Journal of Experimental Psychology:
Learning, Memory, and Cognition
2009, Vol. 35, No. 1, 196–204

© 2009 American Psychological Association
0278-7393/09/$12.00 DOI: 10.1037/a0014104

Conducting the Train of Thought: Working Memory Capacity, Goal Neglect, and Mind Wandering in an Executive-Control Task

Jennifer C. McVay and Michael J. Kane
University of North Carolina at Greensboro

On the basis of the executive-attention theory of working memory capacity (WMC; e.g., M. J. Kane, A. R. A. Conway, D. Z. Hambrick, & R. W. Engle, 2007), the authors tested the relations among WMC, mind wandering, and goal neglect in a sustained attention to response task (SART; a go/no-go task). In 3 SART versions, making conceptual versus perceptual processing demands, subjects periodically indicated their thought content when probed following rare no-go targets. SART processing demands did not affect mind-wandering rates, but mind-wandering rates varied with WMC and predicted goal-neglect errors in the task; furthermore, mind-wandering rates partially mediated the WMC–SART relation, indicating that WMC-related differences in goal neglect were due, in part, to variation in the control of conscious thought.

Keywords: working memory, executive control, mind wandering

Why does working memory capacity (WMC), as measured by complex memory-span tasks, predict individual differences in fluid cognitive abilities? Attentional theories argue that WMC tasks' predictive power derives largely from their tapping domain-general, executive-control capabilities, which are also widely important to complex cognition (e.g., Hasher, Lustig, & Zacks, 2007; Kane, Conway, Hambrick, & Engle, 2007). Some supporting evidence comes from correlations between WMC and simple attention tasks that make limited memory demands, such as the antisaccade task. Here, higher WMC subjects better restrain the habitual response of orienting toward a visual-onset cue than do lower WMC subjects, allowing them to more successfully act according to the task goal of looking in the opposite direction (e.g., Kane, Bleckley, Conway, & Engle, 2001; Unsworth, Schrock, & Engle, 2004).

Some of the WMC-related variation in attention-task performance seems attributable to individual differences in maintaining sufficient access to the current task goals so that they, rather than habit, control responding (see Kane, Conway, et al., 2007). In the Stroop task, for example, which elicits habit–goal conflict (i.e., word reading vs. color naming), Kane and Engle (2003) presented subjects with either many incongruent, mismatching trials (the word *BLUE* in red) or many congruent, matching trials (the word *RED* in red). With many incongruent trials, the context reinforced the color-naming goal because most trials presented word–color conflict and thus demanded ignoring the words; active goal maintenance was thus aided (or supplanted) by environmental support. In contrast, with many congruent trials, goals were not contextually reinforced. Word reading allowed correct responses on most trials, so subjects had to actively maintain goal access to respond appropriately to the rare incongruent trials. Indeed, WMC-related differences were strongest in high-congruent conditions, where goal maintenance was most critical: Lower WMC subjects committed 50–100% more errors than did higher WMC subjects on incongruent trials, apparently maintaining less suitable access to goal-relevant information.

We suggest that lower WMC subjects show frequent *goal neglect* (Duncan, 1995) because goal maintenance fluctuates across trials depending, in part, on the ability to resist interference from task-unrelated thoughts (TUTs). Simply put, lower WMC subjects seem less able than higher WMC subjects to sustain attention to the demands of the ongoing task. This intuitive view is not universally accepted, however. Oberauer and colleagues (e.g., Oberauer, Süβ, Wilhelm, & Sander, 2007; Wilhelm & Oberauer, 2006) instead attribute such goal-neglect failures to insufficient binding of stimulus–response (S-R) mappings. At a broad level, Oberauer et al. claimed that WMC variation reflects the ability to establish, maintain, and decouple mental bindings among a limited number of activated representations, as in associating auditory stimuli with temporal sequences, visual stimuli with locations, or novel responses with imperative stimuli. In tasks such as Stroop and antisaccade, then, lower WMC subjects, as compared with higher WMC subjects, may respond slowly or more frequently in error

Jennifer C. McVay and Michael J. Kane, Department of Psychology, University of North Carolina at Greensboro.

We are grateful to Jessica Allen, Cody Burns, Tammy Core, Amy Crompton, Rick Fisher, Robin Grochowski, Brittany Lambeth, Katy Beth Loner, Jody Ortman, Amanda Page, Nina Powell, Dara Rogers, Lindsay Squires, Alanna Stockford, and Lucas Thomas for assistance in data collection.

Portions of this work were supported by a University of North Carolina at Greensboro Regular Faculty Grant awarded to Michael J. Kane and a National Institutes of Health Ruth L. Kirschstein National Research Service Award Individual Predoctoral Fellowship granted to Jennifer C. McVay, Award Number F31MH081344 from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

Correspondence concerning this article should be addressed to Michael J. Kane, Department of Psychology, University of North Carolina at Greensboro, P.O. Box 26170, Greensboro, NC 27402-6170. E-mail: mjkane@uncg.edu

because they cannot as effectively bind incompatible S-R mappings, not because their attention cannot be as effectively maintained throughout the task. In fact, Wilhelm and Oberauer (2006) found that both WMC and fluid intelligence correlated strongly with performance of choice-response-time tasks presenting arbitrary S-R mappings.

