (pp. 349–372). Amsterdam: Elsevier Science.

Conway, A. R. A., Tuholski, S. W., Shisler, R. J., & Engle, R. W. (1999). The effect of memory load on negative priming: An individual differences investigation. *Memory & Cognition, 27,* 1042–1050.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

De Jong, R. (2000). An intention-activation account of residual switch costs. In S. Monsell & J. Driver (Eds.), *Control of Cognitive Processes: Attention and Performance XVIII* (pp. 357–376). Cambridge, MA: MIT Press.

De Jong, R. (2001). Adult age differences in goal activation and goal maintenance. *European Journal of Cognitive Psychology, 13,* 71–89.

De Jong, R. D., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: Dissociable causes of interference effects in conflict situations. *Acta Psychologica, 101,* 379–394.

Dempster, F. N. (1991). Inhibitory processes: A neglected dimension in intelligence. *Intelligence, 15,* 157–173.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review, 12,* 45–75.

Duncan, J. (1990). Goal weighting and the choice of behavior in a complex world. *Ergonomics, 33,* 1265–1279.

Duncan, J. (1993). Selection of input and goal in the control of behavior. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control* (pp. 53–71). Oxford, England: Oxford University Press.

Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M.

Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia, 33,* 261–268.

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. (1996). Working memory and retrieval: An inhibition-resource approach. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zacks (Eds.), *Working memory and human cognition* (pp. 89–119). New York: Oxford University Press.

Engle, R. W. (2001). What is working memory capacity? In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Suprenant (Eds.), *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., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 972–992.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). New York: Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128,* 309–331.

Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 257–297). New York: Cambridge University Press.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18,* 1279–1296.

Hallett, P. E., & Adams, B. D. (1980). The predictability of saccadic latency in a novel voluntary oculomotor task. *Vision Research, 20,* 329–339.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109,* 340–347.

Jacoby, L. L., McElree, B., & Trainham, T. N. (1999). Automatic influences as accessibility bias in memory and Stroop tasks: Toward a formal model. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 461–486). Cambridge, MA: MIT Press.

Jensen, A. R. (1965). Scoring the Stroop test. *Acta Psychologica, 24,* 398–408.

Jensen, A. R., & Rohwer, W. D. (1966). The Stroop color–word test: A review. *Acta Psychologica, 25,* 36–93.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183.

Kane, M. J., & Engle, R. W. (2000). Working memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336–358.

Kane, M. J., & Engle, R. W. (in press). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review.*

Klein, K., & Boals, A. (2001). Expressive writing can increase working memory capacity. *Journal of Experimental Psychology: General, 130,* 520–533.

Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments and Computers, 31,* 429–432.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389–433.

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1118–1133.

Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*(A), 29–50.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 219–234.

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color–word interference effect. *Neuropsychologia, 38,* 701–711.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition, 7,* 166–174.

Logan, G. D., & Zbrodoff, N. J. (1998). Stroop-type interference: Congruity effects in color naming with typewritten responses. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 978–992.

Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition, 3,* 294–301.

Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided attention in a Stroop task. *Canadian Journal of Psychology, 36,* 684–700.

Lundh, L. G., Wikstroem, J., & Westerlund, J. (2001). Cognitive bias, emotion, and somatic complaints in a normal sample. *Cognition & Emotion, 15,* 249–277.

MacDonald, A. W., III, Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000, June). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288,* 1835–1838.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition, 26,* 201–211.

MacLeod, C. M., & MacDonald, P. A. (2000a). *Facilitation in the Stroop task is illusory: The inadvertent reading hypothesis.* Manuscript under review.

MacLeod, C. M., & MacDonald, P. A. (2000b). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4,* 383–391.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29,* 10–17.

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Neuroscience Reviews, 1,* 59–65.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11,* 56–60.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88,* 1–15.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York: Plenum Press.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (in press). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence.*

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically-based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375–411). New York: Cambridge University Press.

Pardo, J. V., Pardo, P. J., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences, USA, 87,* 256–259.

Park, S., Püschel, J., Sauter, B. H., Rentsch, M., & Hell, D. (1999). Spatial working memory deficits and clinical symptoms in schizophrenia: A 4-month follow-up study. *Biological Psychiatry, 46,* 392–400.

Perlstein, W. M., Carter, C. S., Barch, D. M., & Baird, J. W. (1998). The Stroop task and attention deficits in schizophrenia: A critical evaluation of card and single-trial Stroop methodologies. *Neuropsychology, 12,* 414–425.

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behavior. *Neuropsychologia, 12,* 323–330.

Posner, M. I., & Peterson, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25–42.

Posner, M. I., & Raichle, M. E. (1994). *Images of mind.* New York: Freeman.

Reason, J. T. (1990). *Human error.* Cambridge, England: Cambridge University Press.

Richer, F., Decary, A., Lapierre, M. F., Rouleau, I., Bouvier, G., & Saint-Hilaire, J. M. (1993). Target detection deficits in frontal lobectomy. *Brain and Cognition, 21,* 203–211.