Here we tested the binding versus attentional views of WMC variation in goal neglect by probing subjects' thoughts during an executive-control task. If insufficient binding or drift in its efficacy is responsible for goal-neglect errors, then subjects' TUT experiences and WMC-related variation in mind wandering should be irrelevant. If, however, lapses of goal maintenance that accompany (or result from) slips of thought actually contribute to goal-neglect errors and if WMC variation predicts subjects' goal-maintenance efficacy, then TUT intrusion rates should mediate (at least partially) the relation between WMC and goal neglect. Indeed, TUTs often predict performance errors (Smallwood & Schooler, 2006) and lower WMC subjects experience more mind wandering during effortful daily life activities than do higher WMC subjects (Kane, Brown, et al., 2007). Indirect evidence thus supports our claim that individual differences in attention control and mind wandering contribute to WMC's association with goal maintenance and neglect.

Our more direct test here attempted to link goal-neglect errors to subjective experience within a task yielding high rates of goal neglect and mind wandering. The sustained attention to response task (SART; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997) is a go/no-go task requiring responses to all stimuli except infrequent targets. Whereas previous SART research has administered thought probes at least several seconds (to half a minute) following critical target events (e.g., Smallwood et al., 2004; Smallwood, McSpadden, & Schooler, 2007), we probed thoughts immediately after no-go targets to link in-the-moment subjective experience to performance. We predicted that TUT reports would be accompanied by more errors than would on-task thought reports and that higher WMC subjects would experience fewer TUTS and commit fewer performance errors than would lower-WMC subjects (moreover, to the extent that response time [RT] variability may also reflect more subtle slips of thought and goal neglect, we also predicted that WMC and TUT rate would predict intraindividual RT variation). Of most importance, we hypothesized that TUT rate would partially mediate the relation between WMC and SART performance (accuracy and RT variability), indicating that attention control contributes to WMC's influence on response-conflict tasks.

Of secondary interest, we adapted the SART to contrast the effects of different ongoing processing demands on mind wandering, namely conceptual versus perceptual judgments (subjects either responded to words from one semantic category and withheld responses to another category or responded to words in one font and withheld responses to another font). This manipulation was motivated by research suggesting that TUT frequency decreases when people engage in more conceptual versus perceptual processing. For example, subjects report fewer TUTs when studying and recalling words according to conceptual versus orthographic dimensions (e.g., *musical instruments* vs. *words beginning with P*; Smallwood, Baracia, Lowe, & Obonsawin, 2003; Smallwood, Obonsawin, & Heim, 2003). Most relevant here, Smallwood, Riby, Heim, and Davies (2006) reported lower TUT rates for subjects completing a semantic SART, where subjects responded to words and withheld responses to *XXXXX* strings, versus a perceptual SART that replaced all words with *OOOOO* strings. This finding warrants further examination, however, because the semantic benefit only occurred for subjects instructed to memorize the SART words for a subsequent test and not for subjects who encoded the words incidentally. We therefore followed up this work by manipulating the SART's conceptual demands while more closely matching other task features.

Finally, we tested the association between in-the-moment TUT reports and general retrospective reports of cognitive failures with the Cognitive Failures Questionnaire (CFQ; Broadbent, Cooper, FitzGerald, & Parkes, 1982). The CFQ assesses everyday attention, memory, and motor failures and modestly predicts SART errors (Robertson et al., 1997) and TUT rates (Smallwood et al., 2004). We sought to replicate these findings while pitting our objective WMC measures against the CFQ in predicting SART performance and TUTs.

Method

Subjects

Two hundred forty-four undergraduates (aged 18–35 years) completed WMC and SART sessions during one semester. We dropped data from 1 subject who did not follow SART instructions.

WMC Screening

In 90-min sessions, we tested 3–6 subjects using three automated complex-span tasks: operation span (OSPAN), symmetry span (SSPAN), and reading span (RSPAN). The tasks required subjects to maintain access to memory items while completing an unrelated processing task with an individualized response deadline (M + 2.5 SDs), calculated during 15 processing-task-only items (Unsworth, Heitz, Schrock, & Engle, 2005). In OSPAN, subjects verified solutions to compound equations. In RSPAN, subjects verified the meaningfulness of sentences. In SSPAN, subjects verified the symmetry of black-and-white matrix patterns. In OSPAN and RSPAN, a capital letter (randomly selected among 12) appeared for 250 ms, 200 ms after either operation–reading verification or response deadline. After 3–7 verification-letter pairs, all 12 letters appeared onscreen and subjects identified, via mouse click, the presented letters in serial order. In SSPAN, one square of a 4 × 4 grid was shaded red for 650 ms, 200 ms after either symmetry verification or response deadline. After 2–5 verification–grid pairs, subjects recalled the locations of the colored squares in serial order by mouse clicking on an empty grid. The tasks presented each set length (3–7 in OSPAN and RSPAN; 2–5 in SSPAN) three times, randomly ordered for each subject.

The span score was the sum of items recalled in serial position (Conway et al., 2005). We converted span scores to z scores and averaged them into a WMC composite. Scores correlated .65 (RSPAN × OSPAN), .56 (OSPAN × SSPAN), and .53 (SSPAN × RSPAN). The WMC composite was normally distributed (skew = −0.64; kurtosis = 0.07).

SART

Design and Materials

The design was a 3 × 2 mixed-model factorial, with SART type (semantic, perceptual, perceptual–semantic) manipulated between subjects and stimulus type (target, nontarget) manipulated within subjects. We defined targets as the no-go trials presenting an infrequent stimulus type and requiring restraint of the prepotent "go" response.

In semantic SART, nontarget words came from one category (e.g., animals) and no-go targets from another (e.g., foods), counterbalanced across subjects. In perceptual SART, nontarget words appeared in lowercase type and no-go targets in uppercase. In a third condition, perceptual–semantic, subjects made perceptual decisions but targets and nontargets differed on both dimensions (e.g., animals vs. foods). Animal and food names (excluding animals commonly eaten) for semantic and perceptual–semantic SARTs came from Battig and Montague (1969). We drew words for the perceptual SART quasi-randomly from all Battig–Montague categories. Stimuli appeared in black against a white background, in 18-point Courier-New font, via CRT or LCD monitors.

Procedure

We tested subjects individually in sound-attenuated rooms with white noise machines. Subjects completed a modified CFQ and then the SART.

CFQ–Memory and Attention Lapses (CFQ-MAL). We modified the CFQ to present only its items about memory and attention lapses; we also created new items and drew others from similar questionnaires (Brown & Ryan, 2003; Reason & Mycielska, 1982; Sunderland, Harris, & Baddeley, 1983; for the full scale, see http://www.uncg.edu/~mjkane/memlab.html). This computerized CFQ-MAL presented 40 questions (with responses on a 1–5 scale: 1 = *never*, 2 = *rarely*, 3 = *once in a while*, 4 = *often*, 5 = *very often*); subjects responded via keypress. Total score reflected the item sum. For our sample (N = 242; data from 1 subject were lost), principal-components analysis yielded a first component (eigenvalue = 11.5) accounting for 29% of the variance; the second (eigenvalue = 2.1) accounted for only 5.3%, so we calculated one score for each subject (M = 111.5, SD = 19.1, skewness = 0.64, kurtosis = 0.67).

SART. An experimenter read aloud onscreen instructions. Subjects were to press the space bar as quickly as possible for nontargets and withhold responses to targets. Subjects completed 10 practice trials before seeing thought-probe instructions, which included a thought-probe screen with the question, "What were you just thinking about?" and seven response options. We instructed subjects to report what they were thinking just before the probe, and the experimenter elaborated on these choices: (a) *task*, that is, thinking about the stimulus words or appropriate response; (b) *task performance*, that is, evaluating one's own performance; (c) *everyday stuff*, that is, thinking about recent or impending life events or tasks; (d) *current state of being*, that is, thinking about conditions such as hunger or sleepiness; (e) *personal worries*, that is, thinking about concerns, troubles, or fears; (f) *daydreams*, that is, having fantasies disconnected from reality; or (g) *other*, that is, other thought types. During the task, thought probes presented the italicized category names; subjects then pressed the corresponding number key.

The SART presented 1,810 words: Each was centered for 300 ms and then followed by a 900-ms mask (12 capitalized *X*s, the length of the longest word). The first 10 (unanalyzed) buffer trials presented nontargets. The remaining trials comprised eight blocks, each presenting 225 trials consisting of 45 words repeated five times in a different random order. Within each set of 45, five targets appeared randomly among 40 nontargets (11% of trials). The same five targets appeared across all blocks. Thought probes followed 60% of the targets within each block. After the first four blocks, subjects took a 30-s break. Because there were only five target events per block, our analyses collapsed the eight task blocks into four task-quarter blocks.

Results

We report nondirectional null-hypothesis significance tests with an alpha of .05 and partial eta-squared (η_p^2) as an effect-size estimate.

SART Performance

Accuracy

Mean accuracy rates for target (no-go) and nontarget (go) trials were .49 and .95, respectively. For each subject, we calculated signal-detection sensitivity (d_L) and bias (C_L) scores, using formulas for logistic distributions (Snodgrass & Corwin, 1988) and adjusting individual hit or false-alarm rates of 0 and 1 by .01. Negative C_L scores reflect a "go" bias. Figures 1A and 1B present d_L and C_L scores by task and block.

A 3 (SART type) × 4 (block) mixed-model analysis of variance (ANOVA) on d_L confirmed a main effect of only block, $F(3, 720) = 35.52$, $\eta_p^2 = .13$, modified by an interaction, $F(6, 720) = 3.64$, $\eta_p^2 = .03$, indicating a more shallow sensitivity decrease for the semantic than the perceptual SARTs. Repeated-measures ANOVAs for each SART, however, indicated significant block effects: semantic, $F(3, 249) = 4.76$, $\eta_p^2 = .05$; perceptual–semantic, $F(3, 231) = 15.79$, $\eta_p^2 = .17$; and perceptual, $F(3, 240) = 25.32$, $\eta_p^2 = .24$. A 3 (SART type) × 4 (block) mixed-model ANOVA on C_L indicated only a block effect, $F(3, 720) = 6.135$, $\eta_p^2 = .03$, and no interaction, $F(6, 720) = 1.22$, $p = .30$, corresponding to a slight decrease in go bias over blocks.

RT

Figures 2A and 2B present two nontarget (go-trial) RT indices: means of individual subjects' means, reflecting central tendency, and means of individual subjects' standard deviations, reflecting intraindividual variability. We were particularly interested in RT variability because it may reflect slight attentional fluctuations over the course of the task and thus might be sensitive to WMC and TUT-rate variation.