Roberts, R. J., Jr., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123,* 374–393.

Roberts, R. J., Jr., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology, 12,* 105–126.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language, 39,* 418–436.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM-PC compatibles. *Behavior Research Methods, Instrumentation, and Computers, 20,* 206–217.

Seddon, K., & Waller, G. (2000). Emotional processing and bulimic psychopathology: Age as a factor among nonclinical women. *International Journal of Eating Disorders, 28,* 364–369.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, England: Cambridge University Press.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727–741.

Shute, V. J. (1991). Who is likely to acquire programming skills? *Journal of Educational Computing Research, 7,* 1–24.

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.

Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 506–526.

Spilsbury, G. (1992). Complexity as a reflection of the dimensionality of a task. *Intelligence, 16,* 31–45.

Spivack, G., Levine, M., & Sprigle, H. (1959). Intelligence test performance and the delay function of the ego. *Journal of Consulting Psychology, 23,* 428–431.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Süß, H.-M., Oberauer, K., Wittman, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30,* 261–288.

Swanson, H. L., & Sachse-Lee, C. (2001). A subgroup analysis of working memory in children with reading disabilities: Domain-general or domain-specific deficiency? *Journal of Learning Disabilities, 34,* 249–263.

Taylor, F. B., & Russo, J. (2001). Comparing guanfacine and dextroamphetamine for the treatment of adult attention-deficit/hyperactivity disorder. *Journal of Clinical Psychopharmacology, 21,* 223–228.

Trainham, T. N., Lindsay, D. S., & Jacoby, L. L. (1997). Stroop process dissociations: Reply to Hillstrom and Logan (1997). *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1579–1587.

Tuholski, S. W., Engle, R. W., & Baylis, G. C. (2001). Individual differences in working memory capacity and enumeration. *Memory & Cognition, 29,* 484–492.

Turley-Ames, K. J., & Whitfield, M. M. (2002). *Strategy training and working memory task performance.* Manuscript submitted for publication.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154.

Watson, W. S., Brown, T. L., & Stott, J. A. (2001, March). *Why are Stroop facilitation and interference effects inversely correlated?* Paper presented at the annual meeting of the Southeastern Psychological Association, Atlanta, GA.

West, R. (2001). The transient nature of executive control processes in younger and older adults. *European Journal of Cognitive Psychology, 13,* 91–105.

West, R., & Alain, C. (1999). Event-related neural activity associated with the Stroop task. *Cognitive Brain Research, 8,* 157–174.

West, R., & Alain, C. (2000a). Effects of task context and fluctuations of attention on neural activity supporting performance of the Stroop task. *Brain Research, 873,* 102–111.

West, R., & Alain, C. (2000b). Evidence for the transient nature of a neural system supporting goal-directed action. *Cerebral Cortex, 10,* 748–752.

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 255–260.

Zysset, S., Müller, K., Lohman, G., von Cramon, D. Y. (2001). Color–word matching Stroop task: Separating interference and response conflict. *NeuroImage, 13,* 29–36.

Appendix A

Complete ANOVA Output for Response-Time Data From Experiment 4

Effects and interactions	*F*	*MSE*	*p*
Main effects			
Span (S)	4.01	38,533.42	.05
Congruency-proportion order (CPO)	0.05	38,533.42	.83
Congruency-proportion (CP)	28.48	4,642.86	.00
Trial type (TT)	1,448.40	3,517.11	.00
2-way interactions			
S × CPO	0.74	38,533.42	.40
S × CP	6.31	4,642.86	.01
S × TT	13.49	3,517.11	.00
CPO × CP	41.59	4,642.86	.00
CPO × TT	5.56	3,517.11	.02
CP × TT	670.28	1,419.43	.00
3-way interactions			
S × CPO × CP	0.32	4,642.86	.57
S × CPO × TT	1.46	3,517.11	.23
S × CP × TT	2.23	1,419.43	.14
CPO × CP × TT	10.85	1,419.43	.00
4-way interaction			
S × CPO × CP × TT	0.03	1,419.43	.86

Appendix B

Complete ANOVA Output for Error-Rate Data From Experiment 4

Effects and interactions	*F*	*MSE*	*p*
Main effects			
Span (S)	11.87	0.007	.00
Congruency-proportion order (CPO)	0.30	0.007	.59
Congruency-proportion (CP)	89.25	0.002	.00
Trial type (TT)	224.87	0.004	.00
2-way interactions			
S × CPO	0.28	0.007	.60
S × CP	3.49	0.002	.06
S × TT	16.32	0.004	.00
CPO × CP	8.53	0.002	.00
CPO × TT	1.95	0.004	.16
CP × TT	132.15	0.002	.00
3-way interactions			
S × CPO × CP	1.47	0.002	.23
S × CPO × TT	0.02	0.004	.88
S × CP × TT	16.76	0.002	.00
CPO × CP × TT	7.40	0.002	.01
4-way interaction			
S × CPO × CP × TT	0.00	0.002	.99