Semantic-based responses were slower than perceptually based responses, with stable RTs over blocks: A 3 (SART type) × 4 (block) mixed ANOVA on mean RT indicated only a main effect of SART type, $F(2, 240) = 21.98$, $\eta_p^2 = .15$, and no interaction, $F(6, 720) = 1.17$, $p = .32$. In contrast, RT variability increased over blocks, but similarly across tasks: A 3 (SART type) × 4

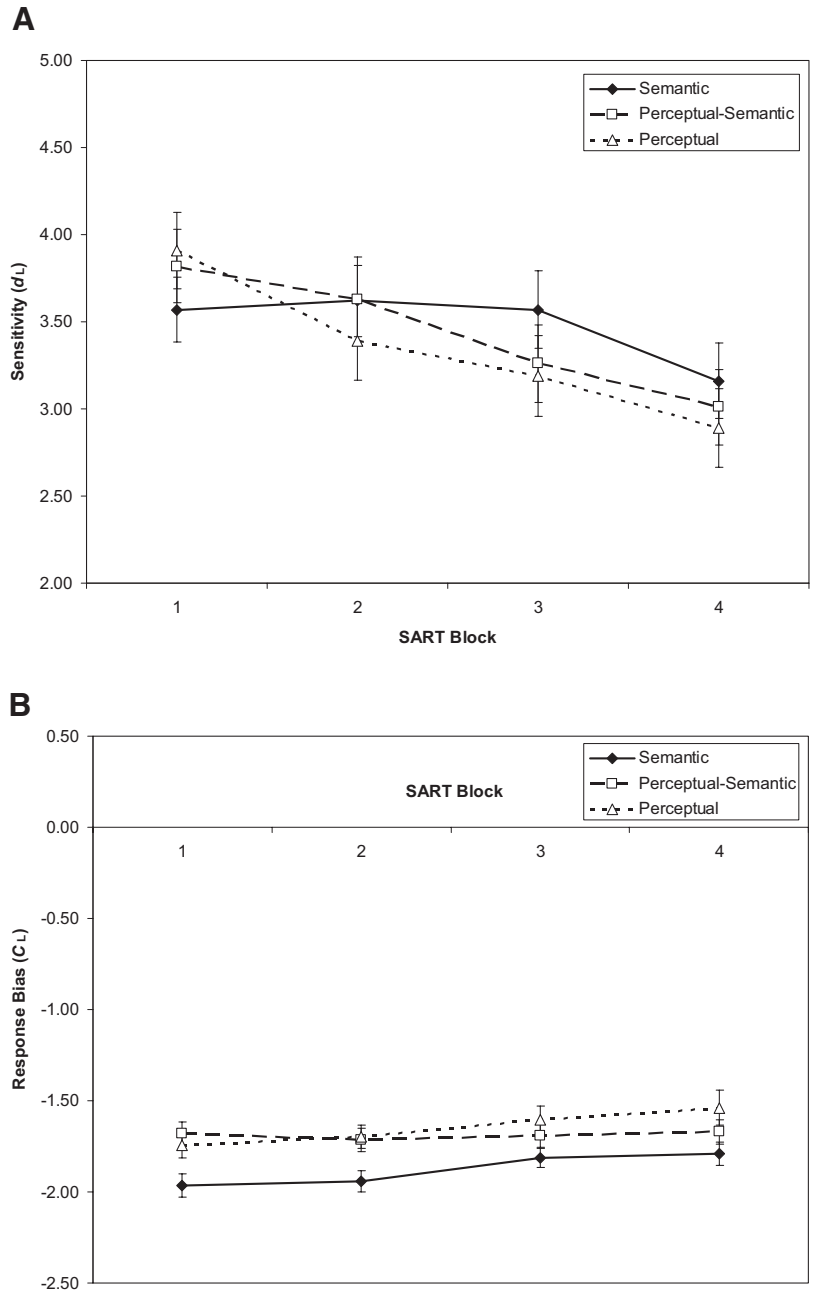

Figure 1. Signal-detection indices from the semantic, perceptual–semantic, and perceptual SARTs (sustained attention to response tasks), across task blocks (N = 243). A: Mean sensitivity (d_L) estimates. B: Mean bias (C_L) estimates. Error bars represent standard errors.

(block) mixed ANOVA confirmed only a block effect, $F(3, 720) = 74.28$, $\eta_p^2 = .24$, and an interaction that approached conventional significance, $F(6, 720) = 1.94$, $p = .07$. Subjects thus became more variable with time on task on all SART types.

In previous SART studies, RTs were shorter preceding target errors than preceding accurate responses, which some investigators have interpreted as habitual, mindless responding (Robertson et al., 1997; Smallwood et al., 2004). Here, too, RTs for the four non-target trials preceding target errors were significantly faster (M = 382 ms) than those preceding correct responses (M = 455 ms), $t(242) = -25.22$.

Thought Reports

Subjects reported task-related and task-unrelated thoughts on 21% and 55% of thought probes, respectively; TUTs were defined as reports of current state (28.4%), daydreams (8.6%), everyday stuff (8.2%), worries (4.7%), and other (5.5%). Thoughts about

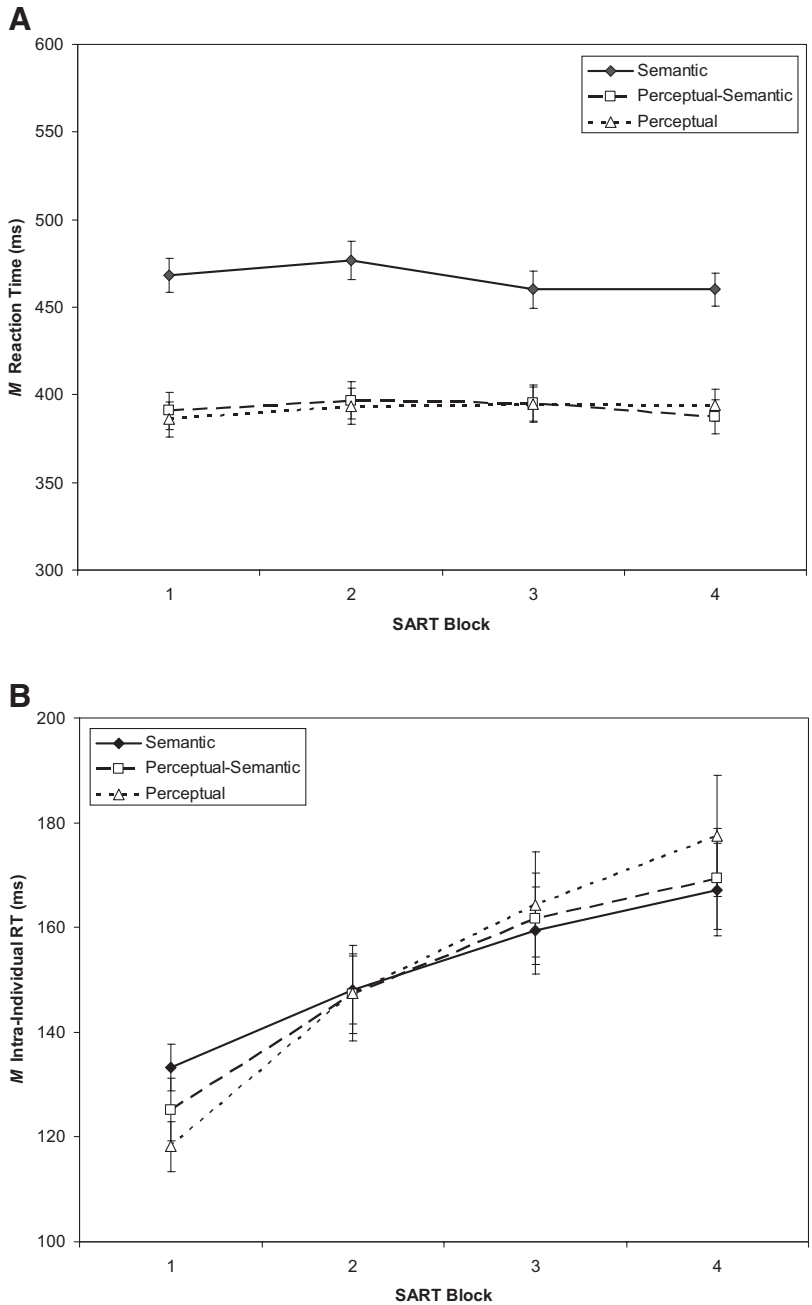

Figure 2. Response time (RT) measures from the semantic, perceptual–semantic, and perceptual SARTs (sustained attention to response tasks), across task blocks ($N = 243$). A: Means of individual subjects' mean RTs. B: Means of individual subjects' RT standard deviations. Error bars represent standard errors.

subjects' performance, sometimes labeled *task-related interference* (e.g., Smallwood et al., 2006), composed 24% of responses. As task-related interference represents an ambiguous intermediary between on- and off-task thought, we do not analyze it further.

Figure 3 illustrates that TUTs increased and on-task thoughts decreased over blocks. For TUTs, a 3 (SART type) × 4 (block) mixed ANOVA indicated a main effect of only block, $F(3, 720) = 223.45$, $\eta_p^2 = .48$, and no interaction, $F(3, 720) < 1$. For on-task thoughts, a parallel ANOVA indicated, again, only a block effect, $F(3, 720) = 44.20$, $\eta_p^2 = .16$, and no interaction, $F(3, 720) = 1.49$, $p = .18$. Because thought reports did not vary by SART type, subsequent analyses collapse over this variable.

We expected RTs to trials preceding a TUT to be shorter than those preceding an on-task thought, indicating attentional lapses and nonreflective responding. Indeed, responses to the four non-target trials preceding TUTs were significantly faster ($M = 415$ ms) than those preceding on-task thoughts ($M = 426$ ms), $t(235) = -2.73$.

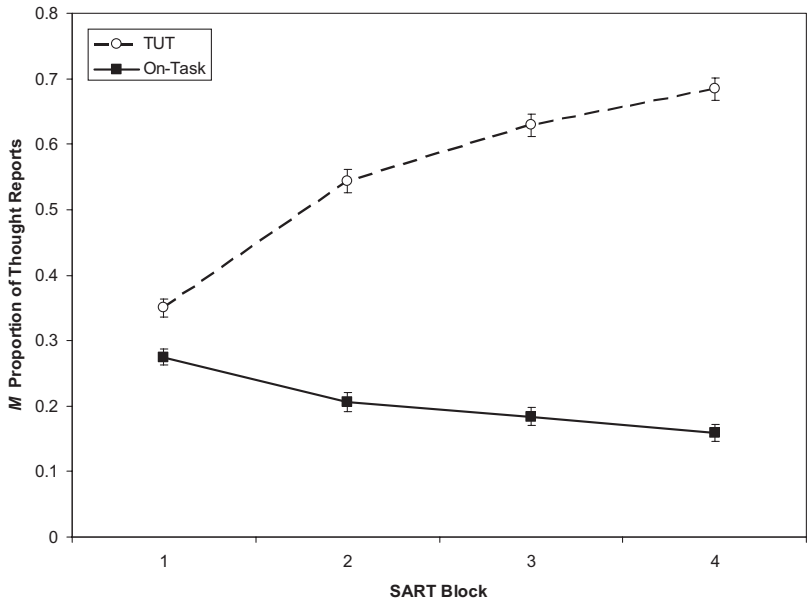

Figure 3. Mean proportion of thought reports by thought category across task blocks (N = 243). Error bars represent standard errors. TUT = task-unrelated thought; on-task = on-task thought, SART = sustained attention to response task.

Performance by Thought Report

No-go accuracy was lower for targets during TUTs (M = .42) than during on-task thoughts (M = .66), $t(231) = -13.83$; a 2 (thought report) × 4 (block) repeated-measures ANOVA indicated that this effect's magnitude persisted across blocks, $F(3, 390) = 1.535$, p = .21. At the level of intratask individual differences, subjects' overall TUT rate predicted d_L ($r = -.37$) and nontarget RT standard deviation (r = .40) but not C_L (r = .11). Moreover, TUT–d_L correlations increased significantly from Block 1 to Block 2 to Block 3 (rs = −.17, −.28, −.39, and −.39 for Blocks 1–4, respectively), as indicated by Williams's t test (Steiger, 1980). Correlations between TUT rate and nontarget RT standard deviations increased significantly from Blocks 1 to 2 only (rs = .19, .35, .37, and .43 for Blocks 1–4, respectively). SART performance thus became more linked to mind wandering as the task progressed.

Intertask Individual Differences

Table 1 presents correlations among all the task variables, along with their reliability estimates. WMC and CFQ-MAL were uncorrelated, and neither score differed among SART groups (Fs < 1). As expected, WMC variation predicted SART performance and thought, correlating significantly with d_L, RT variability, and TUT rate but not with C_L. CFQ-MAL scores showed significant but apparently weaker correlations with SART variables.

Table 2 presents hierarchical-regression analyses predicting SART d_L with WMC, CFQ-MAL, and TUT rate. Considering first WMC and TUTs, each accounted for shared and unique d_L variance: WMC accounted for 8.2%, with about half shared by TUT rate. TUTs predicted 9.8% of the variance independent of WMC (total R^2 = .180). Moreover, WMC, TUT rate, and CFQ-MAL all predicted unique d_L variance, but the three together accounted for little more variance than did WMC and TUT rate alone (total R^2 = .198). Table 3 presents parallel regressions for intrasubject RT variability, where WMC accounted for about 12.3% of the variance, with almost half shared with TUT rate; TUT rate accounted for 10.7% of the variance beyond WMC (total R^2 = .230). Here, WMC and TUT rate again predicted unique variance beyond CFQ-MAL scores, but CFQ-MAL predicted RT variability only beyond WMC, not TUT rate (total R^2 = .240).

Discussion

Subjects who differed in WMC, as measured by complex-span tasks, also varied in SART performance and subjective experience.

Table 1
Correlations Among WMC, Mind Wandering, and Performance Measures

Variable	1	2	3	4	5
1. WMC	—				
2. TUT	−.217**	.885			
3. d_L	.287**	−.368**	.948		
4. RT variation	−.351**	.396**	−.599**	.927	
5. CFQ-MAL	.014	.144*	−.179**	.146*	.930

Note: N = 243. Values on the diagonal reflect Cronbach's alpha for each measure as a reliability estimate; alphas were calculated over task blocks for sustained attention to response task (SART) measures and over items for the CFQ-MAL. WMC = working memory capacity; TUT = proportion self-reported task-unrelated thoughts; d_L = signal-detection sensitivity measure on SART; RT variation = intraindividual standard deviation for nontarget reaction times on SART; CFQ-MAL = Cognitive Failures Questionnaire–Memory and Attention Lapses. N = 242 for CFQ-MAL analyses.
* $p < .05$. ** $p < .01$.

Table 2
Hierarchical Regression Analyses on Sustained Attention to Response Task Signal-Detection Sensitivity Estimate (d_L)

Variable	*B*	*SE*	β	*t*	R^2
Predictors: WMC, TUT					
Step 1: WMC	0.550	0.118	.287	4.651*	.082
Step 2: WMC, TUT	−2.591	0.484	−.321	−5.358*	.180
Predictors: TUT, WMC					
Step 1: TUT	−2.973	0.484	−.368	−6.142*	.135
Step 2: TUT, WMC	0.417	0.115	.217	3.631*	.180
Predictors: CFQ-MAL, TUT					
Step 1: CFQ-MAL	−0.016	0.006	−.179	−2.815*	.032
Step 2: CFQ-MAL, TUT	−2.805	0.487	−.347	−5.762*	.150
Predictors: TUT, CFQ-MAL					
Step 1: TUT	−2.955	0.485	−.366	−6.090*	.134
Step 2: TUT, CFQ-MAL	−0.012	0.006	−.129	−2.136*	.150
Predictors: WMC, CFQ-MAL, TUT					
Step 1: WMC	0.548	0.118	.287	4.636*	.082
Step 2: WMC, CFQ-MAL	−0.017	0.006	−.183	−3.007*	.116
Step 3: WMC, CFQ-MAL, TUT	−2.400	0.486	−.297	−4.941*	.198
Predictors: CFQ-MAL, WMC, TUT					
Step 1: CFQ-MAL	−0.016	0.118	−.179	−2.185*	.032
Step 2: CFQ-MAL, WMC	0.553	0.485	.289	4.755*	.116
Step 3: CFQ-MAL, WMC, TUT	−2.400	0.005	−.297	−4.491*	.198

Note: $N = 243$. WMC = working memory capacity; TUT = proportion self-reported task-unrelated thoughts; CFQ-MAL = Cognitive Failures Questionnaire–Memory and Attention Lapses. $N = 242$ for CFQ-MAL analyses.
* $p < .05$.

Thus, WMC predicted not only attention-task errors and RT variability (see also Kane et al., 2001; Kane & Engle, 2003) but also mind-wandering rates (see also Kane, Brown, et al., 2007); indeed, our objective WMC measure better predicted subjective TUT experiences than did the CFQ-MAL, a subjective self-report measure of everyday attentional failures (ruling out demand characteristics in our WMC effects and attesting to the validity of probed thought reports). Of most importance, however, individual differences in TUT rate accounted for half of WMC's shared variance with SART performance, suggesting that much of WMC's predic-

Table 3
Hierarchical Regression Analyses on Sustained Attention to Response Task Intraindividual Reaction Time Variation

Variable	*B*	*SE*	β	*t*	R^2
Predictors: WMC, TUT					
Step 1: WMC	−25.750	4.430	−.351	−5.813*	.123
Step 2: WMC, TUT	104.000	17.962	.336	5.790*	.230
Predictors: TUT, WMC					
Step 1: TUT	122.665	18.314	.396	6.698*	.157
Step 2: TUT, WMC	−20.397	4.260	−.278	−4.788*	.230
Predictors: CFQ-MAL, TUT					
Step 1: CFQ-MAL	0.510	0.224	.146	2.281*	.021
Step 2: CFQ-MAL, TUT	118.368	18.520	.382	6.391*	.164
Predictors: TUT, CFQ-MAL					
Step 1: TUT	122.417	18.376	.395	6.662*	.156
Step 2: TUT, CFQ-MAL	0.317	0.209	.091	1.516	.164
Predictors: WMC, CFQ-MAL, TUT					
Step 1: WMC	−25.717	4.437	−.350	−5.796*	.123
Step 2: WMC, CFQ-MAL	0.528	0.209	.151	2.520*	.145
Step 3: WMC, CFQ-MAL, TUT	98.799	18.143	.319	5.446*	.240
Predictors: CFQ-MAL, WMC, TUT					
Step 1: CFQ-MAL	0.510	0.224	.146	2.281*	.021
Step 2: CFQ-MAL, WMC	−25.874	4.389	−.353	−5.895*	.145
Step 3: CFQ-MAL, WMC, TUT	98.799	18.143	.319	5.446*	.240

Note. $N = 243$. WMC = working memory capacity; TUT = proportion self-reported task-unrelated thoughts; CFQ-MAL = Cognitive Failures Questionnaire–Memory and Attention Lapses. $N = 242$ for CFQ-MAL analyses.
* $p < .05$.

tive power is attributable to its reflecting people's ability to simply keep their thoughts focused on the task at hand, a notion consistent with our executive-attention view.

WMC and Executive Attention

Our individual-differences findings confirm key hypotheses from the executive-attention theory of WMC (e.g., Kane, Conway, et al., 2007), which holds that WMC's predictive power derives primarily from its tapping attention-control mechanisms that, among other functions, keep novel goals readily maintained to regulate ongoing behavior amid conflict. If goal-neglect errors arise through attention-control failures, and if many attention-control failures are complete enough to result in TUT experiences, then WMC-related variation in TUT rate should partially mediate WMC-related variation in performance. It did, and these findings seem inconsistent with a binding explanation of WMC-related variation in goal neglect (e.g., Wilhelm & Oberauer, 2006). If lower WMC subjects more often fail to act according to goals because they less effectively bind response productions to stimulus classes (e.g., "press key for animals"), then performance differences between lower and higher WMC subjects need not have anything to do with mind wandering, nor should WMC even predict TUT rates during cognitive tasks.

Important questions remain, however, regarding the SART variance explained by WMC independent of TUT rate. Kane and Engle (2003) concluded that (a) WMC predicts attention-task performance via goal-maintenance and competition-resolution mechanisms, the latter of which only engages subsequent to the former (e.g., in Stroop, resolving conflict between color and word dimensions only proceeds if the color-naming goal is accessible), and (b) higher WMC subjects are superior to lower WMC subjects in both processes. In the SART, subjects must not only keep the no-go goal in mind throughout long sequences of go trials but also successfully inhibit this prepotent response when required. Indeed, go-trial RTs preceding errors were 73 ms faster here than those preceding correct responses, whereas RTs preceding TUT reports were only 11 ms faster than those preceding on-task thoughts. Fast, erroneous responding clearly occurs even when subjects are reportedly task focused, presumably reflecting within- and between-subject variation in competition resolution.

We therefore suggest that WMC's TUT-independent prediction of SART performance is largely due to its relation to competition resolution. If so, two predictions follow: (a) A SART that induces weaker prepotencies to overcome should correlate less strongly with WMC (due to a minimization of competition-resolution variance) and (b) SART variance that is predicted by WMC should be more fully mediated by TUT rate, as subjects must maintain goal activation that is not externally reinforced. We are currently testing these predictions with SARTs that present mostly no-go trials—requiring no overt responses—so the "go" goal requires active maintenance but accurate responding requires little competition resolution. We believe that this experiment also tests further the binding theory (Oberauer et al., 2007). If WMC's non-TUT prediction of SART performance derives from S-R binding effectiveness instead of reflecting competition-resolution processes (Wilhelm & Oberauer, 2006), then a SART with weak prepotencies should still correlate substantially with WMC, because subjects must still bind no-go and go responses to stimulus categories. Moreover, SART variance predicted by WMC should not be mediated by TUT rate.

Theoretical and Methodological Issues in Mind Wandering

Our results not only inform WMC theory, but they also raise concerns about current theoretical and methodological approaches to mind wandering. Of most importance, a major theory of mind wandering seems to predict the reverse of our central WMC finding. Smallwood and Schooler (2006) argued that mind wandering draws heavily on WMC and executive resources, largely on the basis of findings that TUTs decrease during demanding tasks and that performance errors increase during TUTs. By this view, as primary tasks consume more resources, fewer remain to support mind wandering and vice versa. Moreover, subjects who have more resources available (e.g., higher WMC) should be able to mind wander more during ongoing tasks than should subjects with fewer resources (e.g., lower WMC).

Of course, we found the opposite: Lower WMC subjects mind wandered more during a demanding primary task than did higher WMC subjects (see also Kane, Brown, et al., 2007). We therefore suggest that TUTs represent an executive-control failure to maintain on-task thoughts and that the generation and persistence of TUTs do not require executive resources. Rather, TUTs are automatically and continually generated as part of the thought stream (e.g., Bar, 2007; James, 1890) in response to internal and external cues (e.g., Klinger, 1971), and executive-control processes keep these thoughts out of the focus of attention during resource-demanding tasks. Neuroscience research connecting TUTs to a default-mode network of the brain (e.g., Mason et al., 2007) suggests that mind wandering may be a return of attention to the type of thoughts produced while subjects are at rest. By this provisional view (which requires further refinement and test), TUTs either cause performance errors by displacing stimulus and goal representations from attentional focus or correlate with errors as a signal (or side effect) of failed attention control. Difficult tasks minimize TUTs because to meet task demands, they stimulate engagement of control processes, one function of which is to sustain conscious focus and actively prevent TUTs from occurring.

Regarding a secondary motivation for the present study, we failed to replicate prior findings that mind wandering varies with the conceptual-processing demands of ongoing tasks: TUT rates were equivalent for semantic and perceptual SARTs, rather than being reduced in the semantic task. It may be important that most experiments showing reduced TUTs during conceptual processing have involved intentional memory encoding, retrieval, or both (e.g., Smallwood, Baracia, et al., 2003). Such task requirements may encourage integration or associations across conceptually related items and thus provide a scaffold for maintaining on-task thought. The SART, in contrast, neither requires nor promotes such mental organization, as individual stimuli require independent judgments. Indeed, Smallwood et al. (2006) observed a reduced TUT rate for the semantic versus nonsemantic SARTs only for subjects instructed to commit the stimuli to memory, who therefore may have thought more elaboratively and cohesively about the stimuli.

Although our findings generally support the notion that variation in conscious thoughts predict (if not cause) some variation in task performance, they also indicate that mind wandering and performance errors are not interchangeable indices of attentional lapses, as some researchers have suggested (e.g., Smallwood, McSpadden, Luus, & Schooler, 2008; Smallwood et al., 2007). Responses to targets were appropriately withheld 42% of the time that subjects' thoughts were

off task and inappropriately committed 34% of the time that subjects' thoughts were on task. Moreover, as noted previously, RTs preceding errors were much faster than those preceding accurate responses, but RTs preceding TUTs were only slightly faster than those preceding on-task reports. Thus, there is more to executive-task performance than just goal neglect and mind wandering, and habit-based errors need not reflect only lapses of sustained attention (Kane & Engle, 2003; Logan & Cowan, 1984).

Conclusion

Our findings demonstrate the utility of subjective mind-wandering reports to the experimental and differential study of executive functions (see also Kane, Brown, et al., 2007; Smallwood & Schooler, 2006). Goal-neglect errors and some WMC-related differences in attention-task performance appear to stem in part from momentary failures of conscious thought control. As in the present experiment, further assessment of subjective experience during cognitive tasks (and especially off-task thoughts) should provide evidence for or against particular mechanistic views of executive control and its variation.

References

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11,* 280–289.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80,* 1–46.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21,* 1–16.

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84,* 822–848.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769–786.

Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York: Oxford University Press.

James, W. (1890). *The principles of psychology.* Boston: Holt.

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., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science, 18,* 614–621.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York: Oxford University Press.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132,* 47–70.

Klinger, E. (1971). *Structure and functions of fantasy.* New York: Wiley.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review, 91,* 295–327.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007, January 19). Wandering minds: The default network and stimulus-independent thought. *Science, 315,* 393–395.

Oberauer, K., Süβ, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York: Oxford University Press.

Reason, J. T., & Mycielska, K. (1982). *Absent minded? The psychology of mental lapses and everyday errors.* Englewood Cliffs, NJ: Prentice Hall.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). Oops: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neurospsychologia, 35,* 747–758.

Smallwood, J., Baracia, S. F., Lowe, M., & Obonsawin, M. C. (2003). Task-unrelated-thought whilst encoding information. *Consciousness and Cognition, 12,* 452–484.

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M. V., O'Connor, R. C., & Obonsawin, M. C. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition, 4,* 657–690.

Smallwood, J., McSpadden, M., Luus, B., & Schooler, J. W. (2008). Segmenting the stream of consciousness: The psychological correlates of temporal structures in the time series data of a continuous performance task. *Brain & Cognition, 66,* 50–56.

Smallwood, J., McSpadden, M., & Schooler, J. W. (2007). The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin & Review, 14,* 527–533.

Smallwood, J., Obonsawin, M., & Heim, D. (2003). Task unrelated thought: The role of distributed processing. *Consciousness and Cognition, 12,* 169–189.

Smallwood, J., Riby, L., Heim, D., & Davies, J. B. (2006). Encoding during the attentional lapse: Accuracy of encoding during the semantic sustained attention to response task. *Consciousness and Cognition, 15,* 218–231.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132,* 946–958.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245–251.

Sunderland, A., Harris, J. E., & Baddeley, A. D. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior, 22,* 341–357.

Unsworth, N., Heitz, R. C., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498–505.

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

Wilhelm, O., & Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus–response compatibility in choice reaction time tasks. *European Journal of Cognitive Psychology, 18,* 18–50.

Received January 8, 2008
Revision received August 27, 2008
Accepted August 28, 2008 ■